Now you see it, now you don't: Overlapping neural representations for the position of visible and invisible 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 invisible 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 invisible 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/invisible 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. Monitoring the position of invisible objects thus utilises similar perceptual

processes as processing 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 feedforward 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. Such internally-generated representations have been studied with tasks involving imagery, mental rotation, and perception of occluded objects. It is clear that internally-generated representations rely on similar brain regions to stimulus-driven perceptual 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). 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. Together, these results suggest that imagery originates in higher-level brain regions rather than involving feed-forward visual processes from V1.

One aspect that is likely to affect the top-down generation of internal representations is the ability to predict aspects of the stimulus 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 either long term memory or working memory, which involve different patterns of brain activation (Ishai, 2002), whereas representations in conditions of occlusion often have 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; Hogendoorn & Burkitt (2018) measured EEG from participants who viewed an apparent motion stimulus that was predictable or unpredictable in its motion trajectory. Position-specific representations 80-90ms after stimulus onset were unaffected by the predictability of the motion, but a later stage of processing (typically 140-150ms after a stimulus is presented) was pre-activated for predictable relative to random sequences by approximately 16ms (Hogendoorn & 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. Disentangling how expected stimulus position is represented in the brain, the *internal* spatial representation, 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 invisible objects. Participants covertly tracked the position of a simple moving stimulus and kept tracking its imaginary trajectory after it disappeared. Using invisible objects 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 invisible stimulus presentations. We successfully decoded the position of the stimuli from all phases of the task. Our results show that the visible and invisible stimuli evoked the same neural response patterns, but with very different temporal dynamics. 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 template pattern estimator and the experimental task. The pattern estimator used unpredictable stimulus sequences to obtain position-specific EEG signals that were unlikely to be affected by eye-movements. These were subsequently used to detect position signals in the experimental task.

105 ***Participants***

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

110

111 ***Stimuli and design***

112 While participants maintained fixation in the centre of the monitor, a stimulus appeared in six distinct
113 positions 4 degrees of visual angle from fixation. The stimulus was a black circle with a diameter of 3
114 degrees of visual angle. Six unfilled circles acted as placeholders, marking all possible positions throughout
115 the trial. Every stimulus presentation was accompanied by a 1000 Hz pure tone presented for 100 ms via
116 headphones. All stimuli were presented using Psychtoolbox (Brainard, 1997; Kleiner et al., 2007; Pelli,
117 1997) in MATLAB. In total, there were 8 blocks of trials, each of which contained two template pattern
118 estimator sequences and 36 experimental task sequences.

119

120 *Template pattern estimator*

121 The template pattern estimator sequences were designed to extract stimulus-driven position-specific
122 neural patterns from the EEG signal. Participants viewed 16 pattern estimator sequences (2 per block),
123 each of which consisted of 10 repetitions of the 6 stimulus positions (Figure 1a). The order of stimuli was
124 randomised to ensure that for a given stimulus position, the preceding and following stimuli would not
125 be predictive of that position; for example, comparing the neural patterns evoked by positions 1 and 2
126 could not be contaminated by preceding and following stimuli because they could both be preceded and
127 followed by all six positions. Each stimulus was shown for 100ms and was followed by an inter-stimulus
128 interval of 200ms. Onset of the stimulus was accompanied by a 100ms tone. Participants were instructed
129 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 & 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 median 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.

