Neural dynamics of shifting attention between perception and working-memory contents

Daniela Gresch[1,2], Sage E.P. Boettcher[1,2], Freek van Ede[3], Anna C. Nobre[1,2,4,5]

[1] Department of Experimental Psychology, University of Oxford, Oxford, UK

[2] Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, Oxford, UK

[3] Institute for Brain and Behavior Amsterdam, Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, The Netherlands

[4] Wu Tsai Institute, Yale University, New Haven, CT, USA

[5] Department of Psychology, Yale University, New Haven, CT, USA

Corresponding author

Daniela Gresch
Brain & Cognition Lab
Oxford Centre for Human Brain Activity
Department of Experimental Psychology
University of Oxford

Mail: daniela.gresch@psy.ox.ac.uk
ORCID: https://orcid.org/0000-0001-9072-4958

Abstract

In everyday tasks, our focus of attention shifts seamlessly between contents in the sensory environment and internal memory representations. Yet, research has mainly considered external and internal attention in isolation. We used magnetoencephalography to compare the neural dynamics of shifting attention to visual contents within vs. between the external and internal domains. Participants performed a combined perception and working-memory task in which two sequential cues guided attention to upcoming (external) or memorised (internal) sensory information. Critically, the second cue could redirect attention to visual content within the same or alternative domain as the first cue. Multivariate decoding unveiled distinct patterns of human brain activity when shifting attention within vs. between domains. Crossing domains did not alter the timing of spatial orienting, as indexed by posterior alpha lateralisation. These findings provide important initial insights into the neural dynamics that govern attentional shifts between perception and working memory.

Introduction

Attention can be selectively directed towards sensory information within the external environment or internal representations in memory, phenomena referred to as external and internal attention, respectively[1]. During almost every natural behaviour, our attention fluctuates seamlessly between these attentional domains, rapidly transitioning focus between sensory and mnemonic contents. Despite playing an integral part in daily cognition, the mechanisms for shifting attention between contents in perception and memory have been relatively overlooked.

A large body of research has focused on studying isolated shifts of attention within perception[2,3] or within working memory[4–6]. Some investigations have also considered how shifts within each domain compare, emphasising shared neural systems and dynamics as well as highlighting dissociations[7–15]. These studies have typically employed informative pre- and/or retro-cues in their task design to isolate brain activity specifically associated with shifting attention in either domain[4,8–10,14,15].

In contrast, the investigation of between-domain shifts – transitioning attention between perceptual and mnemonic contents – is still at a nascent stage. A limited number of behavioural studies have highlighted higher costs when shifting across different attentional domains compared to shifting within the same domain[16–22]. Brain imaging studies which compared task phases that rely more heavily on sensory or internalised stimuli have highlighted the role of the lateral rostral prefrontal cortex in shifting attention between the external and internal domains[23,24]. Other studies have proposed the hippocampus[25,26] or the fronto-insular cortex[27,28] might serve as the interface between perception and memory. Beyond these initial clues, the dynamic neural mechanisms of cross-domain attention shifts are all to be discovered.

Our aim was to expand upon these initial steps by investigating the brain dynamics associated with shifting attention between the perceptual and working-memory domains and by directly comparing these to the dynamics linked to shifting attention within domains. Building on our recent behavioural study[18], we used a combined perception and working-memory task in which informative cues indicated the location of sensory or mnemonic contents for subsequent report. Successive cues necessitated shifts of attention either within or between perception and working memory. We recorded magnetoencephalography (MEG) to track brain activity with high temporal resolution while participants performed the task and used multivariate pattern analysis to compare activity related to attentional shifts of different types. Based on our and other previous findings of performance costs for shifting attention between relative to within domains[16–22], we hypothesized the engagement of distinct processes when shifting attention between domains.

To preview our results, we demonstrate that dissociable patterns of brain activity accompany shifts between perception and working memory as opposed to shifts within either of these domains. These unique patterns of brain activity invoked by between-domain attention shifts are broadly distributed, temporally dynamic, predict behavioural performance, and occur without interfering with the timing of spatial attention shifts as indexed by the lateralisation of posterior 8-12 Hz alpha activity. In addition, we show distinct patterns of brain activity are associated with external and internal attention states themselves. These findings provide important first clues about the dynamic neural bases that govern shifts of attention between perception and working-memory contents.

Results

Twenty-five healthy human volunteers performed a combined perception and working-memory task (Fig. 1a). Two separate bilateral arrays each containing two randomly oriented bars were presented, with one array occurring before and the other after one or two spatially informative colour cues. That is, two bars appeared in an early, internal working-memory array, encoded before the cues, and two bars in a later, external perceptual array, appearing after the cues. At the end of each trial, participants reproduced the orientation of the bar that was last cued. Cues indicating the location of a previously encoded bar guided attention internally (i.e., retro-cues, internal cues), instructing which item in working memory was relevant for reporting. In contrast, cues indicating a previously unoccupied

location directed attention externally (i.e., pre-cues, external cues), instructing which item in the upcoming perceptual array would be relevant. Half of the trials were no-shift trials in which only a single cue was presented, and participants reported the initially cued item. These randomly intermixed no-shift trials were included only to ensure that participants relied on the initial cue when performing the task. In the remaining trials, participants were tasked with shifting their attention to another item, which could be either an internal or external item, as indicated by a second cue. These double-cue trials were of central interest to us. The second cue always pointed to another item in the opposite hemifield, but important, equally often indicated that attention needed to be reoriented to an item of the same domain (i.e., within-domain shift) or to an item in the other domain (i.e., between-domain shift). This resulted in four possible shift conditions: external-to-external, internal-to-internal, external-to-internal, and internal-to-external. The orthogonal manipulation of the to-be-reported *target domain* (external vs. internal) and the *shift type* (within-domain shift vs. between-domain shift) enabled us to independently investigate these two factors in a systematic manner. The first cue in single-cue trials and the second cue in double-cue trials were 100% instructive and, thus, always indicated the target for report (see Methods for the full task description).

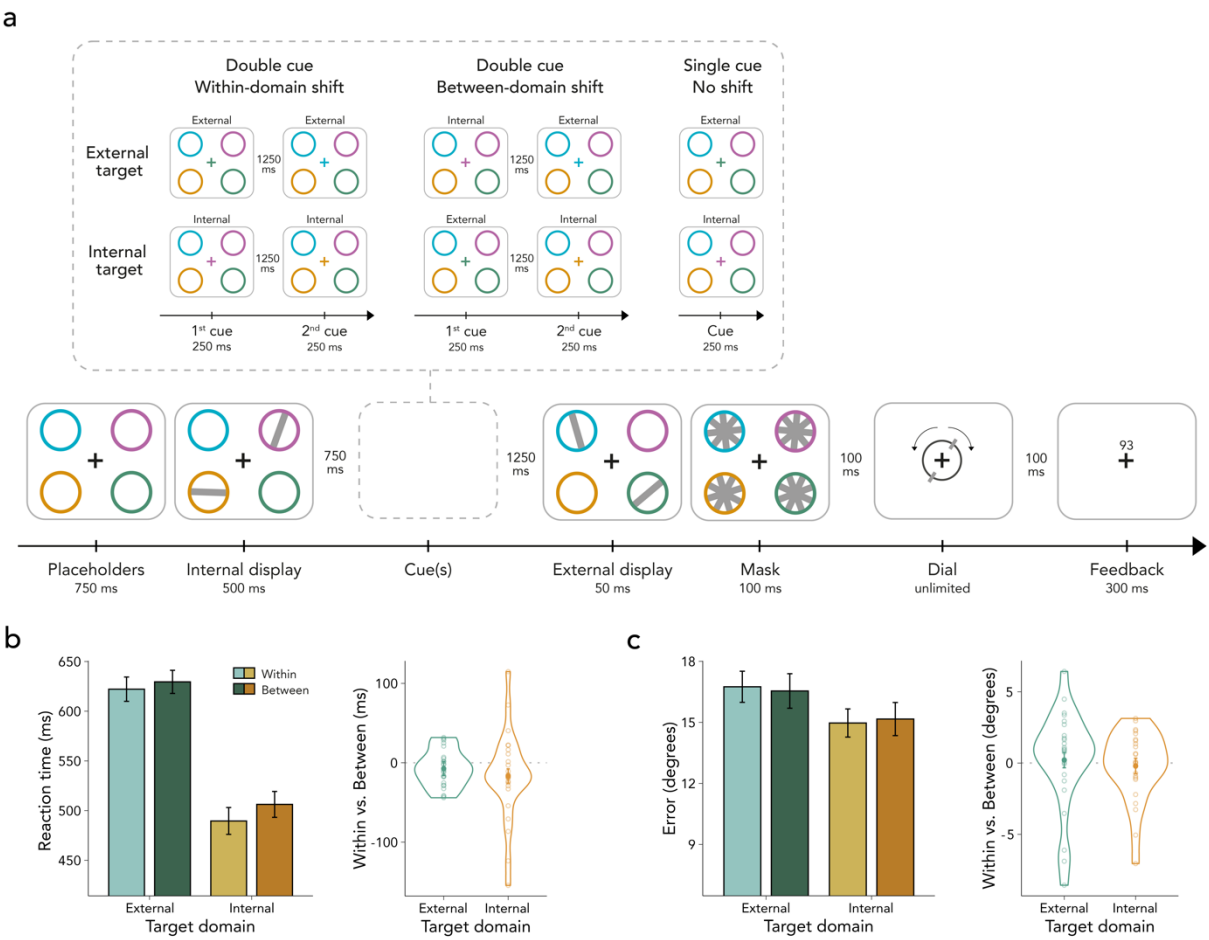


Figure 1. Task schematic and behavioural results. (a) Tilted bars were presented across a working-memory array, encoded before the cues (i.e., internal display), and a perceptual array, appearing after the cues (i.e., external display). In double-cue (i.e., shift) trials, two cues appeared between the presentation of the internal and external display. At the time of the second cue, attention needed to be reoriented either to an item of the same domain (i.e., within-domain shift) or to an item in the other domain (i.e., between-domain shift). To ensure participants relied on the first cue, 50% single-cue trials were included in which only one cue was presented, and thus no attentional reallocation was required. **(b)** Left panel: Average reaction times (RTs) as a function of target domain and shift type. RTs were faster for internal than external targets and for within- than between-domain shifts. The interaction between target domain × shift type was not significant. Error bars represent *SEM*. Right panel: Shift cost in RTs defined as the difference between shift types (i.e., within minus between) for each target domain. Dots represent individual participants. **(c)** Left panel: same as left panel in (b), but for reproduction errors. There were neither significant main nor interaction effects. Right panel: same as right panel in (b), but for reproduction errors.

Shifting attention between domains incurs a behavioural cost

In our analysis, we focused on double-cue trials, as these trials uniquely enabled us to compare within- and between-domain shifts. **Figure 1b** depicts the average response-initiation times (response times, RTs) relative to dial onset as a function of the target domain and shift type. We found RTs to be increased when reorienting attention between domains compared to reorienting within the same domain ($F_{(1,24)}$ = 4.581, p = 0.043, η^2_G = 0.002). This between-domain shift cost replicates previous behavioural findings[16–22]. Moreover, RTs to the internal targets were faster than to external targets ($F_{(1,24)}$ = 40.411, p < 0.001, η^2_G = 0.192), likely owing to the fact that their behavioural response could be prepared ahead of the response stage[29,30]. The interaction between target domain and shift type was not significant ($F_{(1,24)}$ = 0.534, p = 0.472, η^2_G < 0.001), suggesting that the effect of shift type did not depend on target domain (i.e., whether participants reported an internal or an external item).

As shown in **Figure 1c**, reproduction accuracy was consistently high, yielding no significant main effect (target domain: $F_{(1,24)}$ = 1.697, p = 0.205, η^2_G = 0.019; shift type: $F_{(1,24)}$ < 0.001, p = 0.990, η^2_G < 0.001) or interaction ($F_{(1,24)}$ = 0.280, p = 0.601, η^2_G < 0.001).

