Overlapping neural representations for the position of visible and imagined objects

Amanda K. Robinson[a,b,c], Tijl Grootswagers[a,b,c], Sophia M. Shatek[a], Jack Gerboni[a], Alex O. Holcombe[a], Thomas A. Carlson[a,b]

[a] School of Psychology, University of Sydney, NSW, 2006, Australia

[b] Perception in Action Research Centre, Macquarie University, Sydney, NSW, Australia

[c] Department of Cognitive Science, Macquarie University, NSW, 2109, Australia

Abstract

Humans can covertly track the position of an object, even if the object is temporarily occluded. What are the neural mechanisms underlying our capacity to track moving objects when there is no physical stimulus for the brain to track? One possibility is that the brain "fills-in" information about imagined objects using internally generated representations similar to those generated by feed-forward perceptual mechanisms. Alternatively, the brain might deploy a higher order mechanism, for example using an object tracking model that integrates visual signals and motion dynamics (Kwon et al., 2015). In the present study, we used electroencephalography (EEG) and time-resolved multivariate pattern analyses to investigate the spatial processing of visible and imagined objects. Participants tracked an object that moved in discrete steps around fixation, occupying six consecutive locations. They were asked to imagine that the object continued on the same trajectory after it disappeared and move their attention to the corresponding positions. Time-resolved decoding of EEG data revealed that the location of the visible stimuli could be decoded shortly after image onset, consistent with early retinotopic visual processes. For processing of unseen/imagined positions, the patterns of neural activity resembled stimulus-driven mid-level visual processes, but were detected earlier than perceptual mechanisms, implicating an anticipatory and more variable tracking mechanism. Encoding models revealed that spatial representations were much weaker for imagined than visible stimuli. Monitoring the position of imagined objects thus utilises similar perceptual and attentional processes as monitoring objects that are actually present, but with different temporal dynamics. These results indicate that internally generated representations rely on top-down processes, and their timing is influenced by the predictability of the stimulus. All data and analysis code for this study are available at https://osf.io/8v47t/.

Introduction

Internally-generated representations of the world, as opposed to stimulus-driven representations, are important for day-to-day tasks such as constructing a mental map to give a stranger directions, remembering where you last saw a lost item, or tracking the location of a car that becomes occluded by another vehicle. In these cases, there is little or no relevant perceptual input, yet the brain successfully constructs a picture of relevant visual features such as object form and spatial position. Internally-generated representations have been studied with tasks involving imagery, mental rotation, and perception of occluded objects. Such tasks seem to involve different cognitive strategies, yet their neural mechanisms have some similarities. It is clear that internally-generated representations rely on similar brain regions to stimulus-driven representations (Lee et al., 2012; Reddy et al., 2010) but they appear to have different temporal dynamics (Dijkstra et al., 2018), raising the question of how exactly these internal representations are formed.

Top-down processing appears to play an important role in generating internally representations. Current theories of mental imagery are based on similarities between perception and imagery, with a greater focus on bottom-up processing in perception and top-down processing in imagery (for review, see Pearson, 2019). Neuroimaging work has shown increases in brain activation within early visual cortical regions when participants engage in imagery, in a similar way to viewing the same stimuli (Kosslyn et al., 1993; Le Bihan et al., 1993), but there is more perception-imagery overlap in higher level brain regions such as ventral temporal cortex (Lee et al., 2012; Reddy et al., 2010). Imagery involves greater flow of information from fronto-parietal to occipital regions than perception, indicating that top-down or feedback-like processes mediate internally generated representations (Dentico et al., 2014; Dijkstra et al., 2017; Mechelli, 2004). During imagery, neural activation within the ventral stream is consistent with generative feedback models of information flow from higher-level to low-level visual regions (Breedlove et al., 2020). Consistent with this account, recent work using magnetoencephalography and time-resolved decoding showed that imagery of faces and houses involves similar patterns of activation as viewing those stimuli, but with different temporal dynamics (Dijkstra et al., 2018). In the Dijkstra et al. (2018) study, imagery-related processing was delayed and more diffuse than perception, which showed multiple distinct processing stages. A follow-up study suggested that the order of perceptual processes is reversed in imagery (Dijkstra et al., 2019). Together, these results suggest that imagery uses at least some of the same mechanisms as perception but is initiated in higher-level brain regions rather than being driven by perceptual input.

Another mechanism originating in higher-level brain regions that might be intrinsically linked to internal representations is spatial attention. Directing attention to a location enhances processing of stimuli that appear there (Posner, 1980). Reduced amplitude alpha-band (~10Hz) oscillations in visual cortex have been linked to covertly attending to a specific region in space (Worden et al., 2000). Additionally, time-resolved decoding has found that attended locations could be decoded from the neural signal even before a stimulus appeared (Goddard et al., 2019). It follows that spatial imagery tasks that require internal representations of objects with specific positions or orientations, such as in occlusion or mental rotation, might also inherently involve spatial attention. Indeed, alpha-band activity has been found to track spatial locations held in working memory (Foster et al., 2016). Interestingly, a recent study found evidence that imagery and perception share neural processes in the alpha-band frequency linked to high-level visual processing, using a task that did not involve an explicit spatial component (Xie et al., 2020). Imagery and spatial attention therefore seem to share common features; they both appear to rely on top-down processing, with one consequence that perception seems to have higher spatial resolution than both spatial attention (Intriligator and Cavanagh, 2001) and imagery (Breedlove et al., 2020). It is very difficult to untangle the contributions of perceptual processes and spatial attention to internal representations. It seems likely that imagery involves mechanisms related to perception and attention, relying on top-down processing from high-level brain regions.

One aspect that is likely to affect the top-down generation of an internal representation is how it is prompted and the ability to predict its features in advance, for example when objects become occluded. The processes underlying the representation of occluded objects may be closely related to those in conventional imagery tasks (Nanay, 2010). However, there are some important differences between imagery and occlusion. Imagery can be prompted from short- or long-term memory, which involve different brain regions (Ishai, 2002). Mental imagery can be considered to encompass situations in which there is a visual percept that is not produced via current sensation. In this view, representations held in working memory can therefore involve mental imagery; indeed, percepts in working memory resemble those arising from mental rotation (Albers et al., 2013). In conditions of occlusion, as well as in the case of visual working memory, there is usually some sensory support, such as from a fragment of the object not occluded or full view of the object immediately before occlusion. One possibility is that internally generated representations utilise the same brain networks as perceptual representations, but the temporal dynamics vary with the ability to predict and anticipate details of the stimulus to be generated.

Tracking the position of a predictably moving object is a common task that may share some top-down processes with static imagery tasks. In particular, prediction is likely to play an important role in both imagery and visual tracking. The ability to predict the movement of a stimulus influences perceptual processing during visual tracking (Blom et al., 2020; Hogendoorn and Burkitt, 2018). Hogendoorn & Burkitt (2018) measured EEG from participants who viewed an apparent motion stimulus that was predictable or unpredictable in its motion trajectory. The position-specific representations occurring 80-90ms after stimulus onset were unaffected by motion predictability, but a later stage of processing (typically 140-150ms after a stimulus is presented) occurred earlier for predictable relative to random sequences by approximately 16ms (Hogendoorn and Burkitt, 2018). Predictability therefore has a marked effect on the temporal dynamics of spatial representations for visible stimuli. For an object appearing in an unpredictable location, the resulting position representation must be a combination of the internal representation of the expected location and the stimulus-driven response to the actual object location (Blom et al., 2020). Disentangling a stimulus prediction from a stimulus-driven response is an important next step in understanding how and when internal representations are formed. Anticipatory mechanisms are likely to influence internally generated spatial representations, but might interact with other effects, for example the delayed processes observed during imagery (Dijkstra et al., 2018).

In the current study, to understand the nature of internal representations in the brain, we investigated the neural processes underlying visual tracking for visible and imagined objects. Participants covertly tracked the position of a simple moving stimulus and kept tracking its trajectory after it disappeared. Using spatial imagery allowed us to assess the temporal dynamics of internal representations during object tracking in the absence of a stimulus-driven response. EEG and time-resolved multivariate pattern analysis were used to assess the position-specific information contained within the neural signal during visible and imagined stimulus presentations. We successfully decoded the position of the stimuli from all phases of the task. Our results show that the visible and imagined stimuli evoked the same neural response patterns, but with very different temporal dynamics. Further, multivariate encoding models revealed that the spatial representations of imagined stimuli were much weaker than those of visual stimuli. These findings suggest that overlapping mid- and high-level visual processes underlie perceptual and internally generated representations of spatial location, and that these are pre-activated in anticipation of a stimulus.

Methods

All stimuli, data and analysis code are available at https://osf.io/8v47t/. The experiment consisted of two types of sequences: a pattern estimator and the experimental task. In the pattern estimator sequences, the order of the stimuli was unpredictable, whereas in the experimental task the order was predictable. The pattern estimator sequences were used to obtain position-specific EEG signals that were unlikely to be affected by eye-movements, and were subsequently used to detect position signals in the experimental task.

Participants

Participants were 20 adults recruited from the University of Sydney (12 females; age range 18-52 years) in return for payment or course credit. The study was approved by the University of Sydney ethics committee and informed consent was obtained from all participants. Four participants were excluded from analyses due to excessive eye movements during the pattern estimator sequences.

Stimuli and design

While participants maintained fixation in the centre of the monitor, a stimulus appeared in six distinct positions 4 degrees of visual angle from fixation. The stimulus positions were 0°, 60°, 120°, 180°, 240° and 300° relative to fixation. The stimulus was a black circle with a diameter of 3 degrees of visual angle. Six unfilled circles acted as placeholders, marking all possible positions throughout the trial. Every stimulus presentation was accompanied by a 1000 Hz pure tone presented for 100 ms via headphones. All stimuli were presented using Psychtoolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) in MATLAB. In total, there were 8 blocks of trials, each of which contained two pattern estimator sequences and 36 experimental task sequences.