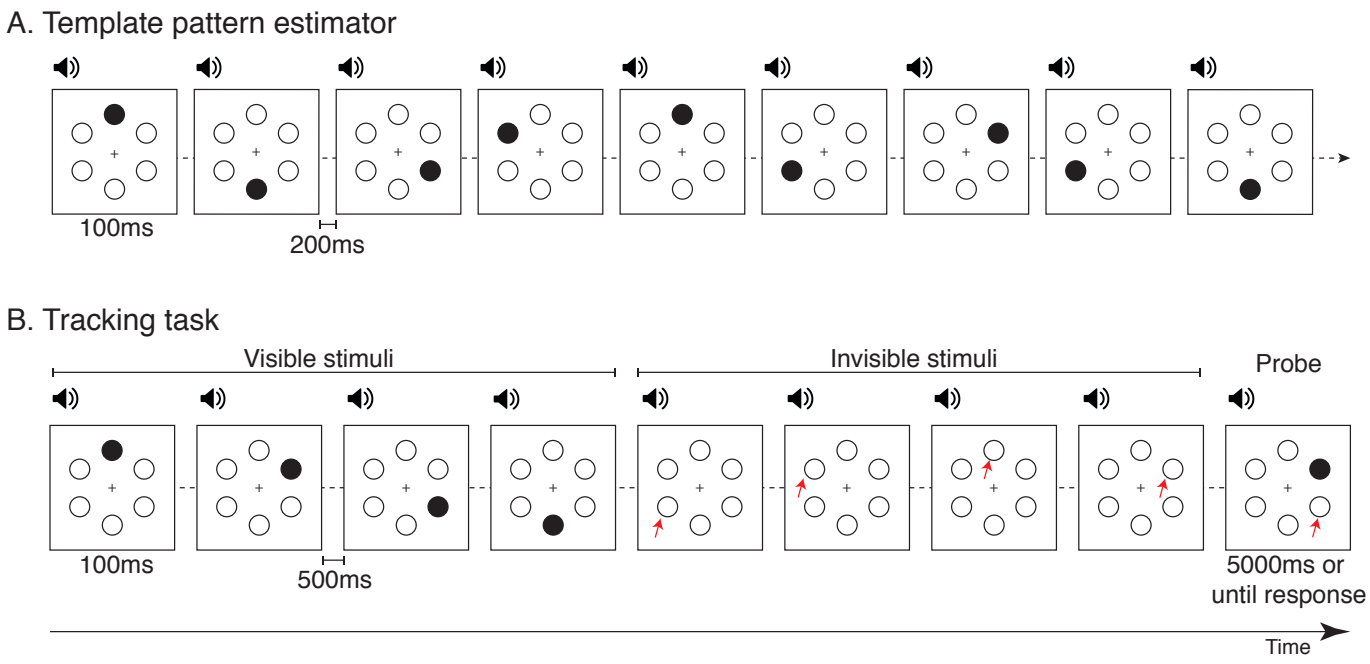

Figure 1. Stimuli and design. a) Template pattern estimator. Participants passively viewed rapid sequences in which a black circle stimulus appeared in six locations in random order. A tone accompanied every stimulus onset. b) Tracking task. The stimulus was presented in different locations in predictable sequences. After 4-6 visible locations, participants had to track the location of the "invisible" 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 invisible stimulus.

Tracking task

For the experimental task, participants viewed sequences consisting of 4-6 visible stimuli and 4-6 "invisible" 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 & Praamstra, 2001) using a BrainVision ActiChamp system, digitized at a 1000-Hz sample rate. Scalp electrodes were referenced to Cz during recording. EEGLAB (Delorme & Makeig, 2004) was used to pre-process the data offline, where data were filtered 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.

Decoding analyses

An MVPA decoding pipeline (Grootswagers et al., 2017) was applied to the EEG epochs to investigate position representations of visible and invisible 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 analyses. A linear discriminant analysis classifier was trained using the template pattern estimator data to distinguish between all pairs of positions. The classifier was trained with balanced numbers of trials per stimulus position from the template pattern estimator sequences. The classifier was then tested separately on the visible and invisible 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 template 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

197 the tracking task, all sequences were included in the decoding analyses regardless of whether the
198 participant correctly classified the position of the probe (i.e., correct and incorrect sequences were
199 analysed). When only correct trials were included, the trends in the results remained the same (see Figure
200 S1, https://osf.io/8v47t/).
201
202 To assess whether neighbouring stimulus positions evoked more similar neural responses, we also
203 calculated decoding accuracy as a function of the distance between position pairs. Each position pair had
204 a radial distance of 60°, 120° or 180° apart. There were six pairs with a distance of 60° (e.g., position 1 vs
205 position 2, position 2 vs position 3), six pairs with a distance of 120° (e.g., position 1 vs position 3, position
206 2 vs position 4), and three pairs with a distance of 180° (directly opposing each other, e.g., position 1 vs
207 position 4, position 2 vs position 5). Decoding accuracy for each pair distance was calculated as the mean
208 of all relevant pair decoding and compared to chance (50%).
209
210 As a final set of analyses, time generalisation (King & Dehaene, 2014) was used to assess whether the
211 patterns of informative neural activity occurred at the same times for the pattern localiser and the visible
212 and invisible stimuli on the tracking task. Classification was performed on all combinations of time points
213 from the pattern estimator epochs and the visible or invisible epochs. The classifier was trained on all
214 trials from the localiser sequences and tested on visible and invisible stimulus positions. To reduce
215 computation time, instead of the 15 pairwise tests conducted for the time-resolved decoding analyses,
216 we performed six-way position decoding for the time generalization analyses, so chance was 16.66%.
217
218 **Statistical inference**
219 To assess the evidence that decoding performance differed from chance, we calculated Bayes factors
220 (Dienes, 2011; Jeffreys, 1961; Kass & Raftery, 1995; Rouder et al., 2009; Wagenmakers, 2007). A JZS prior
221 (Rouder et al., 2009) was used with a scale factor of 0.707 to test the alternative hypothesis of above-

chance decoding (Jeffreys, 1961; Rouder et al., 2009; Wetzels & Wagenmakers, 2012; Zellner & 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 a threshold of BF > 3 as evidence for the alternative hypothesis, and BF < 1/3 as evidence in favour of the null hypothesis (Jeffreys, 1961; Kass & Raftery, 1995; Wetzels et al., 2011). BFs that lie between those values indicate insufficient evidence to favour a hypothesis.