Distinct patterns of brain activity for within- vs. between-domain shifts

Our main aim was to investigate whether and when within- and between-domain shifts evoked reliably different neural activity patterns. To this end, we applied a time-resolved cross-validation decoding approach to MEG data collected while participants performed the task. To target our primary question, we first trained a Linear Discriminant Analysis (LDA) classifier to distinguish external-to-external and internal-to-internal trials (i.e., within-domain shifts) from external-to-internal and internal-to-external trials (i.e., between-domain shifts) on a timepoint-by-timepoint basis. We used the Area Under the Curve of the Receiver Operating Characteristic Curve (ROC-AUC) as the evaluation metric[31]. The orthogonal manipulation of the target domain, shift type, and cued location ensured that the decoding of within- vs. between-domain shifts could be isolated, without being influenced by merely detecting the currently cued attentional domain or cued location.

We found reliable decoding of within- vs. between-domain shifts. As expected, no decoding of reorienting attention within vs. between domains was observed until after the second cue **(Fig. 2a)**. Within- vs. between-shift decoding reached significance approximately 360 ms after the second-cue onset and persisted until external display presentation (~1860-3000 ms, cluster p < 0.001).

Since involuntary eye movements may contaminate MEG recordings and lead to decodable non-brain signals[32,33], we trained another classifier to distinguish within- vs. between-domain shifts but this time using only the x- and y-coordinates of the eye-tracking data. We could not decode the shift type based on the eye-tracking signal (**Supplementary Fig. 1a**). Moreover, we assessed whether there was mutual information between the eye-tracking and the MEG data. The mutual information is shown in **Supplementary Figure 1b**, with mutual information values of zero indicating that the MEG and eye-position signals are independent and larger values indicating a greater relationship between them. There was almost no mutual information between the eye-tracking and the MEG data when decoding within- vs. between-domain shifts. However, we noted that two participants were outliers, exhibiting stronger mutual information than the rest (**Supplementary Fig. 1c**). We therefore repeated the MEG decoding analysis while excluding these participants. This revealed that MEG-based orientation decoding was not driven by participants with strong mutual information (**Supplementary Fig. 1d**).

To assess if the pattern of neural activity underlying the classification performance was stable or evolved dynamically over time, we performed a generalisation-across-time analysis[34,35]. Significant points off the diagonal would indicate that the classifier, when trained on data from time point A, can generalise to data from time point B. In our case, the evolving pattern of neural activity that distinguished within- from between-domain attention shifts was highly dynamic. The classifier only significantly generalised over brief amounts of time (**Fig. 2b**; 1st cluster p < 0.001, 2nd cluster p = 0.027). This indicates that although we can consistently distinguish between trials in which participants shifted attention within vs. between domains, the neural code driving this distinction varied moment-to-moment. Our primary decoding analysis included data from all MEG sensor locations as features in the

classifier. A searchlight analysis (i.e., iterative decoding for subsets of MEG sensors) was used to characterise the scalp distribution of sensors sensitive to within vs. between domain shifts. **Figure 2c** shows that the neural patterns distinguishing within- vs. between-domain shifts were broadly distributed.

Together, these data thus reveal distinct patterns of brain activity associated with between-domain attention shifts as compared to within-domain attention shifts and reveal how these patterns occur early, are highly dynamic, and are broadly distributed.

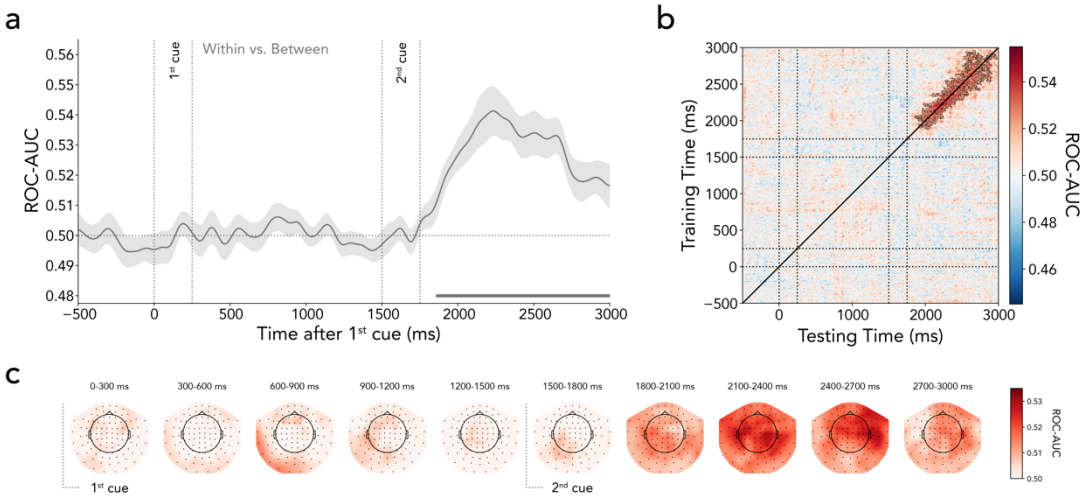


Figure 2. Time-resolved MEG decoding of within- vs. between-domain shifts of attention. (a) Average classifier performance for discriminating within- from between-domain shifts plotted over time. Time course shows *M* ± *SEM* across participants. Cluster-permutation corrected significant time points are indicated with horizontal lines. **(b)** Temporal generalisation matrix reflecting classifier performance in discriminating the two shift types. The diagonal corresponds to the time course illustrated in (a). Black outlines indicate significant decoding-clusters. **(c)** The MEG topography averages in 300-ms steps showing which sensors most strongly contribute to the overall decoding performance.

Shift-type decoding influences subsequent behavioural performance

To ascertain whether patterns of brain activity discriminating within- from between-domain shifts had functional consequences for behavioural performance, we obtained single-trial decoding scores and compared RTs as a function of low vs. high decoding performance. We expected a higher behavioural between-domain shift cost when classifier performance was high as compared to low. To address this hypothesis, we categorised trials based on their average classification score within the interval between 360 to 1500 ms after the second cue (i.e., significant cluster interval in **Fig. 2a**). **Figure 3** depicts RTs as a function of shift type (within vs. between) and median split (low shift-type decoding vs. high shift-type decoding). Planned comparisons revealed a significant between-domain shift cost only when comparing trials with high classifier performance ($t_{(24)}$ = 2.348, p = 0.027, d = 0.470). No difference occurred between shift types when considering trials with low classifier performance ($t_{(24)}$ = -0.049, p = 0. 961, d = 0.010). These results tentatively suggest that the ability to decode within- vs. between-domain shifts of attention had consequences for later behavioural performance.

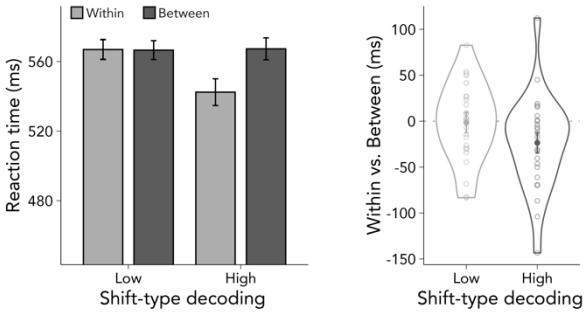


Figure 3. Decoding performance is behaviourally relevant. Left panel: Reaction times (RTs) as a function of shift type (within vs. between) and median split (low vs. high). Error bars represent *SEM*. Right panel: Shift cost in RTs defined as the difference of within- and between-shift trials for each half of decoding strength. Dots represent individual participants.

Decoding external vs. internal states associated with the first and second cues

Having uncovered evidence for distinct patterns of brain activity during shifts of attention within vs. between domains, we investigated whether and when the brain differentiated the attended domain before and after the shifts occurred. As we noted previously, owing to our experimental design, decoding of the currently attended domain (external vs. internal) was orthogonal to the above decoding of within- vs. between-domain shifts.

We started our investigation by decoding external vs. internal attentional states, based on the labels of the first cue. To this end, we trained a classifier to distinguish external first-cue trials (i.e., external-to-external, external-to-internal) from internal first-cue trials (i.e., internal-to-internal, internal-to-external). **Figure 4a** shows successful decoding of the attentional domain from approximately 310 ms after first-cue onset (cluster $p < 0.001$). Interestingly, the significant decoding period extended into the second-cue phase, revealing that information about the initially cued attentional state persisted throughout the delay following the second cue.

Temporal-generalisation analysis unveiled an off-diagonal shape of generalisation following the first cue, indicating temporal stability of the neural pattern evoked by the initially cued domain (**Fig. 4b**; cluster $p < 0.001$). In contrast, after second cue onset, decoding performance occurred mainly along the diagonal, with a narrower spread to neighbouring timepoints. In addition, we found off-diagonal below-chance decoding (i.e., the classifier systematically guessed incorrectly; 1st cluster $p = 0.024$, 2nd cluster $p = 0.025$). The searchlight decoding analysis demonstrated that the contribution of sensors to the external vs. internal decoding was broadly distributed, with peaks observed in the posterior and left-central sensors occurring between 600 and 1200 ms (**Fig. 4c**).

To determine whether the decoding of the first-cued attentional domain was predominantly influenced by a particular trial type, we employed two additional classifiers: one to distinguish between external-to-external and internal-to-external shifts, and another to distinguish between internal-to-internal and external-to-internal shifts. **Supplementary Figure 2a** shows that classifier performance did not differ between these two decoding analyses, thus resembling the general external- vs. internal-domain decoding time course shown in **Figure 4a**. In addition, we again took precautions to ensure that eye movements did not drive the external vs. internal first-cue MEG decoding results. The strength of the decoding based on gaze position was relatively weak and far from sufficient to explain the effects when decoding brain activity (**Supplementary Fig. 1e**). Furthermore, exclusion of participants who exhibited the highest mutual information between eye position and MEG decoding did not alter the MEG decoding results (**Supplementary Fig. 1f-h**).

Next, we aimed to decode the domain indicated by the second cue. To this end, we trained a classifier to distinguish between external second-cue trials (i.e., external-to-external, internal-to-external) vs. internal second-cue trials (i.e., internal-to-internal, external-to-internal). Our results revealed significant decoding of the second-cued attentional domain from approximately 280 ms after second-cue onset (i.e., 1780 ms after first-cue onset; cluster $p < 0.001$; **Fig. 4d**). As was intended in our design via orthogonally manipulating the first and second-cued attentional domain, the second cued

domain could not be decoded prior to the onset of the second cue. The temporal-generalisation matrix in **Fig. 4e** showed a temporal spread, implying some stability in the activation pattern (cluster $p < 0.001$). Similar to the first-cue analysis, sensor searchlight analysis indicated a distributed pattern of sensor contribution to decoding external vs. internal attentional domains (**Fig. 4f**).

To ensure that the results of the domain-decoding were not driven by the contrast between specific trial types, we trained one classifier to discriminate between external-to-external and external-to-internal shifts and another one to discriminate between internal-to-internal and internal-to-external shifts. While both decoding time courses reached significance following the second cue, the internal-to-internal vs. internal-to-external contrast reached significance earlier than the external-to-external and external-to-internal contrast (**Supplementary Fig. 2b**). These findings revealed earlier domain differentiation when initially cued about the internal domain compared to when previously cued about an external item. In line with the first-cue analysis, we ensured that the second-cue MEG decoding results were not dependent on eye movements (**Supplementary Fig. 1i-l**).