Pattern estimator

The pattern estimator sequences were designed to extract stimulus-driven position-specific neural patterns from the EEG signal. Participants viewed 16 pattern estimator sequences (2 per block), each of which consisted of 10 repetitions of the 6 stimulus positions (Figure 1a). The order of stimuli was randomised to ensure that for a given stimulus position, the preceding and following stimuli would not be predictive of that position; for example, comparing the neural patterns evoked by positions 1 and 2 could not be contaminated by preceding and following stimuli because they could both be preceded and followed by all six positions. Each stimulus was shown for 100ms and was followed by an inter-stimulus interval of 200ms. Onset of the stimulus was accompanied by a 100ms tone. Participants were instructed to passively view the stimuli without moving their eyes from the fixation cross in the centre of the screen.

The stimuli were presented in unpredictable patterns so there was no regularity in the positions of the previous or following stimuli to contribute to the neural patterns extracted for each position. Additionally, the random sequences ensured that any eye movements would be irregular and thus unlikely to contribute to the extracted neural signal. Previous work has shown that even the fastest saccades typically take at least 100ms to initiate (Fischer and Ramsperger, 1984). Furthermore, eye movements do not appear to affect decoding of magnetoencephalography data until 200ms after a lateralised stimulus is presented (Quax et al., 2019). Our 100ms stimulus duration was therefore unlikely to generate consistent eye movements that would affect the early, retinotopic EEG signal of stimulus position.

To assess whether participants complied with the fixation instruction, we assessed the EEG signal from electrodes AF7 and AF8 (located near the left and right eye, respectively) as a proxy for electrooculogram measurements. We calculated the standard deviation of the AF7 and AF8 signals across each of the 16 sequences and then averaged the deviation for the two electrodes. If a participant's average deviation across the 16 sequences exceeded 50μV, that individual was considered to be moving their eyes or blinking too often, resulting in poor signal. An amplitude threshold of 100 μV is commonly used to designate gross artefacts in EEG signal (Luck, 2005), so we adopted an arbitrary standard deviation threshold of 50 μV (50% of the typical amplitude threshold) to indicate that there were too many artefacts across the entire pattern estimator sequences. Four participants exceeded this standard deviation threshold (M = 72.72μV, range = 63.93-82.70μV) and were excluded from all analyses. For each of the remaining 16 participants, the median deviation was well below this threshold (M = 25.92μV, SD = 5.64μV, range = 16.06-37.62μV). Thus, the four excluded participants had far more signal artefacts (probably arising from eye movements) than the other participants.

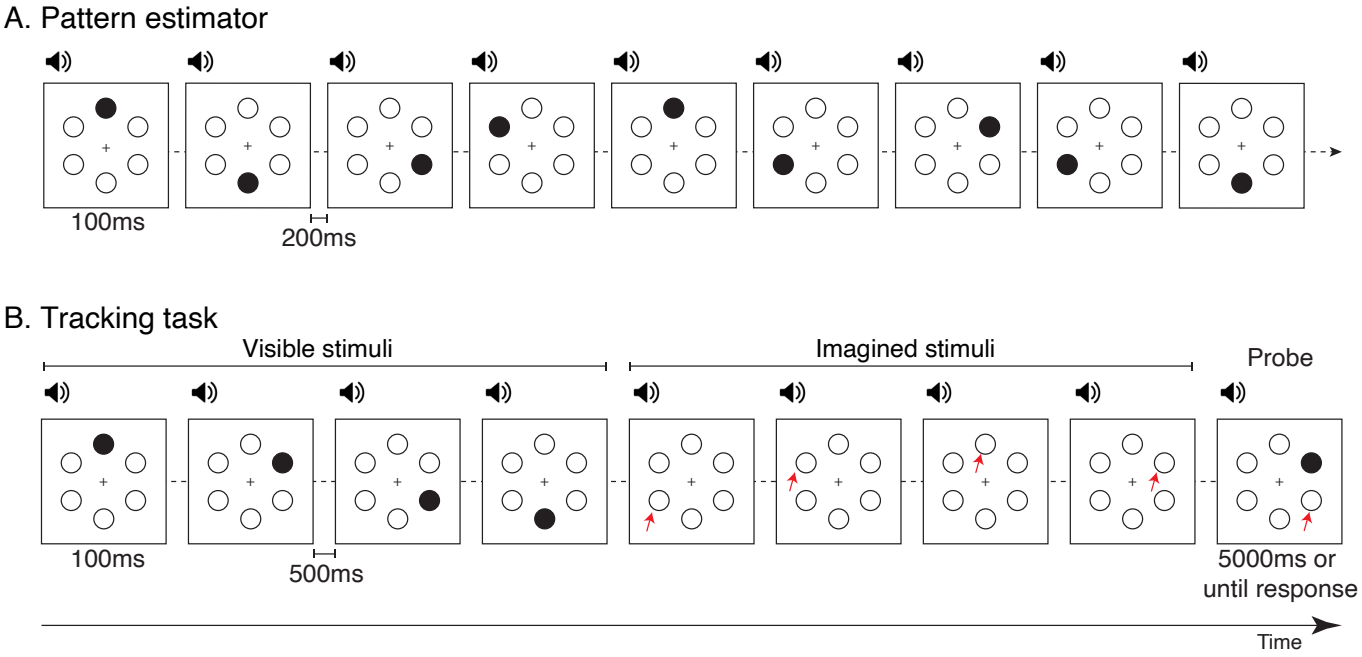

Figure 1. Stimuli and design. A) Pattern estimator. Participants passively viewed rapid sequences in which a black circle stimulus appeared in six locations in random order. A tone accompanied every stimulus onset. B) Tracking task. The stimulus was presented in different locations in predictable sequences. After 4-6 visible locations, participants had to track the location of the "imagined" stimulus by imagining the continuation of the sequence. A tone accompanied every stimulus onset. During the 4-6 "imagined" positions, the auditory stimulus continued at the same rate, but only the six placeholder locations were shown. At the end of the sequence, a probe appeared, and participants had to respond if it was in the expected position or whether it was trailing or leading the sequence. This example shows a clockwise sequence with trailing probe. Red arrows (not shown in experiment) designate the expected position of the imagined stimulus.

Tracking task

For the experimental task, participants viewed sequences consisting of 4-6 visible stimuli and 4-6 "imagined" presentations simulating occluded stimuli (Figure 1b). The positions of the visible stimuli were predictable, presented in clockwise or counter-clockwise sequences. Participants were asked to covertly track the position of the stimulus, and to continue imagining the sequence of positions when the stimulus was no longer visible. At the end of each sequence, there was a 1000 ms blank screen followed by a probe stimulus that was presented in one of the 6 locations. Participants categorised this probe as either (1) *trailing*: one position behind in the sequence, (2) *expected*: the correct location, or (3) *leading*: one position ahead in the sequence. Participants responded using the Z, X or C keys on a keyboard, respectively. Each response was equally likely to be correct, so chance performance was 33.33%.

EEG recordings and preprocessing

EEG data were continuously recorded from 64 electrodes arranged in the international 10–10 system for electrode placement (Oostenveld and Praamstra, 2001) using a BrainVision ActiChamp system, digitized at a 1000-Hz sample rate. Scalp electrodes were referenced to Cz during recording. EEGLAB (Delorme and Makeig, 2004) was used to pre-process the data offline, where data were re-referenced to the average of all electrodes. We filtered the data using a Hamming windowed sinc FIR filter with highpass of 0.1Hz and lowpass of 100Hz and then downsampled to 250Hz as in our previous work (Grootswagers et al., 2019; Robinson et al., 2019). Epochs were created for each stimulus presentation ranging from -200 to 1000ms relative to stimulus onset. No further preprocessing steps were applied. Mean neural responses of these epochs show clear event-related potentials in response to the visual and auditory stimuli (see Supplementary Material S5).

Decoding analyses

An MVPA decoding pipeline (Grootswagers et al., 2017) was applied to the EEG epochs to investigate position representations of visible and imagined stimuli. All steps in the decoding analysis were implemented in CoSMoMVPA (Oosterhof et al., 2016). A leave-one-block-out (i.e., 8-fold) cross-validation procedure was used for all time-resolved decoding analyses. For each time point, a linear discriminant analysis (LDA) classifier was trained using the pattern estimator data to distinguish between all pairs of positions. LDA covariance was regularised by .01. Channel voltages from the 64 EEG channels were used as features for classification. Each classifier was trained with balanced numbers of trials per stimulus position from the pattern estimator sequences. The classifier was then tested separately on the visible and imagined positions in the experimental task. This provided decoding accuracy over time for each condition. At each time point, mean pairwise accuracy was tested against chance (50%). Importantly, because all analyses used the randomly-ordered pattern estimator data for training the classifier, above chance classification was very unlikely to arise from the predictable sequences or eye movements in the experimental task. For the tracking task, all sequences were included in the decoding analyses regardless of whether the participant correctly classified the position of the probe (i.e., correct and incorrect sequences were analysed). When only correct trials were included, the trends in the results remained the same (see Supplementary Material S1).

To assess whether neighbouring stimulus positions evoked more similar neural responses, we also calculated decoding accuracy as a function of the distance between position pairs. Each position pair had a radial distance of 60°, 120° or 180° apart. There were six pairs with a distance of 60° (e.g., 0° vs 60° , 60° vs 120°, 0° vs 300°), six pairs with a distance of 120° (e.g., 0° vs 120°, 60° vs 180°), and three pairs

with a distance of 180° (directly opposing each other, e.g., 0° vs 180°, 60° vs 240°). Decoding accuracy for each pair distance was calculated as the mean of all relevant pair decoding and compared to chance (50%).

As a final set of decoding analyses, time generalisation (King and Dehaene, 2014) was used to assess whether the patterns of informative neural activity occurred at the same times for the pattern localiser and the visible and imagined stimuli on the tracking task. Classification was performed on all combinations of time points from the pattern estimator epochs and the visible or imagined epochs. Classifiers were trained on all trials from the pattern estimator sequences and tested on visible and imagined stimulus positions.