Results

Behavioural results

Participants performed well on the tracking task. Mean accuracy was high for all probe positions (Fig 2a), and 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. Therefore, participants allocated their attention appropriately to the expected position of the stimulus during the invisible portion of the tracking task.

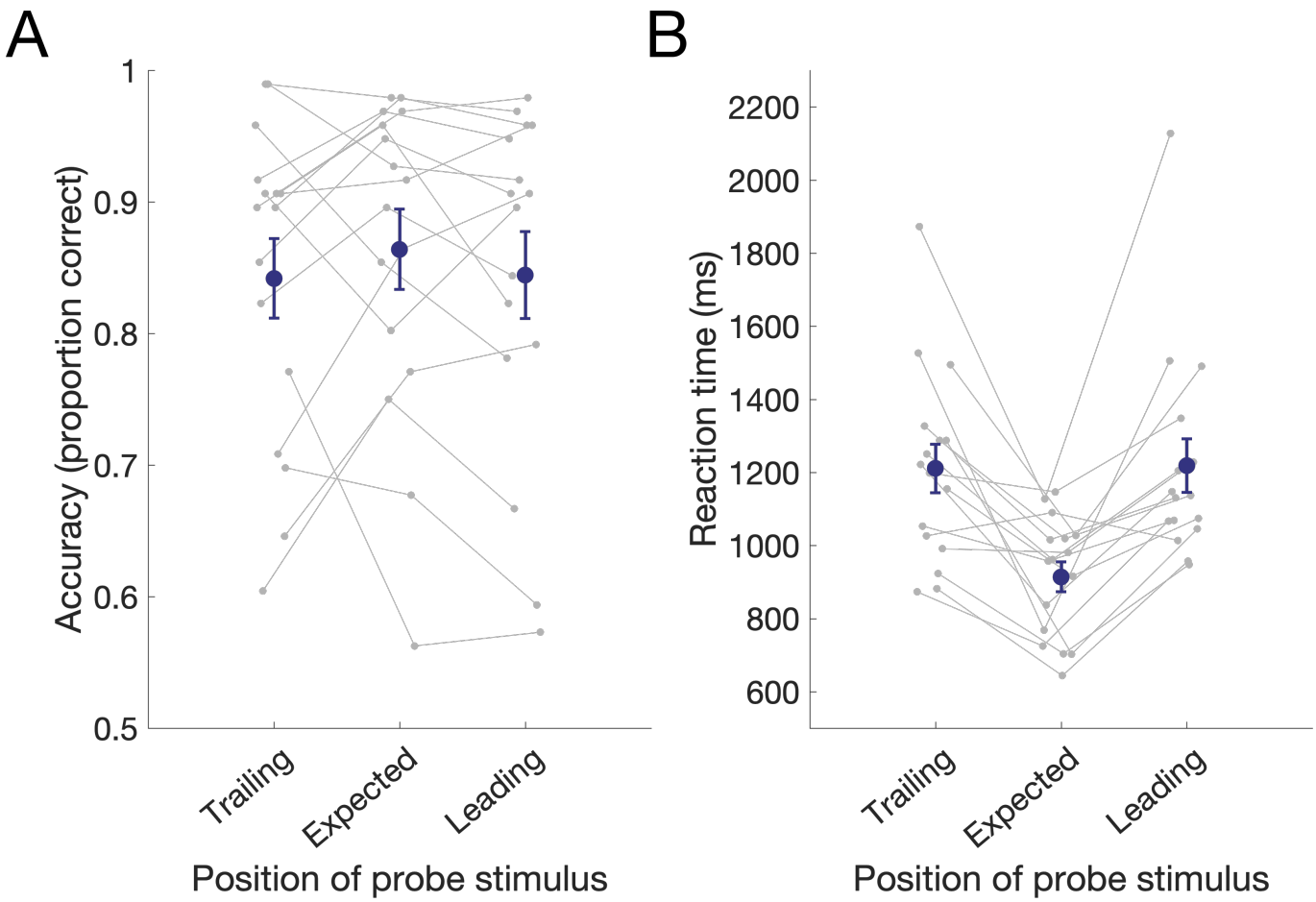

Figure 2. Behavioural results. a) Accuracy, and b) Reaction time on the tracking task as a function of final probe position. Individual subject data are plotted in grey, with group mean in navy. Error bars depict one standard error of the mean across participants (*N* = 16).

Position decoding using the template pattern estimator sequences

The template 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.