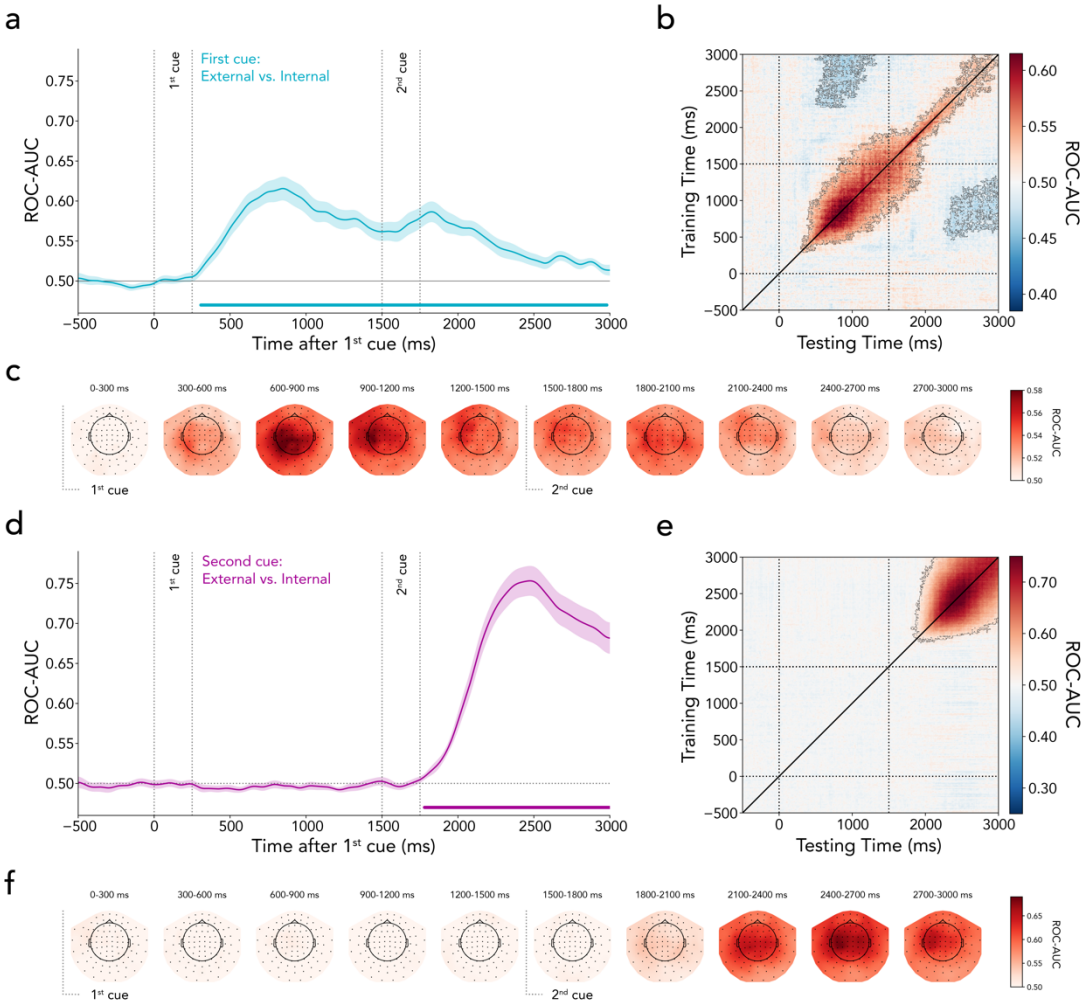


Figure 4. Time-resolved MEG decoding of external vs. internal attention. (a) Average classifier performance for distinguishing external from internal first cues plotted over time. Time course shows *M* ± *SEM* across participants. Cluster-permutation corrected significant time points are indicated with horizontal lines. **(b)** Temporal generalisation matrix reflecting classifier performance in discriminating the first-cued attentional domain. The diagonal corresponds to (a). Black outline indicates significant decoding-clusters. **(c)** The MEG topography averages in 300-ms steps showing which sensors most strongly contribute to the overall decoding performance in discriminating external- and internal first cues. **(d)** Same as (a) but based on second-cue labels. **(e)** Same as (b) but based on second-cue labels. **(f)** Same as (c) but based on second-cue labels.

Lingering first-cued attentional domain is linked to within- vs. between-domain shift decoding

Our results demonstrate that attentional domains indicated by the first and second cue could be decoded. Interestingly, the significant decoding of the first-cued attentional state persisted following the second-cued attentional domain. Supplementary analyses revealed that during the period following the second cue, the decodability of the continued information concerning the first cue was not contingent on the quality of processing the second cue (**Supplementary Fig. 3a, b**). Hence, this finding suggests that lingering processing related to the attentional domain of the first cue and instantiating the attentional domain associated with the second cue proceeded in tandem. In other words, information regarding the past and current attentional state co-existed after the second cue.

Previous task-switch research has argued that switch costs can arise from interference from previous, but now task-irrelevant information[36,37]. We therefore tested whether a similar interaction might arise between lingering activity related to the first-cued domain and the ability to decode within- vs. between-domain shifts. To this end, we median-split our data into trials with bad vs. good first-cued domain decoding following the onset of the second cue (i.e., 1500-3000 ms). Next, we trained a classifier to distinguish within vs. between-domain shifts based on these two data subsets. The within- vs. between-domain shift type decoding only yielded a significant cluster in trials with high decodability of the lingering first-cued domain (**Fig. 5a**; ~1910-2890 ms, cluster $p < 0.001$). This was corroborated by a direct comparison, showing that trials with stronger lingering first-cue decodability had significantly more successful within- vs. between-domain decoding (~2330-2870 ms, cluster $p = 0.005$). **Figure 5b** displays the within- vs. between-shift decoding topographies for low (top panel) and high (bottom panel) lingering first-cue decoding. Consequently, our data suggest that the lingering first-cued attentional state is linked to the ability to decode the performed shift type.

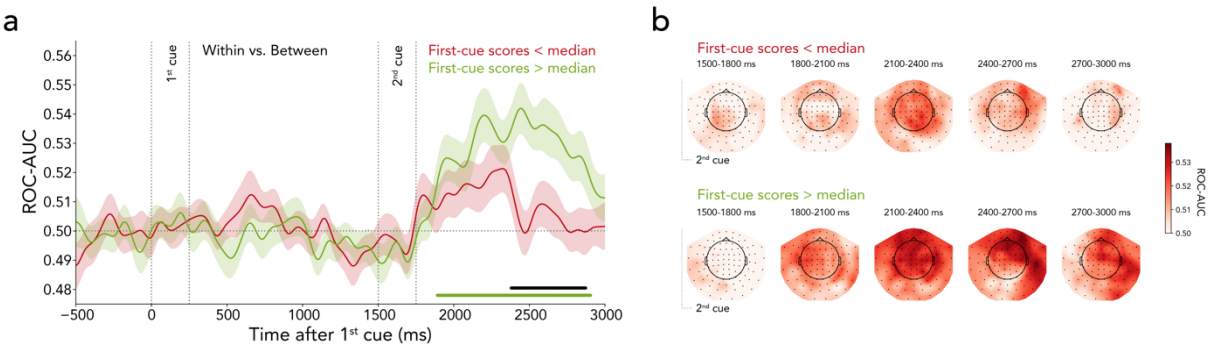


Figure 5. Relationship between lingering first-cued attentional domain and performed shift type. (a) Time-resolved average classifier performance for discriminating within- from between-domain shifts as a function of bad (red) vs. good (green) decoding of the lingering first-cued attention state. Time courses show *M* ± *SEM* across participants. Horizontal lines indicate significant clusters, black line denotes difference between decoding time courses. **(b)** MEG topography averages in 300-ms steps showing which sensors most strongly contribute to the overall decoding performance in discriminating within- and between-domain shifts. Top panel: trials with low decodability of the lingering first-cue domain. Bottom panel: trials with high decodability of the lingering first-cue domain.

Spatial shifts of external and internal attention as tracked in alpha lateralisation

Thus far, we have delved into the concept of shifting from two essential perspectives. Firstly, we contrasted brain activity when transitioning between domains, in comparison to shifting attention within a single domain. Secondly, we examined whether brain activity distinguished between the attentional domain itself, following the first and the second cue. Our task design additionally enabled us to investigate attentional shifts from a third perspective, involving the spatial orienting of attention between locations. We intentionally designed our task to guarantee that shifts occurred between items positioned in distinct hemifields, enabling us to investigate the lateralisation of MEG activity according to the location to which attention was shifted.

A canonical marker for spatial attention shifts is the lateralisation of posterior 8-12 Hz alpha activity. Contra- vs. ipsilateral attenuation of 8-12 Hz alpha oscillations has been linked to relative

increases in excitability in sensory brain areas resulting from shifts of spatial attention in anticipation of relevant sensory stimuli as well as during selection of relevant working-memory content[14,30,38–42]. Here, we therefore focused on alpha lateralisation to delineate spatial shifts of attention after external and internal first and second cues. These additional analyses serve to complement our primary findings from the decoding analyses and to relate our work to the existing neuroscience literature on internal and external spatial shifts of attention.

Figure 6a and 6d show the time- and frequency-resolved lateralisation of spectral power in contra- vs. ipsilateral visual sensors for external and internal first cues, respectively. At the time of the second cue, participants were always required to shift attention from one hemifield to the other. Consequently, relative suppression of contralateral activity in response to the first cue was followed by relative enhancement of contralateral activity (i.e., a reversal of lateralisation) after second-cue onset (external first cued: 1st cluster $p < 0.001$, 2nd cluster $p < 0.001$; internal first cued: 1st cluster $p < 0.001$, 2nd cluster $p = 0.004$).

To chart the time course of alpha lateralisation, we averaged the time-frequency data along the classical alpha band (8-12 Hz; Fig. 6g). We confirmed the typical contralateral alpha attenuation following both external and internal first cues. Clusters of lateralised alpha attenuation alternated between hemispheres after the first (external first cued: ~330-1110 ms, cluster $p < 0.001$; internal first cued: ~430-1730 ms, cluster $p < 0.001$) and second cue (external first cued: ~1910-3000 ms, cluster $p < 0.001$; internal first cued: ~1850-2980 ms, cluster $p = 0.004$). The time course of the initial contralateral alpha attenuation was similar for both external- and internal-domain shifts. However, in line with prior research, the duration of the alpha modulation differed[14,43]. Contralateral alpha suppression was more transient following internally directing first cues and more sustained following externally directing first cues, as indicated by a significant cluster from approximately 1030 to 1740 ms after first-cue onset (cluster $p < 0.001$). A topographic inspection of external and internal first-cued trials confirmed a clear posterior topography for these effects in alpha lateralisation (Fig. 6j upper and lower panel, respectively).

The same time-frequency analysis was conducted based on the attentional domain associated with the second cue. Figure 6b and 6e illustrate that the relative suppression of contralateral activity in response to the second cue was preceded by a relative enhancement of contralateral activity (i.e., a reversed lateralisation) after first-cue onset (external second cued: 1st cluster $p < 0.001$, 2nd cluster $p < 0.001$; internal second cued: 1st cluster $p = 0.002$, 2nd cluster $p = 0.002$). As seen in the alpha-band time-courses (Fig. 6h), there was no difference between alpha-band lateralisation following the first cue (external second cued: ~360-1400 ms, cluster $p < 0.001$; internal second cued: ~360-1300 ms, cluster $p = 0.002$) or following the second cue (external second cued: ~1920-3000 ms, cluster $p = 0.003$; internal second cued: ~1920-2940 ms, cluster $p < 0.001$). Topographies, showing the difference between trials with left- vs. right-cued items, confirmed that these visual signatures were prominent in the corresponding left and right visual sensors (Fig. 6k; top panel: external cue, bottom panel: internal cue).

Together, these complementary analyses of alpha lateralisation reveal that attention lingers longer after the first cue when directed externally vs. internally, while re-orienting after the second cue proceeded similarly regardless of whether spatially re-orienting to an internal or external item.

Alpha lateralisation proceeds similarly during within- and between-domain shifts

Given the effectiveness of alpha attenuation in tracking the spatial orienting and re-orienting of attention, we finally turn to the important question of whether spatial orienting following the second cue would be delayed when shifting attention between as compared to within domains. If shifting between domains engages additional processes prior to implementing a spatial shift, we would expect to see a delay in the onset of spatial orienting when shifting attention between domains.