Multivariate encoding analyses

As exploratory analyses prompted by reviewers' comments, we used forward encoding models to investigate the spatial selectivity of visible and imagined representations across time. Encoding models can be used with neuroimaging data to investigate neural encoding of many visual feature dimensions (Sprague and Serences, 2015). Such models have been applied to fMRI data to assess encoding of features such as colour (Brouwer and Heeger, 2009), orientation (Scolari et al., 2012) and position (Sprague and Serences, 2013). These methods have also been adapted for use with EEG using neural responses in the frequency (Foster et al., 2016; Garcia et al., 2013) and temporal domains (Smout et al., 2019; Tang et al., 2020, 2018). Here, we used data from the pattern estimator task and multivariate linear regression to model the EEG responses per time point as a weighted sum of six position "channels", each tuned to the experimental positions of 60°, 120°, 180°, 240°, 300°, 360°. These models were then used to estimate channel responses for visible and imagined positions on the tracking task, in order to assess the selectivity of the position representations. Analyses were adapted from encoding analyses of EEG data implemented in Smout, Tang, Garrido and Mattingley (2019) using scripts on the Open Science Framework (https://doi.org/10.17605/osf.io/a3pfq) and functions from https://github.com/Pim-Mostert/decoding-toolbox.

Results from encoding analyses are activations (rather than predictions as in decoding), so encoding is more sensitive to noise and artefacts in the data. Additional data cleaning steps were applied to remove noise and artefacts. After epoching, we interpolated electrodes that exceeded 5 standard deviations from the mean kurtosis value. For one dataset, we interpolated one additional channel that remained extremely noisy by visual inspection. In total, six or fewer channels were interpolated per dataset (<10%, M = 3.5, SE = .56). To remove artefacts, any epochs that exceeded +/-100μV at any time across the epoch

were excluded from the analyses, and for every training/testing fold we randomly subsampled the remaining clean trials so there were equal numbers per position for the pattern estimator (total *M* = 640.13*, SE* = 46.77) and equal numbers per position, condition (visible/imagined) and movement direction (clockwise/counter-clockwise) on the tracking task (total *M* = 1957.50, *SE* = 177.11). These steps ensured that the position encoding analyses were based on clean EEG data and could not be biased due to unequal trial numbers.

For each participant and time point, encoding models were trained using four-fold cross-validation, each time training on 75% of the pattern estimator data and testing on 25% of the test data. This procedure was repeated 100 times with different trial subsampling every time (Smout et al., 2019). These analyses resulted in response profiles across the six stimulus positions (encoding "channels"; 0, 60, 120, 180, 240 and 300°) for each trial. Channel responses were then realigned to positions -120 to 180°, where the 0° position channel reflected the correct stimulus position for the trial. We expected that the position representations on the tracking task might also include representations for the previous and next stimuli in the sequence, so we collated the data separately for clockwise and counterclockwise sequences and relabelled the position channels to reflect position relative to stimulus movement. Thus, channels +60, +120 and +180 degrees reflect positions of the next three stimuli in the sequence, and channels -60 and -120 reflect positions of the preceding two stimuli. Mean position channel responses were then calculated per time point for the visible and imagined stimuli.

To assess the position representations in the neural signal, exponentiated cosines were fit to the encoding response profiles across the six position channels for each participant, condition and time point using the equation:

$$y(x) = A * e^{\kappa(\cos(x-\mu)-1)} + B$$

which models the expected response profile for position angle *x* with a distribution with amplitude *A* (peak response amplitude) with κ concentration (sharpness of the distribution, analogous to standard deviation) that clusters around μ (peak of the function) with baseline offset *B*. The fitting was implemented using lsqcurvefit in MATLAB with starting values *A* = 0.2 (range -5 to 10), κ = 1 (0 to 10), μ = 0 (-60° to 60°) and *B* = 0 (-5 to 2). We analysed the amplitude *A* and peak μ over time for position representations of visible and imagined stimuli.

Statistical inference

To assess the evidence that decoding performance or parameter values differed from chance, we calculated Bayes factors (Dienes, 2011; Jeffreys, 1961; Kass and Raftery, 1995; Rouder et al., 2009; Wagenmakers, 2007). A JZS prior (Rouder et al., 2009) was used with a scale factor of 0.707, meaning that for the alternative hypothesis of above-chance decoding, we expected to see 50% of parameter values falling within -.707 and .707 standard deviations from chance (Jeffreys, 1961; Rouder et al., 2009; Wetzels and Wagenmakers, 2012; Zellner and Siow, 1980). The Bayes factor (BF) indicates the probability of obtaining the group data given the alternative hypothesis relative to the probability of the data assuming the null hypothesis is true. We used thresholds of BF > 3 and BF > 10 as increasing evidence for the alternative hypothesis, and BF < 1/3 as evidence in favour of the null hypothesis (Jeffreys, 1961; Kass and Raftery, 1995; Wetzels et al., 2011). BFs that lie between those values indicate insufficient evidence to favour either of the two hypotheses.

Results

Behavioural results

Participants performed well on the tracking task, with high mean accuracy for all probe positions (Fig 2A). Response time was calculated within participant as the mean correct response time per probe position. At the group level, response time was faster for the expected probe position relative to the unexpected probe positions (trailing or leading) (Fig 2B). These results indicate that on most trials participants knew where the probe was meant to appear, which required tracking the expected location of the object. Evidently, participants allocated their attention appropriately to the expected position of the stimulus during the imagined portion of the tracking task.

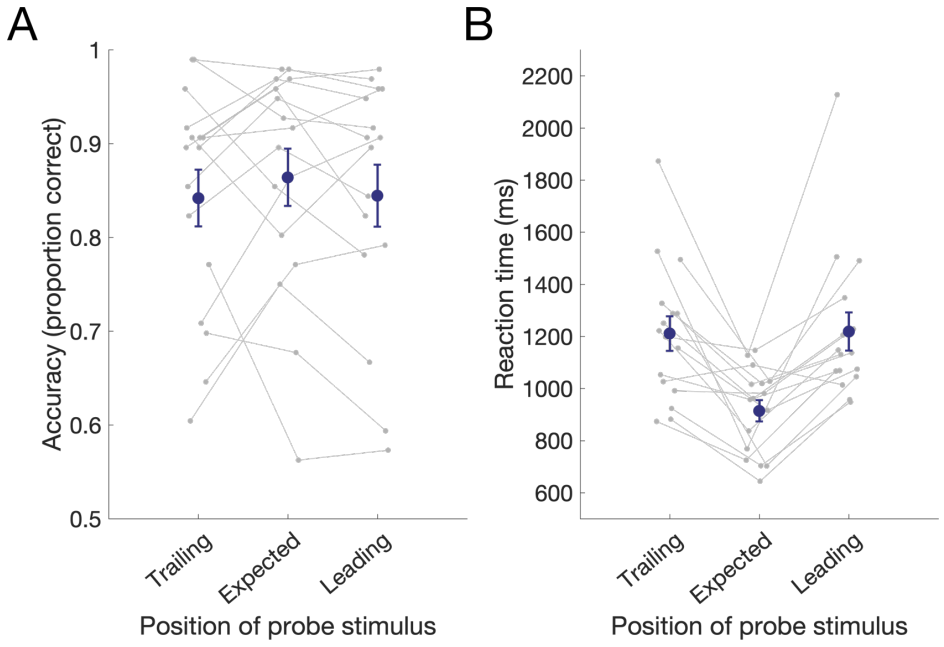

Figure 2. Behavioural results. A) Accuracy, and B) Response time on the tracking task as a function of final probe position. Individual participant data are plotted in grey, with group mean in navy. Error bars depict one standard error of the mean across participants (*N* = 16).

Position decoding using the pattern estimator sequences

The pattern estimator sequences were designed to extract position-specific neural patterns of activity from unpredictable visible stimuli. Time-resolved multivariate pattern analysis (MVPA) was applied to the EEG data from the pattern estimator, which revealed that stimulus position could be decoded above chance from approximately 68ms after stimulus onset and peaked at 150ms (Figure 3), consistent with initial retinotopic processing of position in early visual areas (Di Russo et al., 2003; Hagler et al., 2009). To assess how the physical distance between stimulus positions influenced the neural patterns of activity, we compared the pairwise decodability of position according to the relative angle between stimulus position pairs (i.e., angle of 60°, 120° or 180° between two stimulus positions). The greatest decoding performance was observed for larger angles between stimulus positions.

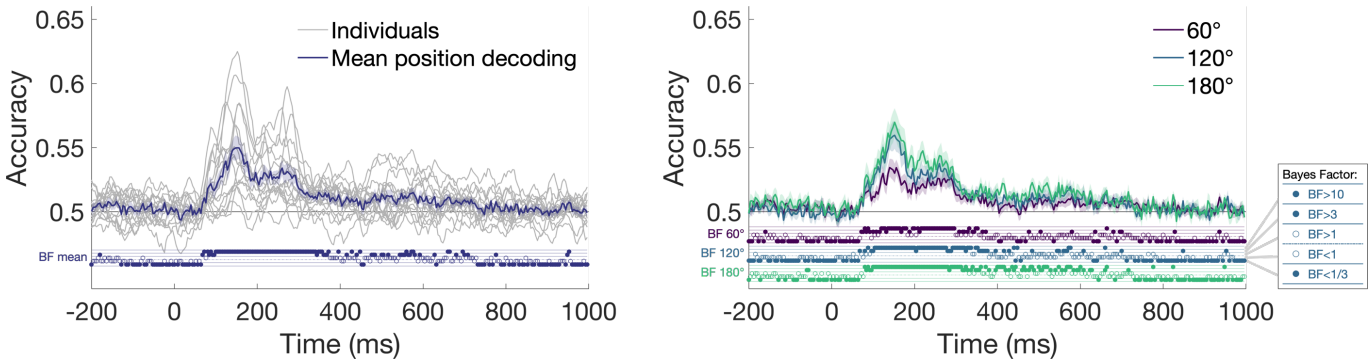

Figure 3. Position decoding using pattern estimator sequences. Left plot shows group mean decoding and smoothed individual participant decoding for all pairs of positions, and right plot shows mean position decoding as a function of the angular distance between stimulus pairs. Shaded areas show standard error across participants (*N* = 16). Thresholded Bayes factors (BF) for above-chance decoding are displayed above the x-axes for every time point as an open or closed circle in one of four locations (see inset).