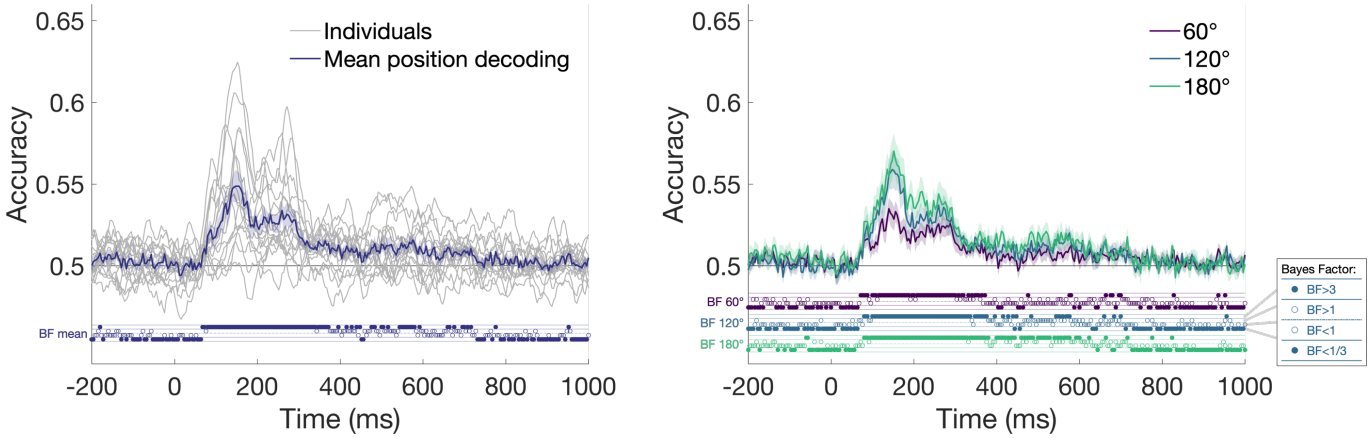

Figure 3. Position decoding using template pattern estimator sequences. Left plot shows group mean decoding and smoothed individual participant decoding for all pairs of positions, and right plot shows mean position decoding as a function of the angular distance between stimulus pairs. Shaded areas show 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 invisible (simulated occluded) stimuli, the classifier was trained on data from the visible template pattern estimator stimuli and tested on data from the tracking task for the visible and invisible stimuli. Crucially, position could be decoded for both visible and invisible 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 invisible stimuli. Here, decoding was not above chance until approximately 152ms and peaked at 176ms (Figure 4b). The above chance cross-decoding from the visible pattern estimator stimuli to the invisible 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 invisible stimuli, unlike for the visible stimuli.