Figure 6c and f depict the alpha lateralisation relative to the second cue, split by whether the second cue indicated a within- or a between-domain shift (within-domain shift: 1st cluster $p = 0.003$, 2nd cluster $p < 0.001$; between-domain shift: 1st cluster $p < 0.001$, 2nd cluster $p < 0.001$). As evident form the alpha-band time-courses (Fig. 6i; within-domain shift: 1st cluster ~340-1300 ms, cluster $p = 0.002$ 2nd

cluster ~1910-3000 ms, cluster $p < 0.001$; between-domain shift: 1st cluster ~370-1380 ms, cluster $p < 0.001$, 2nd cluster ~1860-3000 ms, cluster $p < 0.001$), we observed no difference between shift types in alpha lateralisation. Again, topographical maps confirmed the distribution of the effects over corresponding left and right visual sensors (**Fig. 6l**, top panel: within-domain shift, bottom panel: between-domain shift). These findings suggest that the initiation of the spatial shift remained unaffected, despite our decoding analyses revealing distinguishable neural processing engaged during within vs. between-domain shifts. The results offer no evidence for an additional stage invoked by between-domain shifts that must be resolved before attention can be shifted in space.

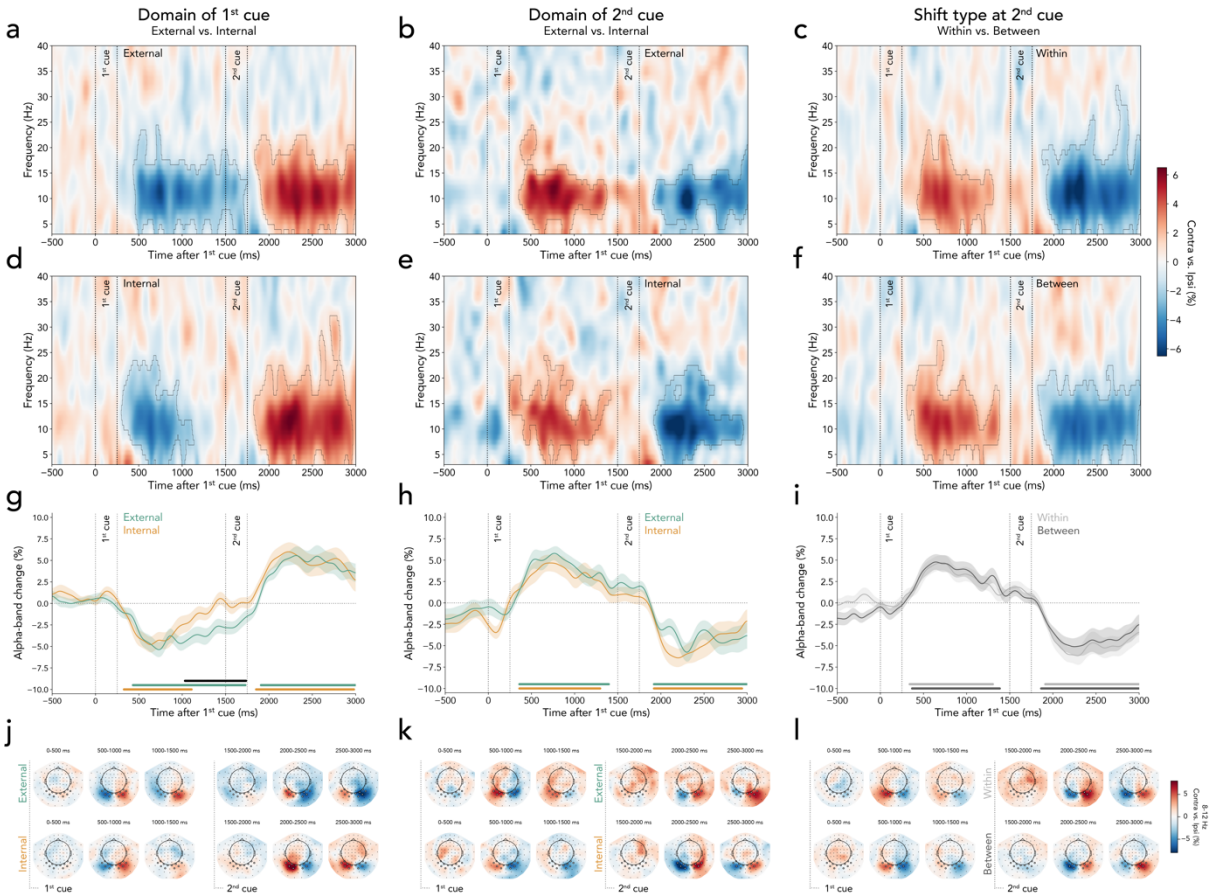


Figure 6. Lateralised neural activity relative to the location of the first-cued and second-cued item. (a) Time-frequency plots of neural lateralisation following external first cues. Plots show the difference in spectral power contra- vs. ipsilateral to the cued side, extracted over the predefined left and right sensor clusters. Black outlines indicate significant clusters. **(b)** Same as (a) but for external second cues. **(c)** Same as (a) but for second cues indicating within-domain shifts **(d)** Same as (a) but for internal first cues. **(e)** Same as (a) but for internal second cues. **(f)** Same as (a) but for second cues indicating between-domain shifts. **(g)** Time courses of neural lateralisation in the 8-12 Hz alpha band relative to the first cue. Time courses show *M* ± *SEM* across participants. Horizontal lines indicate significant clusters, black line denotes difference between time courses. **(h)** Same as (g) but relative to second cues. There was no significant difference between time courses. **(i)** Same as (g) but relative to second cues indicating within- and between-domain shifts. There was no significant difference between time courses. **(j)** Topographies of power differences following left- vs. right-directing external first cues in the 8-12 Hz alpha band. Top panel shows topographies for external first cues, bottom panel for internal first cues. **(k)** Same as (j) but for second cues. **(l)** Same as (j) but for second cues indicating within- and between-domain shifts. Top panel shows topographies for within-domain shifts, bottom panel for between-domain shifts.

Discussion

Our study uncovers direct neurophysiological evidence for distinct brain activity when shifting between perception and working-memory contents as compared to when shifting between contents within either domain. The utilisation of a carefully balanced task design – which orthogonally manipulated

shift type and target domain while also necessitating spatial shifts of attention – provided us with additional valuable insights into the dynamic neural bases that govern shifts between perception and working memory.

In principle, multiple scenarios can explain the behavioural costs and neural differences between shift types, and these need not be mutually exclusive. One obvious scenario is the presence of an additional, supraordinate supervisory control process that acts upon the neural systems that control between-domain attentional shifts[23,24,44,45]. This supraordinate control mechanism could be invoked when initiating a shift between perception and working memory but not when shifting between contents within either domain. Although it is most intuitive to consider a supraordinate control mechanism occurring when attention shifts between domains are initiated, it is also possible to imagine supraordinate control mechanisms operating at other stages of attention deployment. Attention comprises several functions to guide adaptive behaviour, such disengaging, orienting, selecting, prioritising, and output gating[46,47]. Accordingly, additional supraordinate mechanisms could operate at any or multiple points along the way. An alternative type of scenario is one that invokes no additional control process. Instead, the systems for controlling external vs. internal shifts of attention may coexist and compete for access to the relevant sensory and/or motor representations. In this scenario, the additional cost would arise from resolving a stronger degree of competition between as compared to within domains. As we outline below, our findings offer important initial clues regarding the factors that distinguish within- and between-domain shifts.

We unequivocally demonstrated distinct neural processes during within- and between-domain attention shifts. By tracking neural dynamics, we showed that differences occur early following the second cue (**Fig. 2**). These neural dynamics suggest that the shifting process itself is distinct when shifting between domains, and that differences between shift types thus unlikely emerge exclusively at the response stage. In line with that, we showed that this differential cue-related activity is behaviourally relevant, predicting later between-domain shift costs in RTs (**Fig. 3**). We cautiously argue, therefore, that previously observed between-domain shift costs in performance[17–21,24,48] are unlikely explained exclusively by interference at a late response-level response stage, but rather emerge already during the attention shifting process itself.

Interestingly, by analysing alpha-band lateralisation, we found no evidence for a change in the timing of the spatial biasing of visual activity when orienting spatial attention within vs. between domains (**Fig. 6c, f, i, l**). This observation challenges the intuitive scenario proposing that supraordinate control mechanisms are invoked when initiating an attention shift that crosses domains. Such a scenario would predict a delayed shift of spatial attention. As far as the spatial orienting functions of attention – as reflected in alpha lateralisation – we found no evidence for this. These neural alpha-lateralisation results complement a similar observation that between-domain shifts did not delay shifts in gaze biases associated with spatial orienting of attention[18], reinforcing the suggestion that there is no time-consuming additional gating step interposed before shifting spatial attention between domains.

Neural patterns differentiating within- vs. between-domain evolved relatively dynamically and were broadly distributed (**Fig. 2b**). Differences between shift types are thus not solely attributable to a single tonic state that differs between shift types but likely results from the dynamic interplay of multiple states. Moreover, the distributed nature of the within- vs. between-shift decoding suggests that large-scale networks may be involved (**Fig. 2c**). It will be important for future brain-imaging investigations using functional Magnetic Resonance Imaging (fMRI) and/or animal models to substantiate this broad distribution and to pinpoint the precise spatial underpinnings of the here-reported findings.

It is useful to consider that each within- and between-domain shift is a consequence of the attended domains before and after the shift. Therefore, to understand the nature of the two shift types, it is also essential to investigate the neural dynamics evoked by the first- and second-cued attentional domains. While some studies have emphasised the shared neural systems and dynamics between selection mechanisms in external and internal attention[10,15], others have also highlighted important differences[7,11,46,47,49–51]. Underscoring the differences, our data demonstrate successful decoding of

external vs. internal attention (**Fig. 4**). Moreover, we show that decoding of the first-cued domain persisted even after the presentation of the second cue.

The carry-over decoding effect of the first-cued domain occurred regardless of the domain indicated by the second cue. The finding highlights the fact that, depending on the nature of the preceding attentional domain, the brain exists in a different attentional state even after the second cue (i.e., the attentional state invoked by the second cue did not "overrule" or "wipe-out" the lingering state invoked by the first cue). Multiple processes could have contributed to lingering decoding of the first cued domain, and we are unable to adjudicate among these without further investigation. For example, anticipating the spatial location or sensory features of a target in the external domain may have elicited sustained neural activity according to spatial or feature-based receptive field properties [52–54]. In contrast, cueing internal representations in working memory often triggers more transient sensory modulation[14,43]. Our alpha lateralisation findings hint at these effects with more prolonged modulation following spatial cues indicating targets in the external than internal domains (**Fig. 6a, d, g, k**). Alternatively, the second cue may have activated latent functional states in working memory through a 'pinging' mechanism[55–57]. Although the second cue indicated a different item than the first cue, its associated sensory input may have interacted with latent or activated working-memory states, thus acting as a ping for the initially cued item. Another possibility is that the first cue may have placed the system in different states of sensorimotor preparation. If an internal item was cued, its representation could have been readily transformed into a motor code[29,30,58], whereas, this was not possible when sensory items were anticipated by external attention. Therefore, the sustained decoding of the first-cued attentional domain following the second cue could reflect a lasting effect of output gating following internal but not external first cues. Any, or all, of these differences may have contributed to the successful decoding of the first-cued domain after the second cue.

The decodability of the lingering first-cued attentional domain did not trade-off against the decoding of the second-cued attentional domain (**Supplementary Fig. 3**). Instead, after the second cue, information processing unfolded for both the first- and second-cued attentional domains in tandem, implying the co-existence of brain states associated with preceding and current attention states. Though our data do not directly tap into the nature or potential functional relevance of these co-existing states, they did reveal how the degree to which the first state lingered (i.e., could still be decoded after the second cue) predicted how well our classifier could distinguish within- from between-domain shifts (**Fig. 5**). This suggests that attentional states preceding shifts shape upcoming attentional states in the nearby future. What role such lingering states play, and whether/how they mediate parallel processing of contents from preceding and current attention states remains an exciting endeavour for future studies.