Position decoding on the tracking task

To assess the similarity in position representations for visible and imagined (simulated occluded) stimuli, the classifier was trained on data from the visible pattern estimator stimuli and tested on data from the tracking task for the visible and imagined stimuli. Crucially, position could be decoded for both visible and imagined stimuli, suggesting that similar neural processes underpin perceptual and internal representations of stimulus position. For visible stimuli, the pattern of decoding results echoed those of the pattern estimator, with decoding evident from approximately 76ms and peaking at 152ms,

presumably reflecting visual coding of position in ventral visual areas of the brain (Figure 4a, left). When decoding was split according to the distance between the pair of positions, results looked similar to the pattern estimator results (Figure 4A, right).

A different pattern of results was observed for the imagined stimuli. Here, decoding was not above chance until approximately 152ms and consisted of a low, broad "peak" (Figure 4B). There was considerable variation in decoding accuracy across participants (Figure 4B, left; see also Supplementary Material S6). Although decoding accuracy was low, there was considerable evidence that accuracy was above chance (see Supplementary Material S2 for Bayes Factors in more detail). Reliable above chance cross-decoding from the visible pattern estimator stimuli to the imagined stimuli on the tracking task indicates that overlapping processes underlie stimulus-driven and internally-generated representations of spatial location. But this decoding of the internal representation of position was later and less accurate than position decoding for visible stimuli. Similar to the pattern estimator and visible decoding results, positions that were further apart were more decodable (Figure 4b, right). Notably, neighbouring positions (60° apart) showed little evidence of position decoding, suggesting that the representations of position were spatially diffuse for the imagined stimuli, unlike for the visible stimuli.

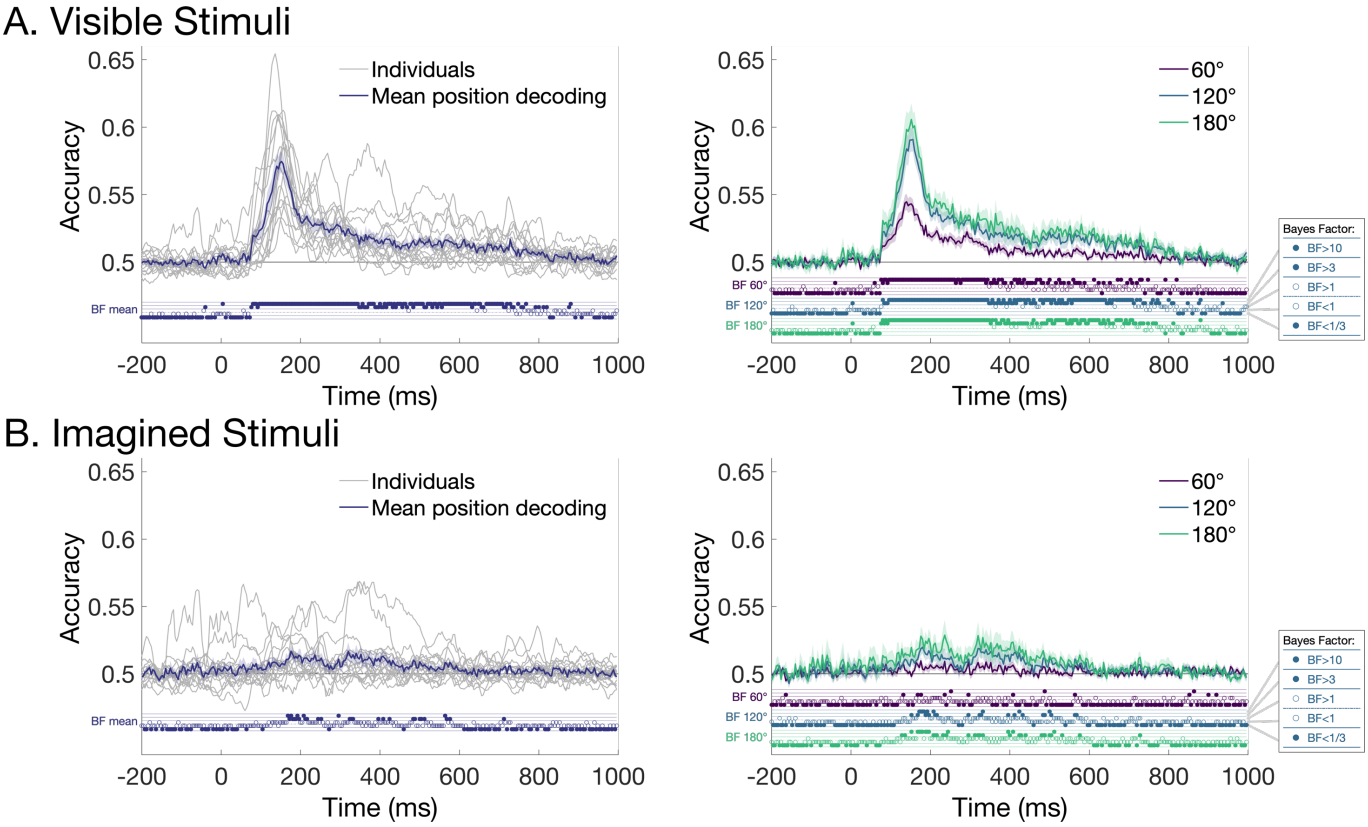

Figure 4. Position decoding from object tracking task. A) Visible stimuli. B) Imagined stimuli. Left plots show group mean decoding and smoothed individual participant decoding for all pairs of positions, and right plots show mean position decoding

as a function of the angular distance between position pairs. Shaded areas show standard error across participants (*N* = 16). Thresholded Bayes factors (BF) for above-chance decoding are displayed above the x-axes for every time point as an open or closed circle in one of four locations (see inset).

The previous analyses were performed using electrodes covering the whole head, which meant that there was a possibility that non-neural artefacts such as eye movements might contribute to the classification results (Quax et al., 2019). Saccadic artefacts tend to be localised to frontal electrodes, close to the eyes (Lins et al., 1993). To assess if the EEG signal contributing to the position-specific neural information originated from posterior regions of the brain (e.g., occipital cortex), as expected, we conducted the same time-resolved decoding analyses using a subset of electrodes from the back half of the head. We used 28 electrodes that were likely to pick up the largest signal from occipital, temporal and parietal areas (and were less likely to be contaminated with frontal or muscular activity). The electrodes were CPz, CP1, CP2, CP3, CP4, CP5, CP6, Pz, P1, P2, P3, P4, P5, P6, P7, P8, POz, PO3, PO4, PO7, PO8, Oz, O1, O2, TP7, TP8, TP9 and TP10. As can be seen in Figure 5, the same trend of results was seen using this subset of electrodes compared with the whole head analyses in Figure 4. Specifically, Bayes Factors revealed evidence that position of imagined stimuli was decodable approximately 136-244 ms, which is slightly earlier than the whole brain results. Decoding was also most evident for positions that were a distance of 120° or 180° apart (Figure 5b). Interestingly, imagery decoding was more prolonged for the whole-brain decoding than posterior analyses, which could reflect higher-order cognitive processing of stimulus position in more anterior regions of the brain, or increased power due to more features (electrodes) included in the whole brain analysis. Analyses restricted to frontal electrodes showed later, more diffuse coding for visible stimuli relative to the posterior analysis, and little evidence for position coding of imagined stimuli (see Supplementary Material S3). Thus, position-specific neural information for visible and imagined stimuli was evident specifically over posterior regions of the brain, consistent with visual cortex representing stimulus-driven and internal representations of spatial location.

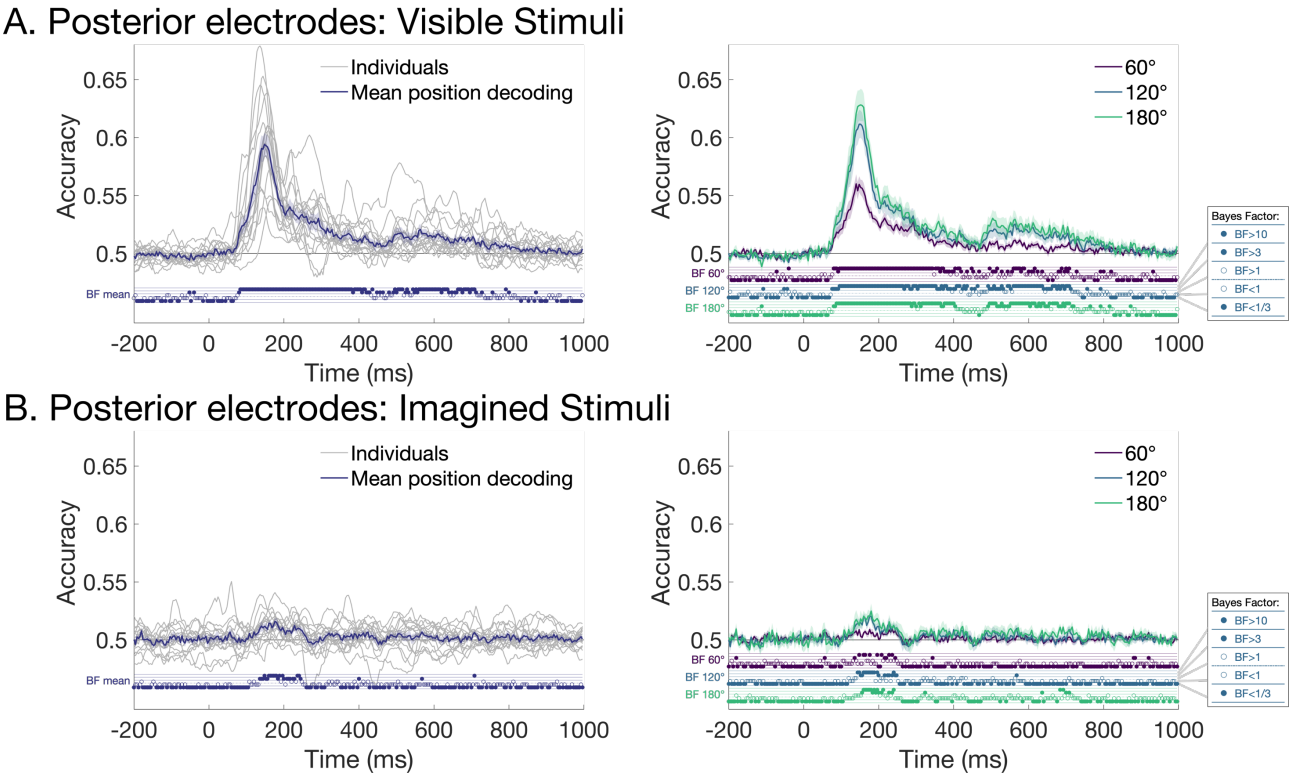

Figure 5. Position decoding from object tracking task using only posterior electrodes. A) Visible stimuli. B) Imagined stimuli. Left plots show group mean decoding and smoothed individual participant decoding for all pairs of positions, and right plots show mean position decoding as a function of the angular distance between stimulus pairs. Shaded areas show standard error across participants (*N* = 16). Thresholded Bayes factors (BF) for above-chance decoding are displayed above the x-axes for every time point as an open or closed circle in one of four locations (see inset).