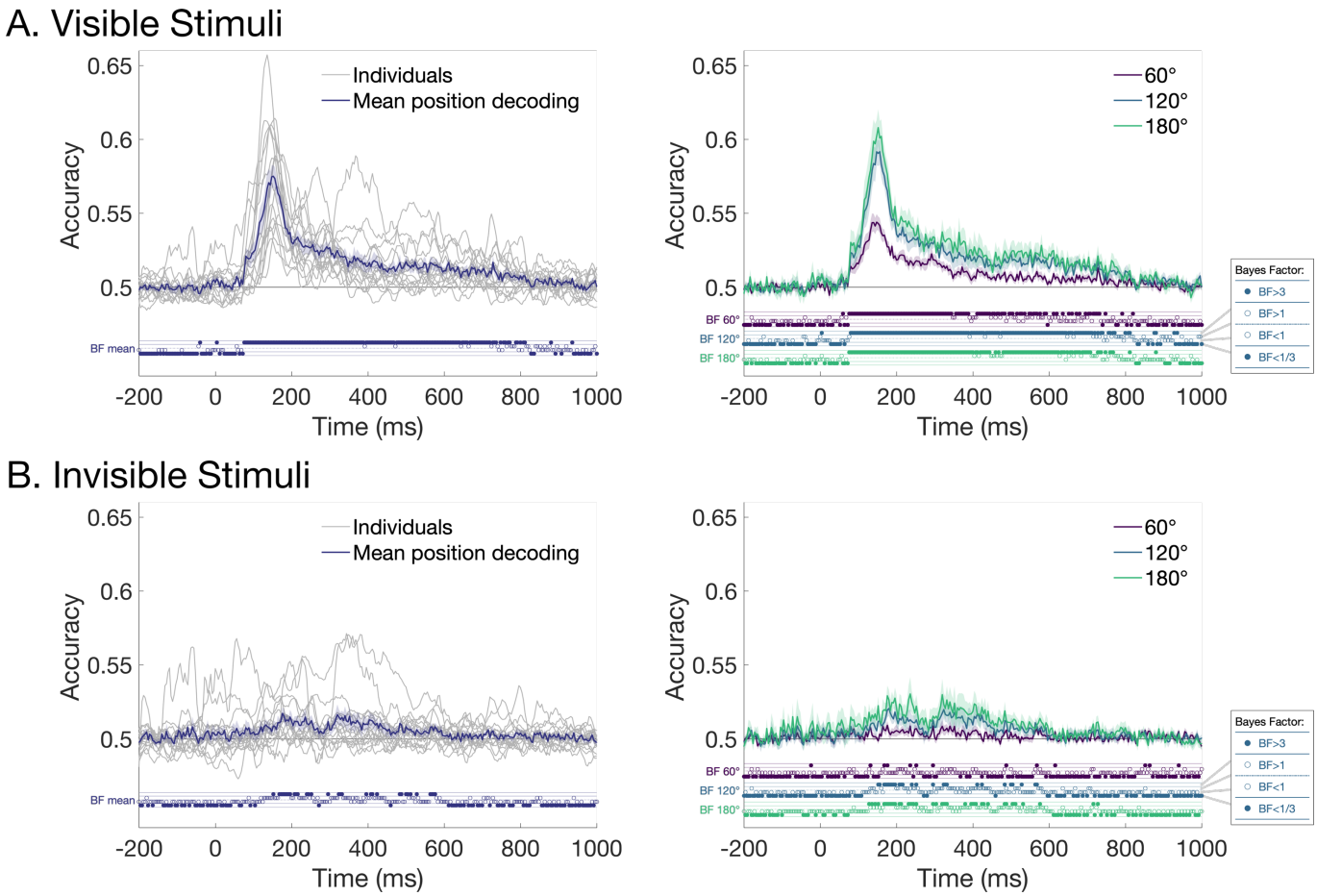

Figure 4. Position decoding from object tracking task. a) Visible stimuli. b) Invisible stimuli. Left plots show group mean decoding and smoothed individual participant decoding for all pairs of positions, and right plots show mean position decoding as a function of the angular distance between position 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 invisible 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). Analyses restricted to frontal electrodes showed later, more diffuse coding for visible stimuli, and little evidence for position coding of invisible stimuli (see Figure S2, https://osf.io/8v47t/). Thus, position-specific neural information for visible and invisible stimuli was evident specifically over posterior regions of the brain, consistent with visual cortex representing stimulus-driven and internal representations of spatial location.