Our task and findings pave the way for investigating the dynamic mechanisms orchestrating the seamless shifts of attention between external and internal contents to guide adaptive flexible behaviour. Our investigation serves as an initial exploration, opening the door to relevant unanswered questions and encouraging further task manipulations. For instance, in our study, when participants were internally cued, they were able to select the relevant content from working memory. Conversely, when externally cued, attention could be directed to a specific spatial location or feature, but the actual attentional selection occurred only after the external display presentation. In future studies, it would also be interesting to investigate the dynamics of transitioning between external and internal content in tasks where external information is presented concurrently with external cues and remains perceptually available until the reporting stage. Another important avenue for future studies would be to situate and study attention shifts in more naturalistic tasks. Such additional manipulations will likely yield complementary insights contributing to a more comprehensive understanding of the principles and mechanisms that govern attention shifts between perception and working memory, in service of adaptive behaviour.

In summary, we revealed distinct neural dynamics for shifting within vs. between perception and working-memory contents. We demonstrated how these distinct neural processes evolve dynamically, are broadly distributed, predict behavioural performance, and occur without delaying the

timing of spatial orienting of attention. Besides successfully decoding the shift type, we were able to decode the attentional domain before and after a shift – representing its starting point and destination. Our results uncovered a direct link between the previously attended domain and the performed shift type, suggesting that sustained processing of the first cue-domain affects processing of the second cued domain, in a way that differentiates between from within-domain shifts. Brain imaging methods with higher spatial resolution should build on these findings to identify the brain networks that orchestrate shifts of attention between domains and investigate how they interact with the more established control networks involved in modulating shifts of attention within the external and internal domains.

Methods

Participants

The study was approved by the Central University Research Ethics Committee of the University of Oxford. The sample size was set to *n* = 25, based on a previous study from our lab targeting a similar research question via behavioural measures[18]. To yield the targeted number of participants, we collected data from 26 participants in total. One participant was excluded following our a-priori behavioural trial-removal procedure (see Behavioural analysis). Of the remaining 25 participants (age range: 19 to 36; mean age: 26.24; 14 female, 11 male), 22 reported being right-handed (three left-handed). Individuals provided written informed consent before participating in the study and were paid £15 per hour.

Task and procedure

The current task design builds on our previous behavioural study[18]. In each trial, four randomly oriented bars appeared within coloured circular placeholders at quadrant locations. Early in the trial, two oriented bars were bilaterally presented and had to be encoded into working memory. Later in the trial, the remaining two bars appeared briefly in the remaining placeholders followed by masks. Between the initial and final presentation of bars, one or two centrally presented colour cues signalled which of the four bars participants had to report by accurately reproducing its exact orientation. Cues that indicated a previously encoded bar directed attention internally (retro-cues or internal cues), indicating which item from working memory should be reported. In contrast, cues that indicated a previously unoccupied placeholder directed attention externally (pre-cues or external cues), specifying which item from the upcoming perceptual array should be reported.

At the start of each trial, a white central fixation cross (RGB value: [255, 255, 255]) surrounded by four coloured circular placeholders appeared against a black background (RGB value: [0, 0, 0]) for 750 ms. Each of the placeholders appeared in one of four highly distinguishable colours (RGB values: light blue [0, 159, 183], orange [217, 120, 0], green [47, 167, 0], pink [247, 35, 255]). The location mapping of the four placeholder colours was randomly determined on each trial. Placeholders subtended 3.5° visual angle and were centred in the top-left, top-right, bottom-left, and bottom-right at a distance of 4.5° visual angle from fixation. Placeholders stayed on the screen until the response was required at the end of the trial.

Next, two tilted grey bars (RGB value: [50, 50, 50]) subtending approximately 0.5°×3.4° appeared for 500 ms inside two of these placeholders (referred to as *internal display*). The orientation of the bars was randomly determined between 0 and 180°. Participants were instructed to memorise the orientation of these two bars. There were four possible configurations of the internal display, which occurred equally often throughout the experiment: the bars could appear in the placeholders located in the (1) top-left and top-right, (2) bottom-left and bottom-right, (3) top-left and bottom-right, and (4) top-right and bottom-left. As such, at the time of the internal display, there was always one bar presented in each visual hemifield.

The internal display was followed by a delay of 750 ms, during which only the fixation cross and placeholders remained on the screen. The central fixation cross then changed colour to match that of one of the placeholders for 250 ms. The colour change acted as an informative cue, indicating the potential item to be reported. This cue could indicate a placeholder that previously contained an item during the internal display (i.e., retro-cue) or a placeholder in which no item had been presented yet (i.e., an item of the upcoming external display; pre-cue). The cued location and attentional domain were pseudo-randomised such that each of the four placeholder locations and each domain was equally often indicated by the cue.

In 50% of the trials, the central fixation cross changed colour again after 1250 ms to match the colour of another placeholder for 250 ms (i.e., double-cue or shift trials). In the remaining 50% of the trials, only a single cue (i.e., single-cue or stay trials) was presented. In single-cue trials, the cue appeared at the same time as the first cue in double-cue trials and was 100% instructive, always

indicating the target for report. Single-cue trials were included to increase the relevance of the first cue and ensure that participants shifted attention to the first-cued item rather than simply ignoring the first cue entirely. Single-cue trials did not serve purpose for behavioural or neural comparison relevant to the research question addressed in the present study.

The double-cue trials were the important trials for our research question since these trials required a shift of attention. In particular, the double-cue condition was further divided into two sub-conditions: within- and between-domain shifts. In within-domain shift trials (i.e., external-to-external or internal-to-internal), the second cue indicated the alternative item within the same domain as the first cue occurring in the other hemifield. In between-domain shift trials (i.e., external-to-internal or internal-to-external), the second cue indicated the item in the alternative domain occurring in the other hemifield. Thus, in within- and between-domain shift trials, attention always needed to be reoriented from one hemifield to the other. The second cued location and domain were pseudo-randomised such that each of the four placeholder locations and each domain were equally often indicated. In double-cue trials, the second cue instructed the target item for report with a certainty of 100%, thus overruling the first cue. Accordingly, the target item was always identical to the second cued item. The colour of the first and second cue always differed in double-cue trials.

The single cue in single-cue trials and the second cue in double-cue trials were followed by an inter-stimulus interval of 1250 ms before the onset of a perceptual display (referred to as *external display*), containing two randomly tilted bars. The new items appeared in the yet-unoccupied placeholders for 50 ms. For example, if the bars in the internal display appeared in the top-left and bottom-right placeholders, bars in the external display occurred the top-right and bottom-left placeholders. To make the perceptual discrimination challenging, items in the external array were masked. The masking array consisted of four overlayed titled bars presented in each of the placeholders for 100 ms. At each mask location the four tilted bars differed 45° from the neighbouring bar. The overall orientation of each mask was randomly drawn at 0 ms and 50 ms, thus creating the impression of a dynamic display. The masking parameters were set to equate the difficulty of internal and external reports, based on piloting efforts.

The offset of the dynamic mask was followed by a 100-ms delay, after which a visual response dial was displayed, always starting in a random position. The response dial surrounding the fixation cross was the same size as the placeholders and included markers along a circle that corresponded to the ends of a bar. The dial rotated leftwards when pressing the left button and rightwards when pressing the right button of the response device held in the dominant hand. The dial rotated with a speed 1/10° per millisecond. Participants had unlimited time to complete the orientation report. When the dial aligned with the orientation of the target item, participants pressed either of the two buttons of the response device held in the non-dominant hand to confirm their response and continue with the task.

After a 100-ms delay in which only the fixation cross stayed on the screen, participants received feedback in the form of a number ranging from 0 to 100, with 100 indicating a perfect report and 0 indicating that the adjusted orientation was perpendicular to the angle of the target item. Feedback was presented 0.7° above the central fixation cross for 300 ms. Trials were separated by an inter-trial interval randomly drawn between 750 and 1000 ms. Between blocks, participants were presented with their average reproduction accuracy in the previous block.

The experiment consisted of 512 trials divided across 8 blocks, each including 64 trials. Of the total trials, 50% (256) were single-cue trials and 50% (256) were double-cue trials. Participants were equally likely to report an internal or external item in single-cue trials (128 each). Double-cue trials varied in the type of shift performed, that is, 50% (128) required a within-domain shift and the other 50% (128) a between-domain shift. Within-domain and between-domain shift trials were further split by the target domain. Ergo, the within-domain shift trials included 50% (64) internal-to-internal and 50% (64) external-to-external shifts, while the between-domain shift trials included 50% (64) external-to-internal and 50% (64) internal-to-external shifts. In each of the four unique shift conditions, participants were equally likely to shift their spatial attention between any of the eight possible cross-

hemifield combinations of placeholders. The orthogonal manipulation of the to-be-reported *target domain* (external vs. internal) and the *shift type* (within-domain shift vs. between-domain shift) enabled us to independently investigate these two factors in a systematic manner. Each block consisted of 50% (32) single- and 50% (32) double-cue trials.

To become familiarised with the procedure of the experiment, participants performed a practice block before starting the data collection. The whole experiment lasted approximately 65 min.

Data acquisition

We obtained whole-head MEG recordings in a magnetically shielded room using a 306-sensor MEG system (204 first-order planar gradiometers, 102 magnetometers; TRIUX neo, MEGIN OY, Espoo, Finland) at the Oxford Centre for Human Brain Activity. The MEG signal was sampled at 1000 Hz, with a high-pass filter at 0.1 Hz and a low-pass anti-aliasing filter at 330 Hz. Each experimental block was recorded separately, resulting in eight runs for most participants. However, due to technical issues, two individuals had fewer recordings, with six and seven recordings obtained, respectively.

Prior to data acquisition, we digitized the individual head shape with an electromagnetic position and orientation monitoring system (FASTRAK, Polhemus, Colchester Vermont, USA). Shapes included fiducial landmarks (nasion, right and left preauricular points) and about 250 additional points evenly spread on the participants' scalp. Five Head Position Indicator (HPI) coils were placed on participants' mastoid bones and forehead to keep track of participants' head position inside the dewar through electromagnetic induction before and after each recording block.

Participants performed the task in a seated position. Participants were instructed to avoid head, body, and limb movements and to keep strict eye fixation during the experimental blocks. In addition to electrocardiography (ECG) and horizontal and vertical electrooculography (EOG), eye positions were continuously monitored by an MEG-compatible eye-tracking device with a sampling frequency of 1000 Hz (Eyelink 1000, SR-Research Ltd., Ottawa, Ontario, Canada). At the beginning of each experimental session, participants performed an eye-tracking calibration task to verify their gaze position on the screen. Calibration was repeated if drift was noticed during the experimental session.

We used the PsychoPy package (version 2021.2.3)[59] in Python for stimulus generation and stimulus delivery. The stimuli were projected on a translucent whiteboard using a DLP LED projector (ProPixx, VPixx Technologies Inc., Saint-Bruno, Quebec, Canada) at a 120-Hz refresh rate. The whiteboard was located at 120-cm distance from the participant, and it provided a projection area of 55×31-cm and 1920×1080-pixel resolution. A bimanual fibre-optic response device was used to collect manual responses.

Behavioural analysis

Behavioural data were analysed in R Studio[60]. During pre-processing (similar to [18]) trials were removed when reaction times (RTs; calculated from dial to response onset) exceeded 5000 ms. Next, we removed trials for which the remaining RTs were 2.5 *SD* above the individual mean across all conditions. Moreover, we removed datasets with average reproduction errors (calculated by averaging the absolute difference between the angle of the target item and the reported angle) equal to or higher than 45° in any of the conditions (n = 1). After these exclusion steps, datasets from 25 participants remained in the main behavioural analysis, with an average of 16.64 ± 3.91 (M ± SD) trials removed. Additionally, we removed the bad MEG trials from the behavioural data analysis.