The results of the time-resolved analyses showed that position-specific neural patterns for visible stimuli generalised to imagined stimuli, but with different temporal dynamics. To assess the possibility that neural processes were more temporally variable for imagined than for visible stimuli, we performed whole brain (64-channel) time-generalisation analyses by training the classifier on all time points of the pattern estimator and testing on all time points from the tracking task. As expected, position could be decoded from both visible and imagined stimulus presentations, but with marked differences in their dynamics (Figure 6). For the visible stimuli, most of the above-chance decoding was symmetric on the diagonal, indicating that the position-specific processes occurred at approximately the same time for visible stimuli in the pattern localiser and the tracking task (Figure 6A, top), even though the inter-stimulus intervals for stimuli in the training and test sets were different. Interestingly, there was also some above-diagonal decoding indicating that some neural signals observed in the pattern localiser occurred substantially *earlier* in the tracking task, which may reflect prediction based on the previous stimuli. Also likely reflecting anticipation of the stimulus position, generalisation occurred for time points prior to onset of the visible stimulus in the tracking task. About 800-1000ms after the tracking stimulus was presented,

there is some evidence of below chance decoding, indicating a different stimulus position was systematically predicted. This is likely to reflect processing of the next stimulus in the tracking task, which was presented at 600ms on the plot (stim +1 vertical line).

Time generalisation for the imagined stimulus position was not centred on the diagonal, reflecting different temporal dynamics for the predicted internal representations than for the stimulus-driven processing of the pattern estimator. Decoding generalisation was also much more diffuse and relied on processes approximately 120-750 ms after stimulus onset in the pattern estimator (Figure 6A, middle). Decoding again preceded the onset of the tone in the tracking task, reflecting an anticipation effect. There was also below chance decoding at later time points, indicating that the classifier was predicting a different stimulus position at times when the next stimulus would be processed. Comparison between visible and imagined position showed higher decoding for the imagined stimuli preceding the tone, but higher decoding for the visible stimuli after the stimulus and tone were presented (Figure 6A, bottom).

The dynamics of the time generalisation results give insight into the processing underlying perceptual and imagined position representations. Using decoding models trained on the pattern estimator at 140-160ms (approximately the time of peak position decoding), we looked at decoding accuracy for each time point on the tracking task. It is clear that visible representations show stimulus-evoked position specific responses, with largest decoding at the same time period as the training times (Figure 6B). Imagined representations, however, show much more diffuse responses that ramp up earlier than those of visible stimuli, with imagined decoding highest before 0ms, the time of the tone. Interestingly, this plot resembled within-condition decoding results (i.e., training and testing on visible or imagined stimuli from the training task; see Supplementary Material S4).

The time generalisation results show that position representations seem to emerge earlier for imagined than visible stimuli. For peak decoding times per participant (Figure 6C, left), visible position was most separable when training and testing approximately the same time points (about 150ms), whereas imagined position relied on later training than testing times, and showed much more variability across participants. To further assess peak decoding times, we bootstrapped the group 1000 times with replacement and calculated the times of peak generalisation to assess the distribution. Figure 6C (right) shows that visible decoding showed training and testing peaks at approximately 150ms with very little variation across the 1000 iterations. Imagined representations, by contrast, peaked after 300ms for training and 0ms for testing. Finally, assessing decoding accuracy by training-testing lag revealed that

imagined decoding was higher when training on later time points than testing times, whereas visible decoding was highest at approximately 0 ms offset (i.e., same training and testing times; Figure 6D). These results suggest that imagined representations rely on high level perceptual and cognitive processes that are implemented earlier in time. Overall, the time generalisation results suggest that during the imagined stimulus portion of the tracking task, which relied on internal representations of position, the neural dynamics were more anticipatory and variable than perceptual processes.

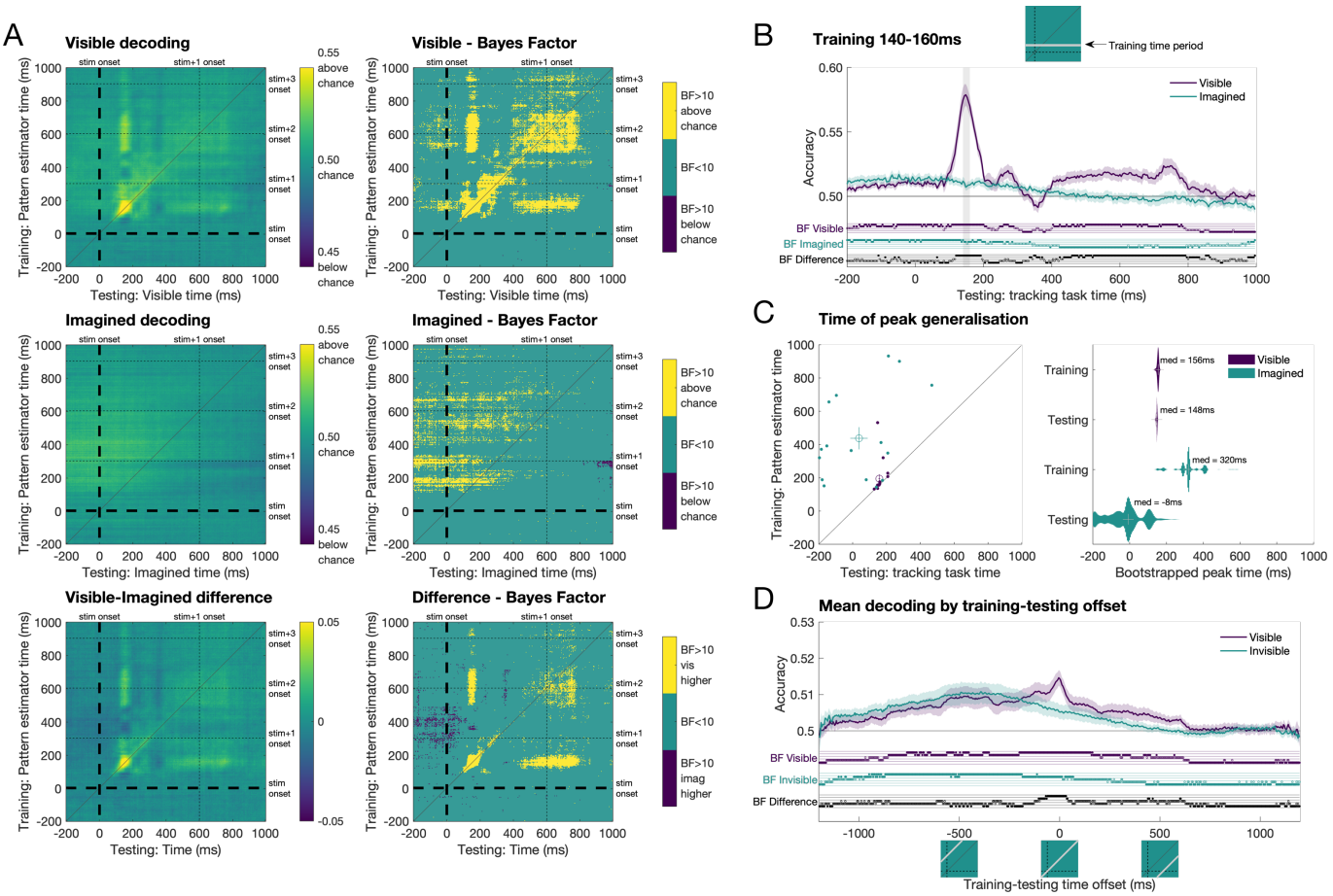

Figure 6. Time generalisation results. A) Decoding stimulus position for visible stimuli and imagined stimuli. Left plots show decoding for visible, imagined and visible-imagined difference, and right plots show associated Bayes Factors. Decoding was performed by training on data from the pattern estimator sequences of visible stimuli and testing on the experimental trials, for all pairs of time points. B) Decoding accuracy using training times 140-160ms on the pattern estimator and testing all time points for visible and imagined stimuli. C) Peak decoding times for training and testing processes. Left plot shows peak decoding times per participant, and right plot shows distribution of peak times after bootstrapping the group 1000 times. D) Mean decoding accuracy for different training-testing time offsets. Highest decoding for visible stimuli occurred around 0ms offset, indicating processes occurred at the same time points for the pattern estimator and visible stimuli on the tracking task. In comparison, the highest decoding for imagined stimuli occurred earlier in the test set than the training set.

Encoding analyses

We used forward encoding to assess the spatial representations in the neural signal for visible and imagined stimuli. Encoding models were trained on EEG data from the pattern estimator sequences and applied to the tracking task separately for each condition and time point. This exploratory analysis resulted in activations per condition for six encoding channels representing the experimental positions. Figure 7A depicts mean response profiles per condition, which show higher responses for the actual stimulus position (0°) relative to the other positions, indicating that visible and imagined position were encoded according to the same processes as the pattern estimator. Plots of the response profiles at representative time periods show position information emerges over time but appears to shift towards to next stimulus position in the sequence (denoted by 60°) at later time periods (Figure 7B). Fitting a model to the channel responses at each time point resulted in two relevant parameters of the spatial coding of the neural signal. The amplitude of the model fit, the peak response amplitude, emerged over time and was reliably above zero for both visible and imagined stimuli, although it emerged slightly later for imagined stimuli (Figure 7C, left). This plot resembled position decoding over time. Importantly, amplitude was reliably higher for visible than imagined position from approximately 100-276ms, indicating that the neural representations of position are stronger for physical rather than internally generated stimuli. Modelling of the response profiles also revealed for peak position (i.e., the centre of the model fit) there was some evidence of a shift away from the current stimulus position in the positive direction for both visible and imagined stimuli (Figure 7C, right). This positive shift was evident from about 300ms, suggesting that the position representations at this time were more consistent with the *upcoming* stimulus position. Together, these encoding analyses complement the decoding results by showing that visible and imagined spatial position are encoded using stimulus-driven processes, and that imagined stimuli elicit considerably weaker spatial representations than visible stimuli.