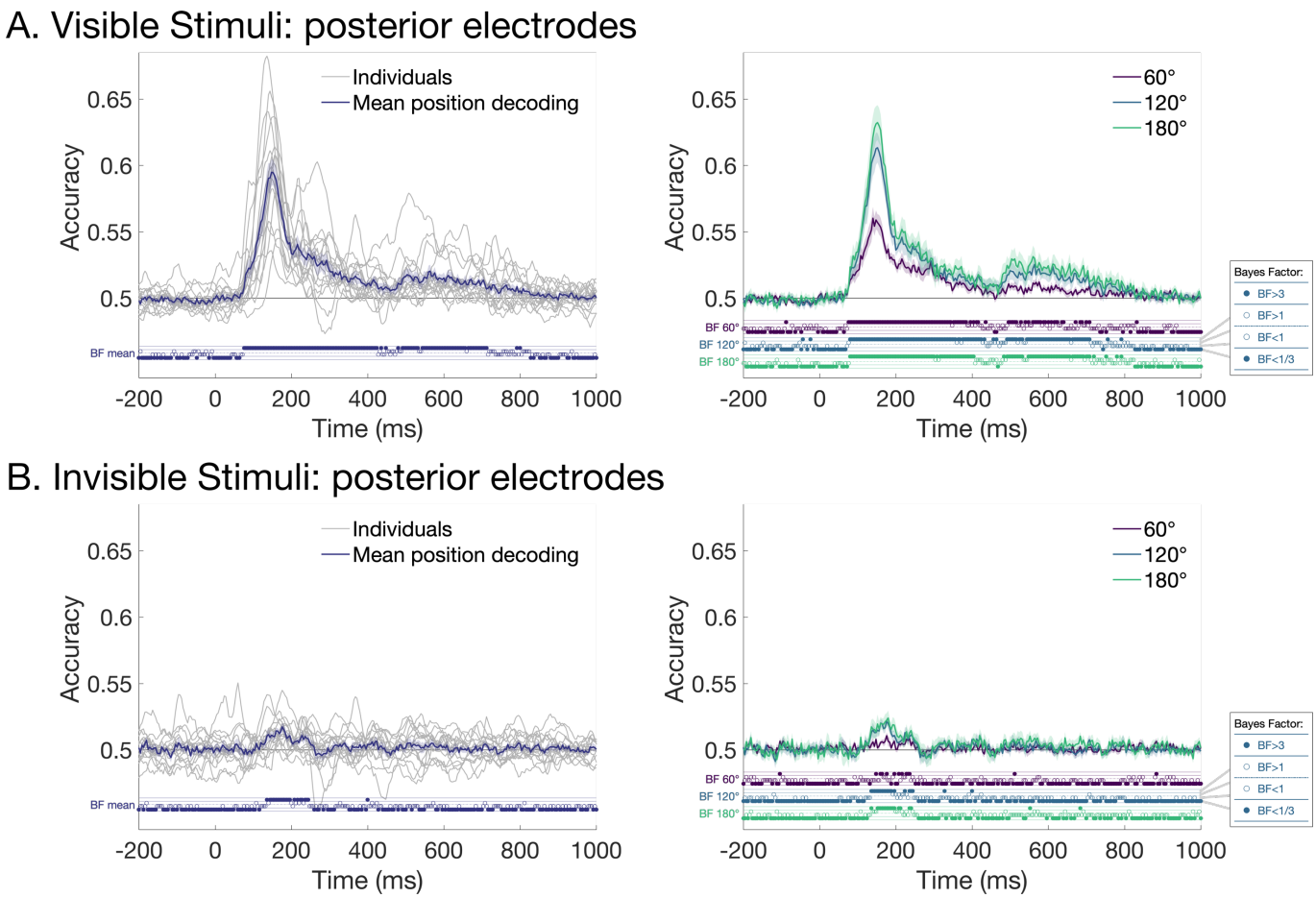

Figure 5. Position decoding from object tracking task using only posterior electrodes. a) Visible stimuli. b) Invisible stimuli. Left plots show group mean decoding and smoothed individual participant decoding for all pairs of positions, and right plots show mean position decoding as a function of the 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 invisible stimuli, but with different temporal dynamics. To assess the possibility that neural processes were more temporally variable for invisible than for visible stimuli, we performed whole brain (63-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 invisible 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), 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. After the tracking stimulus was presented (800-1000ms), 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 700ms on the plot (stim +1 vertical line).

Time generalisation for the invisible 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 template 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 6b). 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. Overall, time generalisation results suggest that during the invisible stimulus portion of the tracking task, which relied on internal representations of position, the neural dynamics were more variable and anticipatory.