The conditions of interest were *target domain* (having to report an external vs. internal item) and *shift type* (within vs. between). When comparing more than two means, we applied a repeated-measures analysis of variance (ANOVA) and reported η^2_G as a measure of effect size. When evaluating only two means, we applied a paired samples t-test and report Cohen's d as a measure of effect size.

MEG pre-processing

MEG data were analysed in Python using MNE-Python (version 1.4.2)[61] and OSL (version 0.2.0)[62] combined with custom code. For each individual run, we applied the Elekta MaxFilter (version 2.2) implementation of Spatiotemporal Signal Source Separation to remove external noise (TSSS)[63] to continuously compensate for head movements occurring during the run and to realign all runs per participant to a participant-specific head position. The reference head position for each participant was determined by minimising the head position distance between runs. Next, data were downsampled to 250 Hz and low-pass filtered at 50 Hz. Importantly, no high-pass filter was applied, since high-pass filtering can temporally displace multivariate information[64]. Bad sensors were identified and interpolated using a generalized extreme studentised deviate test (ESD)[65] at a 0.05 significance threshold. On average, 1.62 ± 3.11 bad gradiometers and 1.42 ± 2.02 bad magnetometers were identified and interpolated per run. We applied an independent component analysis (ICA) to regress out eye-movement and cardiac-related activity. Note that the maximum number of ICA components to be extracted was determined by the residual degrees of freedom after TSSS rank reduction during Maxfiltering[66]. Data were epoched from -500 to 3000 ms relative to the first cue onset without applying baseline correction[64]. Next, we combined all individual runs for each participant by concatenating the data. Additionally, we remapped magnetometers onto gradiometers, allowing us to conduct all subsequent analyses using a unified sensor type. Lastly, we employed an ESD trial rejection approach to detect and exclude trials with high variance. Per participant, 5.24 ± 6.96 (*M* ± *SD*) trials were removed. In addition, we removed the trials that were excluded in the behavioural data analysis.

Time-resolved multivariate pattern analysis

We performed time-resolved multivariate analysis on broadband MEG data using the Scikit-learn toolbox (version 1.3.0)[67] and built-in decoding functions provided by MNE-Python (version 1.4.2)[61]. The decoding pipeline closely followed the approach outlined in previous literature[68].

We extracted patterns of brain activity separately for each participant on a timepoint-by-timepoint basis (*mne.decoding.SlidingEstimator*). First, features (i.e., data at each time point for each sensor) were centred and scaled based on data from all trials (*sklearn.preprocessing.StandardScaler*). The true spatial dimensionality of the MEG data after the Maxfilter preprocessing algorithm is around 70, rather than the 204 gradiometer sensors[66]. Therefore, to reduce the number of redundant features for decoding, we performed Principal Component Analysis on the standardised data while maintaining 99% of the variance (*sklearn.decomposition.PCA*). Training and testing were done on the same data using a 10-fold stratified cross-validation procedure (*sklearn.model_selection.StratifiedKFold*). First, trials were randomised in order, and divided into ten equal-sized folds. Next, a leave-one-out procedure was used on the 10 folds, such that the classifier was trained on 9 folds and tested on the remaining fold. This procedure was repeated 10 times until each fold was used once for testing. A Linear Discriminant Analysis (LDA; *sklearn.discriminant_analysis.LinearDisciminantAnalysis*) classifier was utilised to train on the provided training data and labels. LDA is a supervised dimensionality reduction technique that aims to find a linear combination of features that maximizes class separability which has been recommended for neuroscientific decoding studies[34]. As a scoring metric, we used the Area Under the Curve of the Receiver Operating Characteristic Curve (ROC-AUC). This scoring metric takes into account the trade-off between true and false positive rates is considered a sensitive, nonparametric, and criterion-free measure of classification[31,69]. A ROC-AUC value of 0.5 means chance level classification performance. The obtained classifier scores were averaged over all 10 folds, yielding a single decoding estimate per participant, timepoint, and condition. Lastly, decoding time courses were smoothed with a one-dimensional Gaussian filter with a standard deviation of 10.

In addition, we used temporal-generalisation methods[35] to examine whether there are recurrent neural activation patterns over time, that is, whether an activation pattern observed at one timepoint is observed again at a later timepoint. Generalisation off the diagonal in the temporal generalisation matrix is driven by overlap in the activation pattern at two different timepoints, and therefore implies that the activation pattern recurred. Temporal generalisation allows us to examine

such potential overlaps in representations over time. This approach results in timepoint-by-timepoint decoding matrices in which each cell corresponds to a classification accuracy at a unique training- and test-timepoint combination.

To test which sensors contributed most to the classifier likelihoods observed in our multivariate methods, we ran a sensor-space searchlight decoding analysis[13,34,70]. For this analysis, we ran the same analysis as outlined above across small clusters of 15 neighbouring MEG sensors, resulting in decoding scores for each sensor and timepoint. These decoding scores were averaged into time clusters of 300 ms each, resulting in 10 topographical maps between 0 and 3000 ms. These topographical plots show whether decoding results were primarily driven by specific sensor clusters. We did not test any specific hypotheses regarding the spatial distributions of the effects and therefore did not run any statistical tests on the results.

Even though stratified cross-validation is highly sensitive to class imbalances[31], we additionally applied between-class balancing using under-sampling prior to the decoding analysis, such that the number of trials was equal across the classes. This further ensured that during training the classifier would not only achieve a high decoding performance merely by systematically and blindly voting for the majority class[71]. Note that as the design was balanced in terms of trial counts, the between-class balancing was employed to account for small imbalances due to trial rejection procedures used for data cleaning.

Relationship between decoding and behaviour

To investigate the relationship between the observed decoding of within-domain vs. between-domain shifts of attention and the behavioural shift costs, we employed the following approach. Firstly, we identified trials with low and high classifier performance for each participant by computing timepoint-by-timepoint decoding scores and averaging them across the interval of interest. Subsequently, we performed a median split based on these average decoding scores, creating two groups of trials: one with low decoding scores and the other with high decoding scores. We then included the "median split" factor in our behavioural analysis when comparing within- to between-domain shifts.

Control analysis: Eye movements

Involuntary eye movements are a common problematic covariate in studies using multivariate approaches[32,33]. To test for an influence of gaze on the MEG findings, we ran the same MEG decoding analyses on our eye-tracking data. We acquired a total of 23 eye-tracking datasets, consisting of 15 binocular recordings and 8 monocular recordings. For the binocular recordings, we merged the x- and y-coordinates from both eyes to create a single metric. The data were then epoched from -500 ms to 3000 ms relative to the first cue onset and were baseline-corrected prior to cue onset (-500 to 0 ms). Next, we trained the classifier to distinguish between different conditions at each timepoint using only the x- and y-coordinates of the eye-tracking data. Instead of a LDA classifier, we used a Support Vector Machine classifier (SVC; *sklearn.svm.SVC*) with Radial Basis Function (RBF) kernel. The reason behind opting for SVC with RBF kernel is the likelihood that the distribution of x- and y-coordinates of gaze is not linearly separable. In all conditions, participants were instructed to maintain fixation, making it challenging to distinguish between different classes or conditions using a simple linear boundary. The RBF kernel is a powerful tool for transforming non-linearly separable data into a higher-dimensional space, making it possible to separate the data using a hyperplane [72]. This method is particularly effective when dealing with datasets a smaller number of features than of samples.

By computing mutual information scores over correctly classified MEG trials on a timepoint-by-timepoint basis, we assessed how much information was shared between the MEG decoding results and eye movement decoding results[33]. Mutual information quantifies the amount of information obtained about one variable by observing the other variable. Higher mutual information values indicate a stronger dependence between two variables, suggesting that changes in one variable are informative about the other.

Time frequency analysis

We generated time-frequency representations of power by convolving the data with Morlet wavelets across the frequency range of 3 to 40 Hz. For each frequency, we used a fixed 300-ms time window such that the number of cycles changed with the frequency. The power time series in the planar gradiometer pairs were then combined (i.e., root mean square), resulting in a 102-sensor combined planar gradiometer map in sensor space. For the comparison between contralateral and ipsilateral power, we computed activity following left- and right-directing cues, separately for left sensor clusters (MEG1632+1633, MEG1642+1643, MEG1912+1913, MEG1922+1923, MEG1942+1943, MEG2042+2043) and right sensor clusters (MEG2032+2033, MEG2312+2313, MEG2322+2323, MEG2342+2343, MEG2432+2433, MEG2442+2443), and subsequently pooled the contralateral vs. ipsilateral contrasts between them. We expressed the contra-versus-ipsi contrast as a normalised difference [i.e., ((contra-ipsi)/(contra + ipsi)) × 100]. Sensors were chosen based on prior MEG studies from our lab that have consistently implicated the same set of posterior sensors for capturing neural activity related to lateralised visual stimuli[73,74]. While we selected these sensor clusters based on independent data, we could confirm their validity and appropriateness in the current data (**Fig. 6j**, **k**, **l** overlay of these sensor clusters on the relevant left vs. right topographies).

To extract time courses of alpha lateralisation, we averaged the contra- vs. ipsilateral activity across the predefined alpha band (8 to 12 Hz) in the visual sensors. Topographies of the lateralised visual activity were obtained by contrasting trials in which cues indicated visual content in the left vs. right hemifield. To depict the topographies associated with relevant condition comparisons, we focused on the alpha frequency bands between 0 and 3000 ms in predefined steps of 500 ms.

Statistical analysis

Differences between conditions across the time courses were investigated by means of cluster-based non-parametric permutation testing using 1024 permutations with a cluster alpha of 0.05. This approach circumvents the multiple-comparisons problem by evaluating temporal clusters observed in the original (non-permuted) data against a single permutation distribution of the largest temporal cluster found after each permutation of the condition labels[75].

In a supplementary analysis, we additionally utilised a jack-knife approach to identify differences in decoding-onset times. For the jack-knife approach, we computed decoding time-course averages by excluding each of the *n* data sets once, allowing us to obtain *n* leave-one-out grand averages. Next, we identified the time point at which the decoding time course reached 50% of its peak. Jack-knifed standard errors, *t*-values, and *p*-values were adjusted according to [76].

CRediT authorship contribution statement

Daniela Gresch: Conceptualization, Methodology, Software, Investigation, Formal analysis, Visualization, Writing – original draft. Sage E.P. Boettcher: Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Supervision. Freek van Ede: Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Supervision. Anna C. Nobre: Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Supervision.