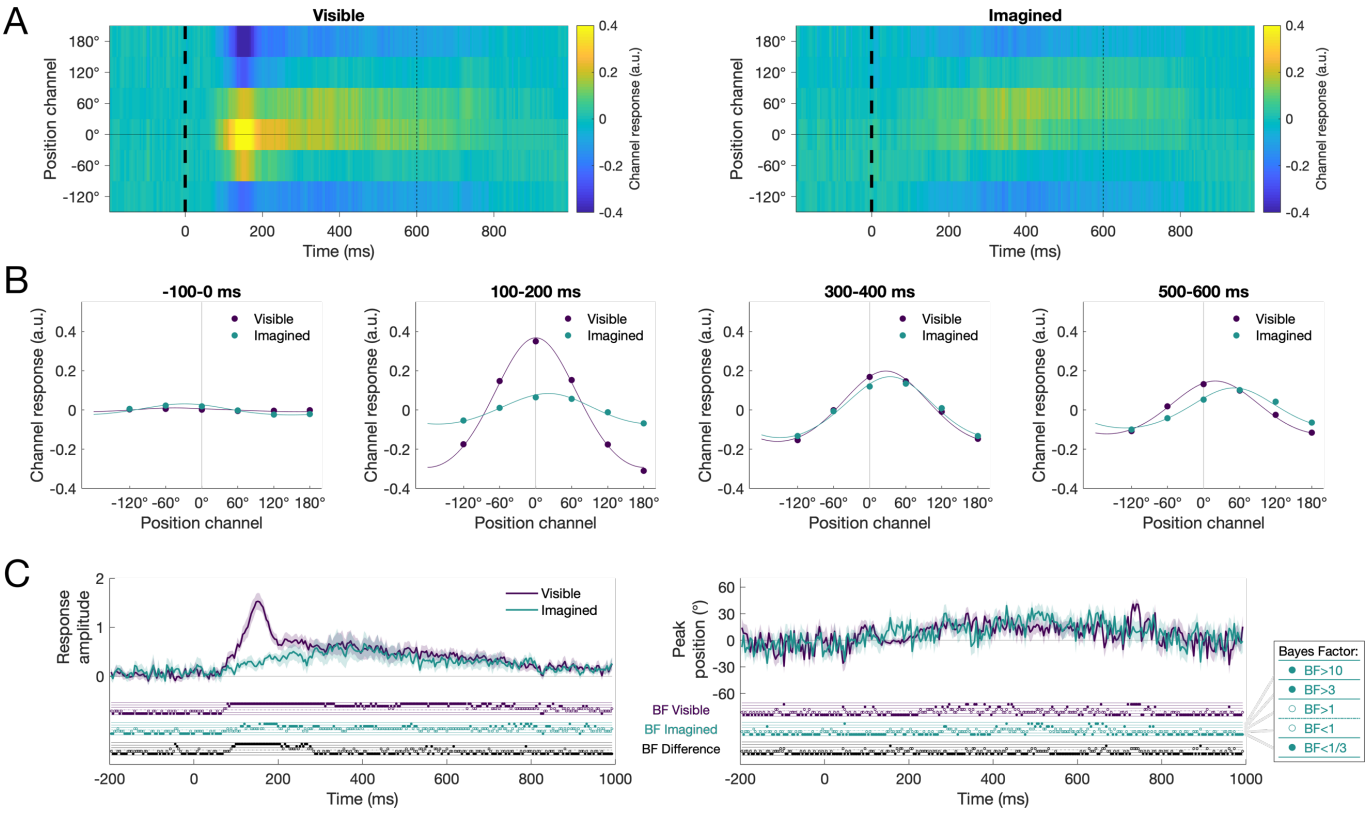

Figure 7. Position response profiles for visible and imagined stimuli using encoding models trained on the random pattern estimator stimuli. A) Activations of each encoding position channel for visible and imagined stimuli, plotted as relationship to the presented stimulus position. For the visible and imagined conditions, there was higher activation for position channels closer to the correct position (0°), indicating that the neural representation of stimulus position was captured by the encoding model. B) Model fitting of channel responses for some representative time periods show the emergence of spatial information over time, with a shift towards the next stimulus position. C) Parameters of the model fits over time. Left: Response amplitude at each time point for visible and imagined stimuli. There was reliable spatial signal for both conditions, but the temporal dynamics varied. Right: Peak position for the model fit generally reflected a peak at 0 degrees for visible and imagined stimuli, although after 250ms there was some evidence of positive shifts towards the upcoming stimulus position.

Discussion

In this study, we assessed the neural underpinnings of internally-generated representations of spatial location. Participants viewed predictable sequences of a moving stimulus and imagined the sequence continuing when the stimulus disappeared. Time-resolved multivariate analyses revealed that patterns of activity associated with visual processing in random sequences were also associated with processing of visible and imagined spatial stimulus positions in the tracking task, but with different temporal dynamics. Specifically, the neural correlates of imagined position (i.e., internally-generated representations) were anticipatory and more temporally diffuse than those of visible position (i.e., sensory-driven representations). Taken together, this study provides evidence that internal representations of spatial

position rely on mechanisms of visual processing, but that these are applied with different temporal dynamics to actual perceptual processes.

The results of this study suggest that similar perceptual and cognitive processes are implemented for processing position of visible and imagined (e.g., occluded) stimuli. This adds to previous neuroimaging work using high level objects by showing that internally-generated spatial representations appear to use the same visual perceptual processes as viewed stimuli (Dijkstra et al., 2018). What neural processes are responsible for this low-level spatial imagery? We found generalisation from the pattern estimator to the visible tracked stimuli began at approximately 76ms, but for imagined stimuli the generalisation did not occur until 120ms (Figure 6A). This suggests that internal spatial representations do not originate with early retinotopic processes such as that of the first stages of processing in V1, but are implemented by higher order processes potentially via feedback and recurrent processes. Above-chance generalisation for visible and imagined stimuli was maintained until approximately 750ms after the pattern estimator stimulus was presented, indicating that position-specific information represented throughout the visual hierarchy has some similarity for stimulus-driven and internally generated representations. It is important to note, however, that the time generalisation results did not show evidence of distinct, progressive stages of processing for the imagined representations. In contrast, the visible stimuli showed different clusters of above-chance decoding on the diagonal of the time-generalisation results, indicating that there were distinct stages of processing. These results are similar to those observed in Dijkstra et al., (2018) during imagery of faces and houses. Recent work has suggested that imagery involves a flow of information from higher- to lower-level brain regions in succession through the ventral stream (Breedlove et al., 2020). We did not find any evidence of this reversal of perceptual processes in imagery; rather, our results suggest that internal representations activate different perceptual stages simultaneously. However, new analysis methods might yield more insight into the information flow through different brain regions during imagery (Dijkstra et al., 2019).

For both visible and imagined stimuli, more distant stimulus positions could more easily be discriminated by the EEG signals. Decoding for neighbouring positions (60° apart) was generally much lower than decoding for positions that were further apart. This is consistent with the retinotopic organization of visual cortices (Tootell et al., 1998), where closer areas of space are represented in neighbouring regions of cortex, leading to more similar spatial patterns of activation that are measured on the scalp with EEG (Carlson et al., 2011). Time generalisation results also showed that neural patterns of activity from the pattern estimator sequences generalised to neighbouring positions in the visible condition, highlighting

the neural similarity for close spatial representations (see ~750ms in Figure 6B). Interestingly, however, decoding for the closest positions was particularly low for the imagined stimuli, indicating that internally generated representations of position are more spatially diffuse than perceptual representations. Multivariate encoding analyses verified that the neural representations of spatial position for both visible and imagined stimuli were encoded using stimulus-driven processes (modelled using data from the pattern estimator sequences), but that imagined representations of position were weaker than stimulus-driven representations from 100-276 ms. Weaker spatial signal in imagery is consistent with representations originating in higher-level regions of the visual hierarchy, which have larger receptive fields (Breedlove et al., 2020). Together, these results suggest that there are common, retinotopic mechanisms for processing position of both visible and imagined stimuli, but with important differences in the origin of the representations leading to much greater precision for visible stimuli.

A cognitive process that might contribute to the extracted position-specific signal in the current study is that of spatial attention. In our experimental task, participants were explicitly asked to track the position of the stimulus, and they performed well, suggesting they were directing their attention to the location of the stimulus. Spatial attention influences the amplitude of early EEG responses (for review, see Mangun, 1995), and MEG classification work has shown that spatial attention enhances object decoding at early stages of processing (Goddard et al., 2019). Top-down spatial attention also results in more diffuse spatial representations than stimulus processing (Intriligator and Cavanagh, 2001). Our decoding and encoding results were obtained from training on the pattern estimator, so our results are focused on processes common to the pattern estimator and the tracking task. In the pattern estimator, there was no explicit task and therefore no incentive to specifically attend to stimulus position. However, there was only one stimulus presented at a time and the saliency of the onsets were likely to attract attention, albeit in a different fashion to the cued positions in the experimental tracking task. As such, the pattern estimator and tracking task had different spatial attention demands, but that does not rule out spatial attention as a source of overlap between the two types of sequences. It is difficult to untangle perceptual and attentional mechanisms during imagery, and it is possible that internal spatial representations rely on processes that are common to perception and attention. The current results are consistent with previous work on imagery using paradigms that are unlikely to rely on spatial attention (Dijkstra et al., 2018; Xie et al., 2020), so it seems likely that perceptual mechanisms are at least a considerable source of overlap for neural patterns on the random pattern estimator sequences and the imagined positions on the tracking task. Future work could attempt to disentangle the role of perceptual and attentional

processes in spatial imagery with a manipulation to reduce attention during the pattern estimator or using valid and invalid cues for spatial position.