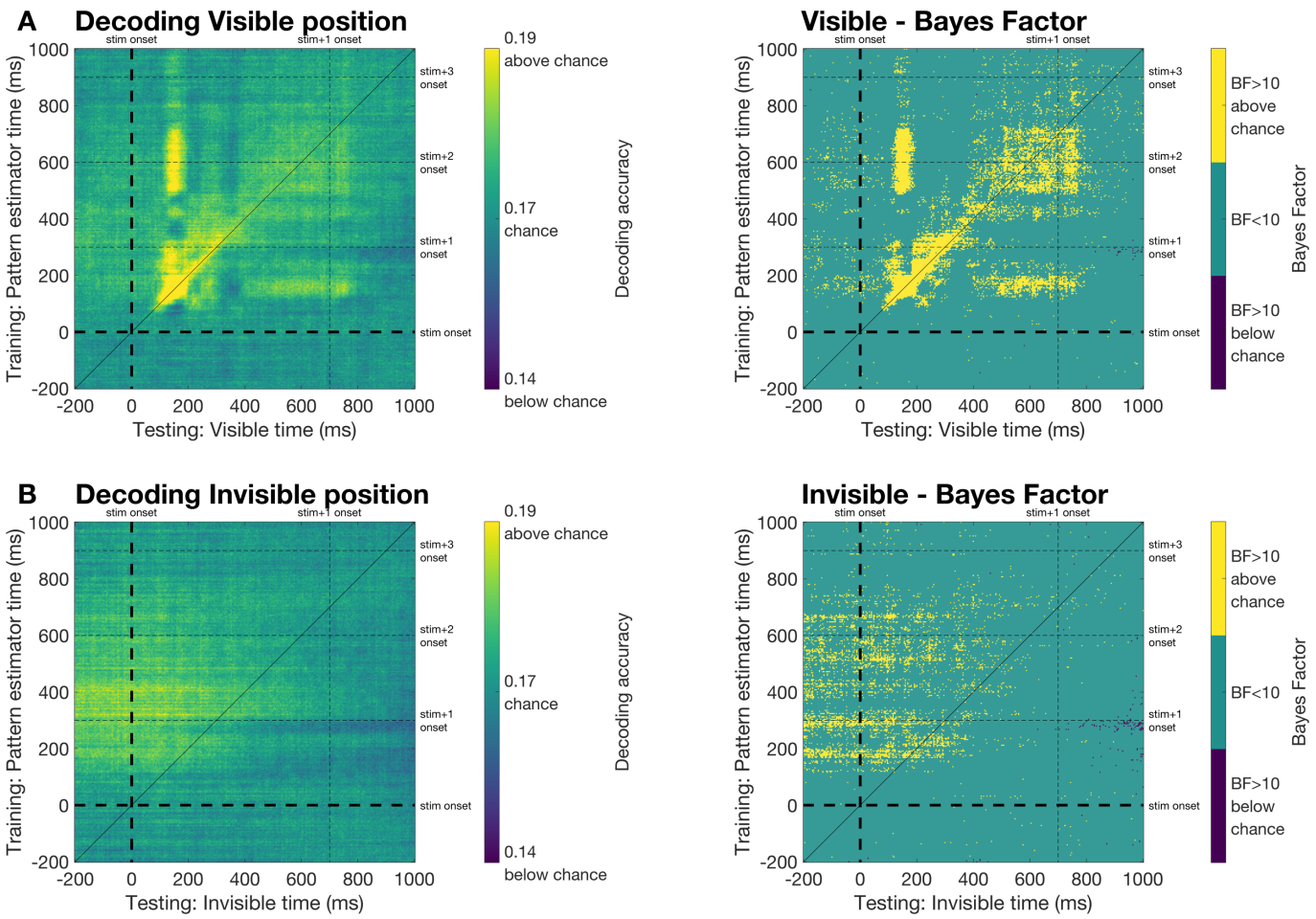

Figure 6. Time generalisation results. a) Decoding visible stimulus position. b) Decoding invisible stimulus position. Decoding was performed by training on data from the template pattern estimator sequences of visible stimuli and testing on the experimental trials for all pairs of time points. Left plots show 6-way decoding accuracy for stimulus position, and right plots show associated Bayes Factors.

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 MVPA revealed that patterns of activity associated with visual processing in random sequences were also associated with processing of visible and invisible spatial stimulus positions in the tracking task, but with different temporal dynamics. Specifically, the neural correlates of invisible position (i.e., internally-generated representations) were anticipatory and more temporally diffuse than those of visible position (i.e., sensory and perceptual

355 representations). Taken together, this study provides evidence that internal representations of spatial
356 position rely on mechanisms of visual processing, but that these are applied with different temporal
357 dynamics to actual perceptual processes.
358
359 The results of this study suggest that similar perceptual processes are implemented for processing
360 position of visible and invisible (e.g., occluded) stimuli. This adds to previous neuroimaging work using
361 high level objects by showing that internally-generated spatial representations appear to use the same
362 visual perceptual processes as viewed stimuli (Dijkstra et al., 2018). What neural processes are responsible
363 for this low-level spatial imagery? We found generalisation from the template pattern estimator to the
364 visible tracked stimuli began at approximately 76ms, but for invisible stimuli the generalisation did not
365 occur until 120ms. This suggests that internal spatial representations do not rely on early retinotopic
366 processes such as that of V1, but are implemented by higher order visual processes. Above-chance
367 generalization for visible and invisible stimuli was maintained until approximately 750ms after the pattern
368 estimator stimulus was presented, indicating that position-specific information represented throughout
369 the visual hierarchy has some similarity for stimulus-driven and internally generated representations. It is
370 important to note, however, that the time generalisation results did not show evidence of distinct,
371 progressive stages of processing for the invisible representations. In contrast, the visible stimuli showed
372 different clusters of above-chance decoding on the diagonal of the time-generalisation results, indicating
373 that there were distinct stages of processing. These results are similar to those observed in Dijkstra et al.,
374 (2018) during imagery of faces and houses. Internal representations thus seem to activate different
375 perceptual processes simultaneously, rather than the representations involving information flow through
376 different brain regions.
377
378 For both visible and imagined stimuli, more distant stimulus positions could more easily be discriminated
379 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 generalization results also showed that neural patterns of activity from the template pattern estimator sequences generalized above chance to neighbouring positions. Interestingly, however, decoding for the closest positions was particularly low for the invisible stimuli, raising the possibility that internally generated representations of position are more spatially diffuse than perceptual representations. Together, increasing decodability of stimulus position with increasing distance between stimuli supports a common, retinotopic mechanism for processing position of both visible and imagined stimuli, but with greater precision for visible stimuli.

Another 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). It is important to note, however, that our classification results were obtained from training on the template pattern estimator, in which there was no explicit task and therefore no incentive to specifically attend to stimulus position. The neural patterns of activity associated with position were therefore more likely to be associated with perceptual rather than attentional mechanisms. A role of spatial attention cannot be ruled out, however. In the pattern estimator there was only one stimulus presented at a time and the onsets were likely to attract attention, albeit in a different fashion to the cued positions in the experimental tracking task. It is possible that a combination of both perceptual and attentional mechanisms is necessary for the generation of internal spatial representations. 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 even make the stimuli invisible.

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 (template 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 template 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 & Ramsperger, 1984). Third, we excluded participants that appeared to move their eyes excessively during the template pattern estimator sequences, which were the sequences used for training the classifier. Finally, we conducted an additional analysis using only posterior electrodes to validate that the neural patterns of activity informative for spatial position were consistent with processes within the visual system (e.g., from occipital cortex). Decoding from posterior electrodes was similar to the whole-brain results. Furthermore, a similar analysis

using only frontal electrodes showed later, more diffuse position decoding for visible stimuli, and insufficient evidence for position decoding of invisible stimuli (see Figure S2, https://osf.io/8v47t/), indicating that frontal signal or artefacts did not drive decoding of spatial position for visible or imagined stimuli. Taken together, our finding that spatial position generalised from the pattern estimator to the tracking task from relatively early stages of processing indicates that it was actually a neural representation of spatial location that was driving the classifier rather than any overt eye movements.