Acknowledgements

This research was supported by the Experimental Psychology Society Postdoctoral Fellowship awarded to S.E.P.B., an ERC Starting Grant from the European Research Council (MEMTICIPATION, 850636) and a NWO Vidi Grant from the Dutch Research Council (14721) awarded to F.v.E., a Wellcome Trust Senior Investigator Award (104571/Z/14/Z) and a James S. McDonnell Foundation Understanding Human Cognition Collaborative Award (220020448) awarded to A.C.N., and by the NIHR Oxford Health Biomedical Research Centre. The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust (203139/Z/16/Z). For the purpose of open access, the authors have applied a CC-BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

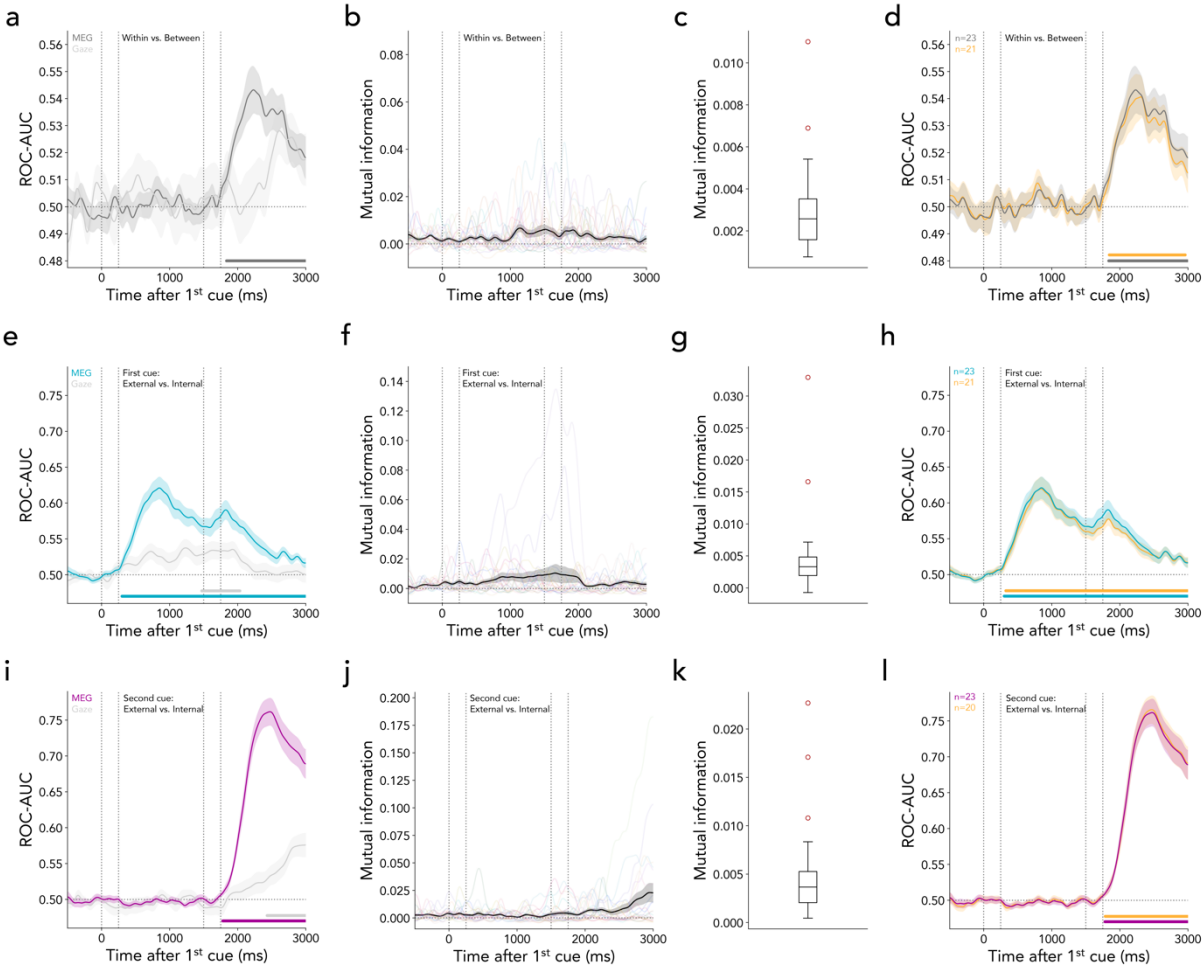
1. Chun, M. M., Golomb, J. D. & Turk-Browne, N. B. A taxonomy of external and internal attention. *Annu Rev Psychol* **62**, 73–101 (2011).
2. Carrasco, M. Visual attention: The past 25 years. *Vision Research* **51**, 1484–1525 (2011).
3. Posner, M. I. Orienting of Attention. *Quarterly Journal of Experimental Psychology* **32**, 3–25 (1980).
4. Griffin, I. C. & Nobre, A. C. Orienting attention to locations in internal representations. *J Cogn Neurosci* **15**, 1176–1194 (2003).
5. Landman, R., Spekreijse, H. & Lamme, V. A. F. Large capacity storage of integrated objects before change blindness. *Vision Res* **43**, 149–164 (2003).
6. Souza, A. S. & Oberauer, K. In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Atten Percept Psychophys* **78**, 1839–1860 (2016).
7. Chota, S., Gayet, S., Kenemans, J. L., Olivers, C. N. L. & Van der Stigchel, S. A matter of availability: sharper tuning for memorized than for perceived stimulus features. *Cereb Cortex* bhad064 (2023) doi:10.1093/cercor/bhad064.
8. Myers, N. E., Walther, L., Wallis, G., Stokes, M. G. & Nobre, A. C. Temporal dynamics of attention during encoding versus maintenance of working memory: complementary views from event-related potentials and alpha-band oscillations. *J Cogn Neurosci* **27**, 492–508 (2015).
9. Nobre, A. C. *et al.* Orienting Attention to Locations in Perceptual Versus Mental Representations. *Journal of Cognitive Neuroscience* **16**, 363–373 (2004).
10. Panichello, M. F. & Buschman, T. J. Shared mechanisms underlie the control of working memory and attention. *Nature* **592**, 601–605 (2021).
11. Tamber-Rosenau, B. J., Esterman, M., Chiu, Y.-C. & Yantis, S. Cortical Mechanisms of Cognitive Control for Shifting Attention in Vision and Working Memory. *J Cogn Neurosci* **23**, 2905–2919 (2011).
12. Tanigawa, H. *et al.* Decoding distributed oscillatory signals driven by memory and perception in the prefrontal cortex. *Cell Reports* **39**, 110676 (2022).
13. Teichmann, L., Moerel, D., Rich, A. N. & Baker, C. I. The nature of neural object representations during dynamic occlusion. *Cortex* **153**, 66–86 (2022).
14. Wallis, G., Stokes, M., Cousijn, H., Woolrich, M. & Nobre, A. C. Frontoparietal and Cingulo-opercular Networks Play Dissociable Roles in Control of Working Memory. *Journal of Cognitive Neuroscience* **27**, 2019–2034 (2015).
15. Zhou, Y., Curtis, C. E., Sreenivasan, K. K. & Fougnie, D. Common Neural Mechanisms Control Attention and Working Memory. *J. Neurosci.* **42**, 7110–7120 (2022).
16. Carlson, R. A., Wenger, J. L. & Sullivan, M. A. Coordinating information from perception and working memory. *J Exp Psychol Hum Percept Perform* **19**, 531–548 (1993).
17. Dark, V. J. Switching between memory and perception: Moving attention or memory retrieval? *Memory & Cognition* **18**, 119–127 (1990).
18. Gresch, D., Boettcher, S. E. P., van Ede, F. & Nobre, A. C. Shifting attention between perception and working memory. *Cognition* **245**, 105731 (2024).
19. Hautekiet, C., Verschooren, S., Langerock, N. & Vergauwe, E. Attentional switching between perception and memory: Examining asymmetrical switch costs. *Atten Percept Psychophys* (2023) doi:10.3758/s13414-023-02665-w.
20. Verschooren, S., Liefooghe, B., Brass, M. & Pourtois, G. Attentional flexibility is imbalanced: Asymmetric cost for switches between external and internal attention. *Journal of Experimental Psychology: Human Perception and Performance* **45**, 1399–1414 (2019).
21. Verschooren, S., Pourtois, G. & Egner, T. More efficient shielding for internal than external attention? Evidence from asymmetrical switch costs. *Journal of Experimental Psychology: Human Perception and Performance* **46**, 912–925 (2020).
22. Weber, R. J., Burt, D. B. & Noll, N. C. Attention switching between perception and memory. *Memory & Cognition* **14**, 238–245 (1986).

23. Burgess, P. W., Dumontheil, I. & Gilbert, S. J. The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends Cogn Sci* **11**, 290–298 (2007).
24. Gilbert, S. J., Frith, C. D. & Burgess, P. W. Involvement of rostral prefrontal cortex in selection between stimulus-oriented and stimulus-independent thought. *European Journal of Neuroscience* **21**, 1423–1431 (2005).
25. Poskanzer, C. & Aly, M. Switching between External and Internal Attention in Hippocampal Networks. *J. Neurosci.* **43**, 6538–6552 (2023).
26. Treder, M. S. *et al.* The hippocampus as the switchboard between perception and memory. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2114171118 (2021).
27. Sridharan, D., Levitin, D. J. & Menon, V. A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proc Natl Acad Sci U S A* **105**, 12569–12574 (2008).
28. Tang, Y.-Y., Rothbart, M. K. & Posner, M. I. Neural correlates of establishing, maintaining, and switching brain states. *Trends Cogn Sci* **16**, 330–337 (2012).
29. Boettcher, S. E. P., Gresch, D., Nobre, A. C. & van Ede, F. Output planning at the input stage in visual working memory. *Sci Adv* **7**, eabe8212 (2021).
30. van Ede, F., Chekroud, S. R., Stokes, M. G. & Nobre, A. C. Concurrent visual and motor selection during visual working memory guided action. *Nat Neurosci* **22**, 477–483 (2019).
31. Thölke, P. *et al.* Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *NeuroImage* **277**, 120253 (2023).
32. Mostert, P. *et al.* Eye Movement-Related Confounds in Neural Decoding of Visual Working Memory Representations. *eNeuro* **5**, (2018).
33. Quax, S. C., Dijkstra, N., van Staveren, M. J., Bosch, S. E. & van Gerven, M. A. J. Eye movements explain decodability during perception and cued attention in MEG. *NeuroImage* **195**, 444–453 (2019).
34. Grootswagers, T., Wardle, S. G. & Carlson, T. A. Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *J Cogn Neurosci* **29**, 677–697 (2017).
35. King, J.-R. & Dehaene, S. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn Sci* **18**, 203–210 (2014).
36. Wylie, G. & Allport, A. Task switching and the measurement of "switch costs". *Psychological Research Psychologische Forschung* **63**, 212–233 (2000).
37. Yeung, N., Nystrom, L. E., Aronson, J. A. & Cohen, J. D. Between-Task Competition and Cognitive Control in Task Switching. *J. Neurosci.* **26**, 1429–1438 (2006).
38. Gould, I. C., Rushworth, M. F. & Nobre, A. C. Indexing the graded allocation of visuospatial attention using anticipatory alpha oscillations. *J Neurophysiol* **105**, 1318–1326 (2011).
39. Jensen, O. & Mazaheri, A. Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition. *Frontiers in Human Neuroscience* **4**, (2010).
40. Thut, G., Nietzel, A., Brandt, S. A. & Pascual-Leone, A. Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *J Neurosci* **26**, 9494–9502 (2006).
41. van Ede, F. Mnemonic and attentional roles for states of attenuated alpha oscillations in perceptual working memory: a review. *Eur J Neurosci* **48**, 2509–2515 (2018).
42. Worden, M. S., Foxe, J. J., Wang, N. & Simpson, G. V. Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. *J Neurosci* **20**, RC63 (2000).
43. Mok, R. M., Myers, N. E., Wallis, G. & Nobre, A. C. Behavioral and Neural Markers of Flexible Attention over Working Memory in Aging. *Cerebral Cortex* **26**, 1831–1842 (2016).