To investigate the neural processes underlying spatial imagery, this study focused on spatial representations that were common to two different types of sequences: the pattern estimator and the tracking task. All analyses were performed by obtaining patterns of neural activity associated with spatial position from the randomly ordered pattern estimator stimuli and assessing how these patterns are similar to the position representations of visible and imagined stimuli during the tracking task. Training encoding and decoding models on an independent task allowed us to draw conclusions about the nature of the position representations during tracking without the confound of prediction. Our results show that spatial imagery implements similar neural processes as viewing stimuli. However, it is possible that spatial imagery also contains *different* information to the pattern estimator. As an exploratory analysis, we investigated the temporal dynamics of position representations by decoding within condition (visible and imagined) on the tracking task (see Supplementary Material S4). This cross-validated decoding is somewhat problematic due to the predictable nature of the sequence, so decoding is above chance throughout the whole time period. However, the dynamics are still informative; specifically, within-condition decoding revealed very similar dynamics to the original decoding analysis. Within-visible decoding had a peak at 150ms, resembling the time-resolved analyses from training the pattern estimator and testing the visible stimuli on the same time points (as in Figure 4A). Within-imagined decoding was highest around 0ms, resembling the time-generalisation results from training on mid- and high-level processes of the pattern estimator (e.g., Figure 6B). These results suggest that neural processes as measured in the pattern estimator do capture most of the relevant neural processes implemented during stimulus tracking for visible and imagined stimuli.

Spatial imagery representations were evident using multivariate decoding and encoding analyses, but the magnitudes of the effects were very small. One likely contributing factor to the small effects is the temporal jitter in the neural representations evoked by imagery both within and across participants. Time-locked analyses assess reliable patterns of neural activity occurring at the exact same time across trials. Imagery, as an internally-generated process, is likely to be much more temporally variable than perception, resulting in smaller, more diffuse time-locked neural signals. There is also the likelihood that different participants will use different strategies, resulting in variation from participant to participant. Temporal variability is a challenge in all research involving mental imagery. Indeed, in a face versus house imagery experiment, Dijkstra and colleagues (2018) found <60% accuracy for imagined decoding

compared with nearly 90% for viewed stimuli. To minimise temporal variation in imagery in the current experiment, we used tones to guide participants in the timing of the task. Participants had to covertly track stimulus position on thousands of trials (>1400 visible + >1400 imagined per participant; see Supplementary Material S1 for details), and this large number of trials ensured that we had the power to capture the neural processes associated with spatial imagery representations, despite the temporal variation. Decoding accuracy was low, but accuracy is not an effect size (Hebart and Baker, 2017). Our analyses show there were reliable spatial imagery representations that shared neural patterns with stimulus-driven representations. Importantly, temporal jitter for imagery cannot explain the observed temporal dynamics for processing of imagined position (as seen in the time generalisation plots in Figure 6), because jitter would predict only the x-axis of the time-generalisation plots being smeared relative to the visible condition. The observed imagery results appear to be diffuse in terms of the contributions of the pattern estimator (training; y-axis) processes, reflecting processing occurring at different times in visible and imagined parts of the task.

One factor that we tried to control in this study was eye movements. Recent work has shown that even when participants were instructed to maintain central fixation, the spatial position of a peripheral stimulus could be decoded from eye movements, and the eye movements appeared to account for variance in the MEG signal from 200ms after the stimulus was presented (Quax et al., 2019). To reduce the likelihood of eye movements influencing our spatial representation results, one countermeasure we implemented was using independent sequences of randomly ordered visible stimuli (pattern estimator sequences) to extract position-specific patterns from the EEG signal and used these to generalise to the tracking task. Thus, only neural signals in common between the pattern estimator and the tracking task could result in above chance decoding. The position sequences in the pattern estimator (training set) were randomised, so any incidental eye movements were unlikely to consistently vary with position. The tracking task implemented both clockwise and counter-clockwise sequences, so if there were eye movements, across the whole experiment a given position would have two completely different eye movement patterns. Above-chance cross-decoding from the pattern estimator to the tracking task was therefore unlikely to be driven by eye movements. Second, all stimuli were presented briefly (100ms duration), and for a short 200ms inter-stimulus interval during the pattern estimator. This rapid presentation rate reduced the likelihood that participants would overtly move their eyes, as even the fastest saccades take at least 100ms to initiate (Fischer and Ramsperger, 1984). Third, we excluded participants that appeared to move their eyes excessively during the pattern estimator sequences, which were the sequences used for training the classifier. Finally, we conducted an additional analysis using only

649 posterior electrodes to validate that the neural patterns of activity informative for spatial position were
650 consistent with processes within the visual system (e.g., from occipital cortex). Decoding from posterior
651 electrodes was similar to the whole-brain results. Furthermore, a similar analysis using only frontal
652 electrodes showed later, more diffuse position decoding for visible stimuli, and insufficient evidence for
653 position decoding of imagined stimuli (see Supplementary Material S3), indicating that frontal signal or
654 artefacts did not drive decoding of spatial position for visible or imagined stimuli. Taken together, our
655 finding that spatial position generalised from the pattern estimator to the tracking task from relatively
656 early stages of processing indicates that it was actually a neural representation of spatial location that
657 was driving the classifier rather than any overt eye movements.
658
659 In conclusion, in this study we successfully showed that the position of predictable visible and imagined
660 stimuli can be modelled using patterns of neural activity extracted from independent visible stimuli. Our
661 findings suggest that internally generated spatial representations involve mid- and high-level perceptual
662 processes. The visible stimuli that we used relied on early retinotopic visual processes, yet we found no
663 evidence of generalisation from very early processes (90-120ms) to the imagined stimuli. The stimuli we
664 used were much simpler than the vivid, complex objects used in previous work, but we found similar
665 stages of processing generalised from perceptual to internally-generated representations (Dijkstra et al.,
666 2018), suggesting a general role of mid- and high-level perceptual processing in internally-generated
667 representations such as those implemented during imagery or occlusion. Our finding that neural
668 representations of spatial location were weaker and occurred earlier for imagined objects than for the
669 unpredictable objects indicates an important role of prediction in generating internal representations.
670 Together, our findings suggest that similar neural mechanisms underlie internal representations and
671 stimulus-driven mechanisms, but the timing of these processes is dependent on the predictability of the
672 stimulus.
673
674 **Acknowledgements**
675 This research was supported by an Australian Research Council Discovery Early Career Research Award
676 (DE200101159) to A.K.R. and Australian Research Council Discovery Projects (DP160101300 and
677 DP200101787) to T.A.C. The authors acknowledge the Sydney Informatics Hub and the University of
678 Sydney's high performance computing cluster Artemis for providing the high performance computing
679 resources that contributed to these research results. We thank Alexander Sulfaro for insightful
680 discussions about the study.
681

682 **Additional information**

683 All data, code, and Supplementary Material can be found on the Open Science Framework:

684 https://osf.io/8v47t/.

References

Albers, A.M., Kok, P., Toni, I., Dijkerman, H.C., de Lange, F.P., 2013. Shared representations for working memory and mental imagery in early visual cortex. Curr. Biol. 23, 1427–1431. https://doi.org/10.1016/j.cub.2013.05.065

Blom, T., Feuerriegel, D., Johnson, P., Bode, S., Hogendoorn, H., 2020. Predictions drive neural representations of visual events ahead of incoming sensory information. Proc. Natl. Acad. Sci. 117, 7510–7515. https://doi.org/10.1073/pnas.1917777117

Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10, 433–436.

Breedlove, J.L., St-Yves, G., Olman, C.A., Naselaris, T., 2020. Generative Feedback Explains Distinct Brain Activity Codes for Seen and Mental Images. Curr. Biol. https://doi.org/10.1016/j.cub.2020.04.014

Brouwer, G.J., Heeger, D.J., 2009. Decoding and Reconstructing Color from Responses in Human Visual Cortex. J. Neurosci. 29, 13992–14003. https://doi.org/10.1523/JNEUROSCI.3577-09.2009

Carlson, T.A., Hogendoorn, H., Kanai, R., Mesik, J., Turret, J., 2011. High temporal resolution decoding of object position and category. J. Vis. 11, 9–9. https://doi.org/10.1167/11.10.9

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. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dentico, D., Cheung, B.L., Chang, J.-Y., Guokas, J., Boly, M., Tononi, G., Van Veen, B., 2014. Reversal of cortical information flow during visual imagery as compared to visual perception. NeuroImage 100, 237–243. https://doi.org/10.1016/j.neuroimage.2014.05.081

Di Russo, F., Martinez, A., Hillyard, S.A., 2003. Source analysis of event-related cortical activity during visuo-spatial attention. Cereb. Cortex 13, 486–499.