In conclusion, in this study we successfully decoded the position of predictable visible and invisible stimuli using patterns of neural activity extracted from independent visible stimuli. Our findings suggest that internally generated spatial representations involve mid- and high-level perceptual processes. The visible stimuli that we used relied on early retinotopic visual processes, yet we found no evidence of generalisation from very early processes (90-120ms) to the invisible stimuli. The stimuli we used were much simpler than the vivid, complex objects used in previous work, but we found similar stages of processing generalised from perceptual to internally-generated representations (Dijkstra et al., 2018), suggesting a general role of mid- and high-level perceptual processing in internally-generated representations such as those implemented during imagery or occlusion. Our finding that mid- and high-level perceptual processes were spatially diffuse and occurred earlier for invisible objects than for the unpredictable objects indicates an important role of prediction in generating internal representations. Together, our findings suggest that similar neural mechanisms underlie internal representations and visual perception, but the timing of these processes is dependent on the predictability of the stimulus.

References

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Carlson, T. A., Hogendoorn, H., Kanai, R., Mesik, J., & Turret, J. (2011). High temporal resolution decoding of object position and category. *Journal of Vision*, *11*(10), 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. *Journal of Neuroscience Methods*, *134*(1), 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. *Cerebral Cortex*, *13*(5), 486–499.

Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspectives on Psychological Science*, *6*(3), 274–290. https://doi.org/10.1177/1745691611406920

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*(e33904), 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. *Scientific Reports*, *7*(1). 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. *Experimental Brain Research*, *57*(1). https://doi.org/10.1007/BF00231145

Goddard, E., Carlson, T. A., & Woolgar, A. (2019). Spatial and feature-selective attention have qualitatively different effects on population-level tuning. *BioRxiv*, *530352*. https://doi.org/10.1101/530352

Grootswagers, T., Robinson, A. K., & Carlson, T. A. (2019). The representational dynamics of visual objects in rapid serial visual processing streams. *NeuroImage*, *188*, 668–679. https://doi.org/10.1016/j.neuroimage.2018.12.046

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068

Hagler, D. J., Halgren, E., Martinez, A., Huang, M., Hillyard, S. A., & Dale, A. M. (2009). Source estimates for MEG/EEG visual evoked responses constrained by multiple, retinotopically-mapped stimulus locations. *Human Brain Mapping*, *30*(4), 1290–1309. https://doi.org/10.1002/hbm.20597

Hogendoorn, H., & Burkitt, A. N. (2018). Predictive coding of visual object position ahead of moving objects revealed by time-resolved EEG decoding. *NeuroImage*, *171*(December 2017), 55–61. https://doi.org/10.1016/j.neuroimage.2017.12.063

Ishai, A. (2002). Visual Imagery of Famous Faces: Effects of Memory and Attention Revealed by fMRI. *NeuroImage*, *17*(4), 1729–1741. https://doi.org/10.1006/nimg.2002.1330

Jeffreys, H. (1961). *Theory of probability*. Oxford University Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

King, J. R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., & others. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., Hamilton, S. E., Rauch, S. L., & Buonanno, F. S. (1993). Visual Mental Imagery Activates Topographically Organized Visual Cortex: PET Investigations. *Journal of Cognitive Neuroscience*, *5*(3), 263–287. 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. *Proceedings of the National Academy of Sciences*, *112*(26), 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. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(24), 11802–11805.

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

Lins, O. G., Picton, T. W., Berg, P., & Scherg, M. (1993). Ocular artifacts in EEG and event-related potentials I: Scalp topography. *Brain Topography*, *6*(1), 51–63. https://doi.org/10.1007/BF01234127

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. The MIT Press.

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, *32*(1), 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. *Cerebral Cortex*, *14*(11), 1256–1265. https://doi.org/10.1093/cercor/bhh087

Nanay, B. (2010). Perception and imagination: amodal perception as mental imagery. *Philosophical Studies*, *150*(2), 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. *Clinical Neurophysiology*, *112*(4), 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. *Frontiers in Neuroinformatics*, *10*. https://doi.org/10.3389/fninf.2016.00027

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

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

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*(2), 818–825. https://doi.org/10.1016/j.neuroimage.2009.11.084

Robinson, A. K., Grootswagers, T., & Carlson, T. A. (2019). The influence of image masking on object representations during rapid serial visual presentation. *BioRxiv*, 515619. https://doi.org/10.1101/515619

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.

Tootell, R. B., Hadjikhani, N. K., Vanduffel, W., Liu, a K., Mendola, J. D., Sereno, M. I., & Dale, a M. (1998). Functional analysis of primary visual cortex (V1) in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(3), 811–817. https://doi.org/10.1073/pnas.95.3.811

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests. *Perspectives on Psychological Science*, *6*(3), 291–298. https://doi.org/10.1177/1745691611406923

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, *19*(6), 1057–1064. https://doi.org/10.3758/s13423-012-0295-x

Zellner, A., & Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. M. Smith (Eds.), *Bayesian statistics: Proceedings of the First InternationalMeeting* (pp. 585–603). University of Valencia Press.