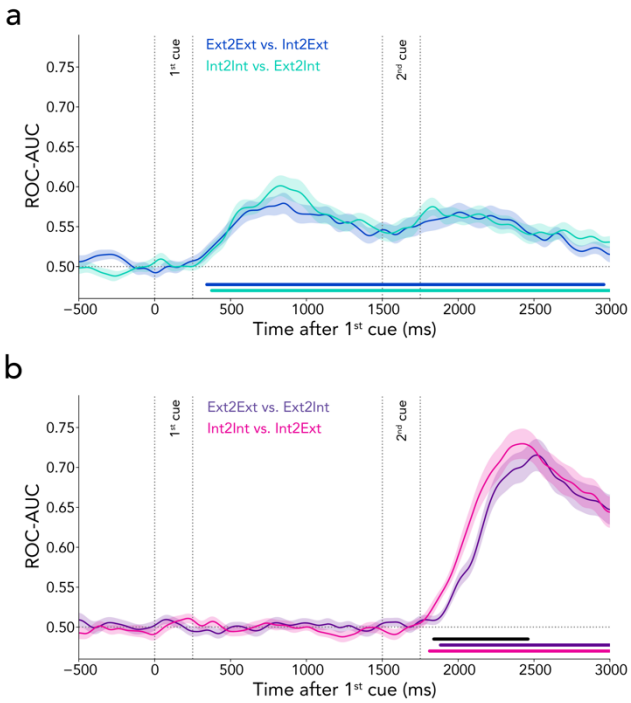
44. Burgess, P. W., Gilbert, S. J., Okuda, J. & Simons, J. S. Rostral prefrontal brain regions (area 10): A gateway between inner thought and the external world? in *Disorders of volition* 373–395 (MIT Press, Cambridge, MA, US, 2006). doi:10.7551/mitpress/2457.001.0001.
45. Verschooren, S., Schindler, S., De Raedt, R. & Pourtois, G. Early reduction of sensory processing within the visual cortex when switching from internal to external attention. *Biological Psychology* **163**, 108119 (2021).
46. Myers, N. E., Stokes, M. G. & Nobre, A. C. Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. *Trends in Cognitive Sciences* **21**, 449–461 (2017).
47. van Ede, F. & Nobre, A. C. Turning Attention Inside Out: How Working Memory Serves Behavior. *Annual Review of Psychology* **74**, 137–165 (2023).
48. Calzolari, S., Boneva, S. & Fernández-Espejo, D. Investigating the shift between externally and internally oriented cognition: a novel task-switching paradigm. *Neuroscience of Consciousness* **2022**, niac016 (2022).
49. Nobre, A. C. & Stokes, M. G. Premembering Experience: A Hierarchy of Time-Scales for Proactive Attention. *Neuron* **104**, 132–146 (2019).
50. Esterman, M., Chiu, Y.-C., Tamber-Rosenau, B. J. & Yantis, S. Decoding cognitive control in human parietal cortex. *Proc Natl Acad Sci U S A* **106**, 17974–17979 (2009).
51. Vortmann, L.-M., Kroll, F. & Putze, F. EEG-Based Classification of Internally- and Externally-Directed Attention in an Augmented Reality Paradigm. *Front Hum Neurosci* **13**, 348 (2019).
52. Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R. & Ungerleider, L. G. Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron* **22**, 751–761 (1999).
53. Silver, M. A., Ress, D. & Heeger, D. J. Neural correlates of sustained spatial attention in human early visual cortex. *J Neurophysiol* **97**, 229–237 (2007).
54. Stokes, M. G., Thompson, R., Nobre, A. C. & Duncan, J. Shape-specific preparatory activity mediates attention to targets in human visual cortex. *Proceedings of the National Academy of Sciences* **106**, 19569–19574 (2009).
55. Stokes, M. G. 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends in Cognitive Sciences* **19**, 394–405 (2015).
56. Wolff, M. J., Ding, J., Myers, N. & Stokes, M. Revealing hidden states in visual working memory using electroencephalography. *Frontiers in Systems Neuroscience* **9**, (2015).
57. Wolff, M. J., Jochim, J., Akyürek, E. G. & Stokes, M. G. Dynamic hidden states underlying working-memory-guided behavior. *Nat Neurosci* **20**, 864–871 (2017).
58. Schneider, D., Barth, A. & Wascher, E. On the contribution of motor planning to the retroactive cuing benefit in working memory: Evidence by mu and beta oscillatory activity in the EEG. *NeuroImage* **162**, 73–85 (2017).
59. Peirce, J. *et al.* PsychoPy2: Experiments in behavior made easy. *Behav Res* **51**, 195–203 (2019).
60. Posit team. *RStudio: Integrated Development Environment for R*. (Posit Software, PBC, Boston, MA, 2022).
61. Gramfort, A. *et al.* MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience* **7**, (2013).
62. Quinn, A. J., van Es, M. W. J., Gohil, C. & Woolrich, M. W. OHBA Software Library in Python (OSL). Zenodo https://doi.org/10.5281/ZENODO.6875060 (2022).
63. Taulu, S. & Simola, J. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Phys. Med. Biol.* **51**, 1759 (2006).
64. van Driel, J., Olivers, C. N. L. & Fahrenfort, J. J. High-pass filtering artifacts in multivariate classification of neural time series data. *Journal of Neuroscience Methods* **352**, 109080 (2021).
65. Rosner, B. Percentage Points for a Generalized ESD Many-Outlier Procedure. *Technometrics* **25**, 165–172 (1983).
66. Woolrich, M., Hunt, L., Groves, A. & Barnes, G. MEG beamforming using Bayesian PCA for adaptive data covariance matrix regularization. *Neuroimage* **57**, 1466–1479 (2011).

67. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. Preprint at https://doi.org/10.48550/arXiv.1201.0490 (2018).
68. Hajonides, J. E., Nobre, A. C., van Ede, F. & Stokes, M. G. Decoding visual colour from scalp electroencephalography measurements. *NeuroImage* **237**, 118030 (2021).
69. Hand, D. J. & Till, R. J. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Machine Learning* **45**, 171–186 (2001).
70. Kaiser, D., Oosterhof, N. N. & Peelen, M. V. The Neural Dynamics of Attentional Selection in Natural Scenes. *J Neurosci* **36**, 10522–10528 (2016).
71. Sun, Y., Wong, A. K. C. & Kamel, M. S. Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence* **23**, 687–719 (2009).
72. Hsu, C.-W., Chang, C.-C. & Lin, C.-J. A Practical Guide to Support Vector Classification. (2003).
73. Heideman, S. G. *et al.* Anticipatory neural dynamics of spatial-temporal orienting of attention in younger and older adults. *NeuroImage* **178**, 46–56 (2018).
74. Heideman, S. G., van Ede, F. & Nobre, A. C. Temporal alignment of anticipatory motor cortical beta lateralisation in hidden visual-motor sequences. *European Journal of Neuroscience* **48**, 2684–2695 (2018).
75. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* **164**, 177–190 (2007).
76. Miller, J., Patterson, T. & Ulrich, R. Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology* **35**, 99–115 (1998).
77. Tukey, J. W. *Exploratory Data Analysis*. (Addison-Wesley, Reading, Mass., 1977).
78. Sassenhagen, J. & Draschkow, D. Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology* **56**, e13335 (2019).

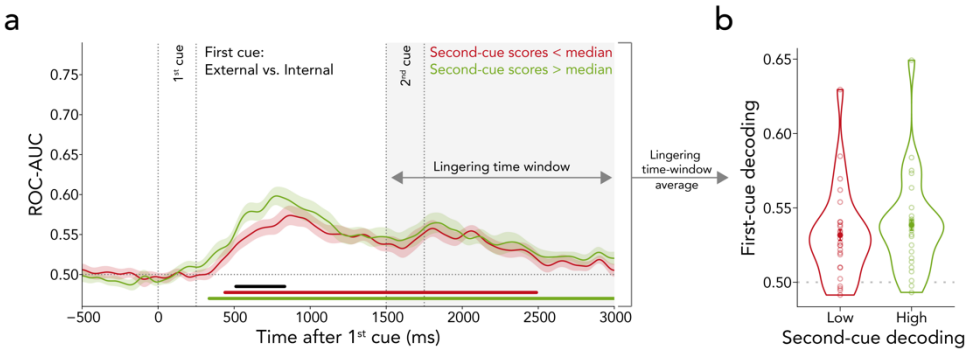
Supplementary material



Supplementary Figure 1. Eye movements do not account for decoding results. (a) Average time-resolved decoding of within- vs. between-domain shifts based on MEG and eye-tracking data. Only MEG data sets were included that had an accompanying eye tracking data set (n = 23). Horizontal lines indicate significant clusters. (MEG decoding: ~1840-3000 ms, cluster $p < 0.001$) **(b)** Average mutual information between MEG and eye-tracking data for the within- vs. between-shift decoding. Faded lines indicate individual participants **(c)** Box-and-whisker plot according to [77] displaying distribution of average mutual information across -500 to 3000 ms relative to first-cue onset. Red dots represent outliers. **(d)** Average time-resolved decoding of within- vs. between-domain shifts based on MEG after removing outliers (n=23: ~1840-3000 ms, cluster $p < 0.001$, n=21: ~1840-3000 ms, cluster $p < 0.001$). **(e)** Same as (a) but for external vs. internal first cues (MEG decoding: ~300-3000 ms, cluster $p < 0.001$, eye decoding: ~1470-2030 ms, cluster $p = 0.042$). **(f)** Same as (b) but for external vs. internal first cues. **(g)** Same as (c) but for external vs. internal first cues. **(h)** Same as (d) but for external vs. internal first cues (n=23: ~300-3000 ms, cluster $p < 0.001$, n=21: ~320-3000 ms, cluster $p < 0.001$). **(i)** Same as (a) but for external vs. internal second cues (MEG decoding: ~1780-3000 ms, cluster $p < 0.001$, eye decoding: ~2430-3000 ms, cluster $p = 0.010$). **(j)** Same as (b) but for external vs. internal second cues. **(k)** Same as (c) but for external vs. internal second cues. **(l)** Same as (d) but for external vs. internal second cues (n=23: ~1780-3000 ms, cluster $p < 0.001$, n=20: ~1780-3000 ms, cluster $p < 0.001$).



Supplementary Figure 2. Time-resolved MEG decoding of different trial types. (a) Average time-resolved classifier performance for distinguishing external-to-external from internal-to-external shifts, and internal-to-internal from external-to-internal shifts. Time courses show *M* ± *SEM* across participants. Cluster-permutation corrected significant time points are indicated with horizontal lines (external-to-external vs. internal-to-external cluster: ~340-2960 ms, cluster $p < 0.001$; internal-to-internal vs. internal-to-external cluster: ~380-3000 ms, cluster $p < 0.001$). **(b)** Same as (a) but decoding internal-to-internal vs. internal-to-external shifts, and external-to-external vs. external-to-internal shifts (external-to-external vs. external-to-internal cluster: ~1880-3000 ms, cluster $p < 0.001$; internal-to-internal vs. internal-to-external cluster: ~1810-3000 ms, cluster $p < 0.001$). Black horizontal line denotes significant difference between decoding time courses (~1840-2460 ms, cluster $p < 0.001$). Since statements regarding the exact onset of an effect are not supported by permutation tests[78], we quantified the timings of the decoding time courses more formally by applying a jack-knife approach to compare the point at which the decoding time courses reached 50% of their peak value. This revealed significantly earlier onset times for internal-to-internal vs. internal-to-external (*M* = 636.560, *SEM* = 5.200) decoding as compared to external-to-external vs. external-to-internal decoding (*M* = 658.480, *SEM* = 4.185; $t_{(24)} = -3.379$, $p = 0.002$).



Supplementary Figure 3. No evidence for a trade-off in decodability between first- and second-cued attentional domains. (a) We median-split the MEG data into trials with low vs. high decoding performance in distinguishing the second-cued domain following the second cue (i.e., 0-1500 ms after second-cue onset). Next, we trained a classifier separately on both data sets to predict the first-cued domain (external vs. internal). In trials with low and high second-cue decoding, we could successfully predict the first-cued attentional domain after first-cue appearance and the decodability remained significant following the second cue (low decoding: ~440-2490 ms, cluster $p < 0.001$; high decoding: ~340-3000 ms, cluster $p < 0.001$). We did not observe any trade-off in the decoding of the first-cued attentional state between low vs. high second-cue decoding following the second cue (i.e., grey area). There was an earlier difference in the decodability of external vs. internal first-cue trials when the decodability of the second cue was low vs. high (~510-830 ms, cluster $p = 0.012$). **(b)** Average first-cued domain decoding in the second-cued period (i.e. grey area in (a)) as a function of low vs. high second-cued domain decoding. There was no difference between low and high second-cue decoding in predicting the first-cued domain ($t_{(24)} = 1.348$, $p = 0.190$, $d = 0.270$).