Dienes, Z., 2011. Bayesian Versus Orthodox Statistics: Which Side Are You On? Perspect. Psychol. Sci. 6, 274–290. https://doi.org/10.1177/1745691611406920

Dijkstra, N., Ambrogioni, L., Gerven, M.A.J. van, 2019. Neural dynamics of perceptual inference and its reversal during imagery. bioRxiv 781294. https://doi.org/10.1101/781294

Dijkstra, N., Mostert, P., de Lange, F.P., Bosch, S., van Gerven, M.A.J., 2018. Differential temporal dynamics during visual imagery and perception. eLife 7, 1–16. https://doi.org/10.7554/eLife.33904.001

Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M.A.J., Friston, K., 2017. Distinct Top-down and Bottom-up Brain Connectivity During Visual Perception and Imagery. Sci. Rep. 7. https://doi.org/10.1038/s41598-017-05888-8

Fischer, B., Ramsperger, E., 1984. Human express saccades: extremely short reaction times of goal directed eye movements. Exp. Brain Res. 57. https://doi.org/10.1007/BF00231145

720 Foster, J.J., Sutterer, D.W., Serences, J.T., Vogel, E.K., Awh, E., 2016. The topography of alpha-band
721 activity tracks the content of spatial working memory. J. Neurophysiol. 115, 168–177.
722 https://doi.org/10.1152/jn.00860.2015
723 Garcia, J.O., Srinivasan, R., Serences, J.T., 2013. Near-real-time feature-selective modulations in human
724 cortex. Curr. Biol. CB 23, 515–522. https://doi.org/10.1016/j.cub.2013.02.013
725 Goddard, E., Carlson, T.A., Woolgar, A., 2019. Spatial and feature-selective attention have qualitatively
726 different effects on population-level tuning. bioRxiv 530352. https://doi.org/10.1101/530352
727 Grootswagers, T., Robinson, A.K., Carlson, T.A., 2019. The representational dynamics of visual objects
728 in rapid serial visual processing streams. NeuroImage 188, 668–679.
729 https://doi.org/10.1016/j.neuroimage.2018.12.046
730 Grootswagers, T., Wardle, S.G., Carlson, T.A., 2017. Decoding Dynamic Brain Patterns from Evoked
731 Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging
732 Data. J. Cogn. Neurosci. 29, 677–697. https://doi.org/10.1162/jocn_a_01068
733 Hagler, D.J., Halgren, E., Martinez, A., Huang, M., Hillyard, S.A., Dale, A.M., 2009. Source estimates
734 for MEG/EEG visual evoked responses constrained by multiple, retinotopically-mapped stimulus
735 locations. Hum. Brain Mapp. 30, 1290–1309. https://doi.org/10.1002/hbm.20597
736 Hebart, M.N., Baker, C.I., 2017. Deconstructing multivariate decoding for the study of brain function.
737 NeuroImage 1–15. https://doi.org/10.1016/j.neuroimage.2017.08.005
738 Hogendoorn, H., Burkitt, A.N., 2018. Predictive coding of visual object position ahead of moving objects
739 revealed by time-resolved EEG decoding. NeuroImage 171, 55–61.
740 https://doi.org/10.1016/j.neuroimage.2017.12.063
741 Intriligator, J., Cavanagh, P., 2001. The spatial resolution of visual attention. Cognit. Psychol. 43, 171–
742 216. https://doi.org/10.1006/cogp.2001.0755
743 Ishai, A., 2002. Visual Imagery of Famous Faces: Effects of Memory and Attention Revealed by fMRI.
744 NeuroImage 17, 1729–1741. https://doi.org/10.1006/nimg.2002.1330
745 Jeffreys, H., 1961. Theory of probability. Oxford University Press.
746 Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.
747 King, J.R., Dehaene, S., 2014. Characterizing the dynamics of mental representations: The temporal
748 generalization method. Trends Cogn. Sci. 18, 203–210. https://doi.org/10.1016/j.tics.2014.01.002
749 Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., others, 2007. What's new in
750 Psychtoolbox-3. Perception 36, 1.
751 Kosslyn, S.M., Alpert, N.M., Thompson, W.L., Maljkovic, V., Weise, S.B., Chabris, C.F., Hamilton,
752 S.E., Rauch, S.L., Buonanno, F.S., 1993. Visual Mental Imagery Activates Topographically
753 Organized Visual Cortex: PET Investigations. J. Cogn. Neurosci. 5, 263–287.
754 https://doi.org/10.1162/jocn.1993.5.3.263

Kwon, O.-S., Tadin, D., Knill, D.C., 2015. Unifying account of visual motion and position perception. Proc. Natl. Acad. Sci. 112, 8142–8147. https://doi.org/10.1073/pnas.1500361112

Le Bihan, D., Turner, R., Zeffiro, T.A., Cuénod, C.A., Jezzard, P., Bonnerot, V., 1993. Activation of human primary visual cortex during visual recall: a magnetic resonance imaging study. Proc. Natl. Acad. Sci. U. S. A. 90, 11802–11805.

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

Lins, O.G., Picton, T.W., Berg, P., Scherg, M., 1993. Ocular artifacts in EEG and event-related potentials I: Scalp topography. Brain Topogr. 6, 51–63. https://doi.org/10.1007/BF01234127

Luck, S.J., 2005. An Introduction to the Event-Related Potential Technique. The MIT Press, Cambridge, Massachusetts.

Mangun, G.R., 1995. Neural mechanisms of visual selective attention. Psychophysiology 32, 4–18. https://doi.org/10.1111/j.1469-8986.1995.tb03400.x

Mechelli, A., 2004. Where Bottom-up Meets Top-down: Neuronal Interactions during Perception and Imagery. Cereb. Cortex 14, 1256–1265. https://doi.org/10.1093/cercor/bhh087

Nanay, B., 2010. Perception and imagination: amodal perception as mental imagery. Philos. Stud. 150, 239–254. https://doi.org/10.1007/s11098-009-9407-5

Oostenveld, R., Praamstra, P., 2001. The five percent electrode system for high-resolution EEG and ERP measurements. Clin. Neurophysiol. 112, 713–719. https://doi.org/10.1016/S1388-2457(00)00527-7

Oosterhof, N.N., Connolly, A.C., Haxby, J.V., 2016. CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. Front. Neuroinformatics 10. https://doi.org/10.3389/fninf.2016.00027

Pearson, J., 2019. The human imagination: the cognitive neuroscience of visual mental imagery. Nat. Rev. Neurosci. https://doi.org/10.1038/s41583-019-0202-9

Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spat. Vis. 10, 437–442.

Posner, M.I., 1980. Orienting of attention. Q. J. Exp. Psychol. 32, 3–25. https://doi.org/10.1080/00335558008248231

Quax, S.C., Dijkstra, N., van Staveren, M.J., Bosch, S.E., van Gerven, M.A.J., 2019. Eye movements explain decodability during perception and cued attention in MEG. NeuroImage 195, 444–453. https://doi.org/10.1016/j.neuroimage.2019.03.069

Reddy, L., Tsuchiya, N., Serre, T., 2010. Reading the mind's eye: Decoding category information during mental imagery. NeuroImage 50, 818–825. https://doi.org/10.1016/j.neuroimage.2009.11.084

789 Robinson, A.K., Grootswagers, T., Carlson, T.A., 2019. The influence of image masking on object
790 representations during rapid serial visual presentation. bioRxiv 515619.
791 https://doi.org/10.1101/515619
792 Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., Iverson, G., 2009. Bayesian t tests for accepting and
793 rejecting the null hypothesis. Psychon. Bull. Rev. 16, 225–237.
794 Scolari, M., Byers, A., Serences, J.T., 2012. Optimal deployment of attentional gain during fine
795 discriminations. J. Neurosci. Off. J. Soc. Neurosci. 32, 7723–7733.
796 https://doi.org/10.1523/JNEUROSCI.5558-11.2012
797 Smout, C.A., Tang, M.F., Garrido, M.I., Mattingley, J.B., 2019. Attention promotes the neural encoding
798 of prediction errors. PLOS Biol. 17, e2006812. https://doi.org/10.1371/journal.pbio.2006812
799 Sprague, T.C., Serences, J.T., 2015. Using Human Neuroimaging to Examine Top-down Modulation of
800 Visual Perception, in: Forstmann, B.U., Wagenmakers, E.-J. (Eds.), An Introduction to Model-
801 Based Cognitive Neuroscience. Springer, New York, NY, pp. 245–274.
802 https://doi.org/10.1007/978-1-4939-2236-9_12
803 Sprague, T.C., Serences, J.T., 2013. Attention modulates spatial priority maps in the human occipital,
804 parietal and frontal cortices. Nat. Neurosci. 16, 1879–1887. https://doi.org/10.1038/nn.3574
805 Tang, M.F., Ford, L., Arabzadeh, E., Enns, J.T., Visser, T.A.W., Mattingley, J.B., 2020. Neural dynamics
806 of the attentional blink revealed by encoding orientation selectivity during rapid visual
807 presentation. Nat. Commun. 11, 434. https://doi.org/10.1038/s41467-019-14107-z
808 Tang, M.F., Smout, C.A., Arabzadeh, E., Mattingley, J.B., 2018. Prediction Error and Repetition
809 Suppression Have Distinct Effects on Neural Representations of Visual Information. bioRxiv
810 213710–213710. https://doi.org/10.1101/213710
811 Tootell, R.B., Hadjikhani, N.K., Vanduffel, W., Liu, a K., Mendola, J.D., Sereno, M.I., Dale, a M.,
812 1998. Functional analysis of primary visual cortex (V1) in humans. Proc. Natl. Acad. Sci. U. S. A.
813 95, 811–817. https://doi.org/10.1073/pnas.95.3.811
814 Wagenmakers, E.-J., 2007. A practical solution to the pervasive problems of p values. Psychon. Bull.
815 Rev. 14, 779–804. https://doi.org/10.3758/BF03194105
816 Wetzels, R., Matzke, D., Lee, M.D., Rouder, J.N., Iverson, G.J., Wagenmakers, E.-J., 2011. Statistical
817 Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests. Perspect.
818 Psychol. Sci. 6, 291–298. https://doi.org/10.1177/1745691611406923
819 Wetzels, R., Wagenmakers, E.-J., 2012. A default Bayesian hypothesis test for correlations and partial
820 correlations. Psychon. Bull. Rev. 19, 1057–1064. https://doi.org/10.3758/s13423-012-0295-x
821 Worden, M.S., Foxe, J.J., Wang, N., Simpson, G.V., 2000. Anticipatory biasing of visuospatial attention
822 indexed by retinotopically specific alpha-band electroencephalography increases over occipital
823 cortex. J. Neurosci. Off. J. Soc. Neurosci. 20, RC63.

824 Xie, S., Kaiser, D., Cichy, R.M., 2020. Visual Imagery and Perception Share Neural Representations in
825 the Alpha Frequency Band. Curr. Biol. 0. https://doi.org/10.1016/j.cub.2020.04.074
826 Zellner, A., Siow, A., 1980. Posterior odds ratios for selected regression hypotheses, in: Bernardo, J.M.,
827 DeGroot, M.H., Lindley, D.V., Smith, A.F.M. (Eds.), Bayesian Statistics: Proceedings of the First
828 InternationalMeeting. University of Valencia Press, Valencia, pp. 585–603.
829