Preparatory neural activity differs between Selective Attention and Perceptual Expectations

José M. G. Peñalver[1], David López-García[1], Carlos González-García[1], Blanca Aguado-López[1], Juan M. Górriz[2,3] and María Ruz[1]*

[1]Mind, Brain and Behavior Research Center (CIMCYC), University of Granada, Granada 18071, Spain
[2]Data Science & Computational Intelligence Institute, Universidad de Granada, CP 18071
[3]Cambridge Neuroscience, University of Cambridge, CB2 0SZ, UK

Correspondence: mruz@ugr.es

Summary

Proactive cognition brain models are mainstream nowadays. Within these, preparation is understood as an endogenous, top-down function that takes place prior to the actual perception of a stimulus and improves subsequent behavior. Neuroimaging has uncovered the existence of such preparatory activity in different cognitive domains, however no research to date has sought to uncover their potential similarities and differences. Two of these domains, often confounded in the literature, are Selective Attention (information relevance) and Perceptual Expectation (information probability). We used EEG to characterize the mechanisms that support preparation in Attention and Expectation. In different blocks, participants were cued to the relevance or to the probability of target categories, faces vs. names in a gender discrimination task. Multivariate Pattern (MVPA) and Representational Similarity Analyses (RSA) during the preparation window showed that both manipulations led to a significant, ramping-up prediction of the relevant or expected target category. However, classifiers trained on data from one condition did not generalize to the other, indicating the existence of different anticipatory neural patterns. Further analyses pointed to a differential involvement of specific frequency bands in each condition. Finally, a Canonical Template Tracking procedure showed that there was stronger perceptual reinstatement for Attention than for Expectation. Overall, results indicate that preparation during attention and expectation acts through different, although partially overlapping, neural mechanisms. These findings have important implications for current models of brain functioning, as they are a first step towards characterizing and dissociating the neural mechanisms involved in top-down anticipatory processing.

Introduction

For decades, research in Cognitive Psychology has studied behavior while manipulating external factors, which led to theoretical models that framed cognition mostly from a reactive point of view. Recent years have witnessed a renaissance of proactive cognition, where endogenous top-down mechanisms play a core role in brain functioning. Within this framework, preparation is conceptualized as an endogenous function that takes place in anticipation of incoming inputs or demands and improves subsequent behavior[1,2]. Neural preparatory activity has been shown for a plethora of processes, including attention[1,3,4], expectation[5,6], working memory[7,8], or cognitive control[2,9,10]. Similarly, influential models have proposed different ways in which anticipatory patterns interact with stimulus inputs to guide perception. Examples of this are the Predictive Coding [11] or the Biased Competition frameworks[12]. However, currently these models are silent as to how different preparatory phenomena relate to each other, and whether they reflect common or diverging underlying top-down mechanisms.

Selective Attention and Perceptual Expectation are complex functions that involve top-down and bottom-up elements. Attention refers to the selection of *relevant* information based on specific goals[3], while expectation involves predictions based on prior *probability*[13]. Studies of selective attention have manipulated the relevance of information using cues that indicate the stimulus or dimensions to respond while ignoring others[1,3,14,15]. Expectation has primarily (but not only, see e.g.[16]) been manipulated with cues that inform about the most probable stimulus[6,17,18]. Previous research focusing on the effect of Attention and Expectations on target processing has shown both diverging and overlapping results. Although both lead to behavioral improvements[19,20], several neuroimaging studies reveal specific oscillatory activity in each context. Selective attention investigations have shown the relevance of Theta (4-7 Hz) on environmental sampling and coding[21–23], Alpha (8-12 Hz) on both target-enhanced representations and inhibition of distractors[24], and have associated Beta (13-30 Hz) and Gamma (>30 Hz) with top-down[25] and bottom-up[26] signal coordination, respectively. On the other hand, predictive coding models[27] have linked top-down perceptual expectations to the Alpha and Beta bands [28,29] and bottom-up prediction error signaling with Gamma[30].

Studies that have tried to overcome frequent confounds between attention and expectation (e.g.[31,32]) have disentangled the separate roles of relevance and probability during target processing[33–36]. On the other hand, research focused on their preparatory activity separately has provided overlapping results. Cues indicating relevance in selective attention[1,3] preactivate relevant regions of space processing [37], specific shape patterns in visual cortex[14], or patterns in category[38,39] and object-selective perceptual regions[40,41]. Similarly, cues providing probabilistic information lead to the preactivation of specific templates of oriented gabors[6], direction[42], motor patterns[18] or abstract shapes[43]. However, all these previous investigations are agnostic regarding the potential similarities or differences in such top-down preparation across relevance and probability anticipation. Unraveling the differences in how anticipatory activity in different contexts reflects the upcoming information is a necessary step to understand the differences between Attention and Expectation, and is also essential for theoretical models that explain the neural basis of these two phenomena[12,27,44].

In our study, we employed different multivariate pattern analyses of EEG data to directly compare top-down preparation in Selective Attention (relevance) and Expectation (probability). To do so, we embedded a sex/gender discrimination in a cue-target paradigm. Here, depending on the block cues provided information about future relevance or probability[34,45] of face or name stimulus categories. In addition, we ran an independent localizer to study similarities between preparation and perception across contexts. Our first goal was to study the anticipatory mechanisms during preparation by means of time-resolved representational similarity analysis (RSA[46]). Then, we used multivariate pattern analysis (MVPA[47]) to examine whether anticipated stimulus categories are represented with differential fidelity during selected compared to probable targets. Next, we used a cross-classification approach to directly contrast the patterns of activity underlying the representation of relevant vs. probable stimuli. Taking into consideration the differences observed during target processing in these contexts[17,48,49], our hypothesis was that anticipated relevant vs. probable stimuli would be differently coded. More

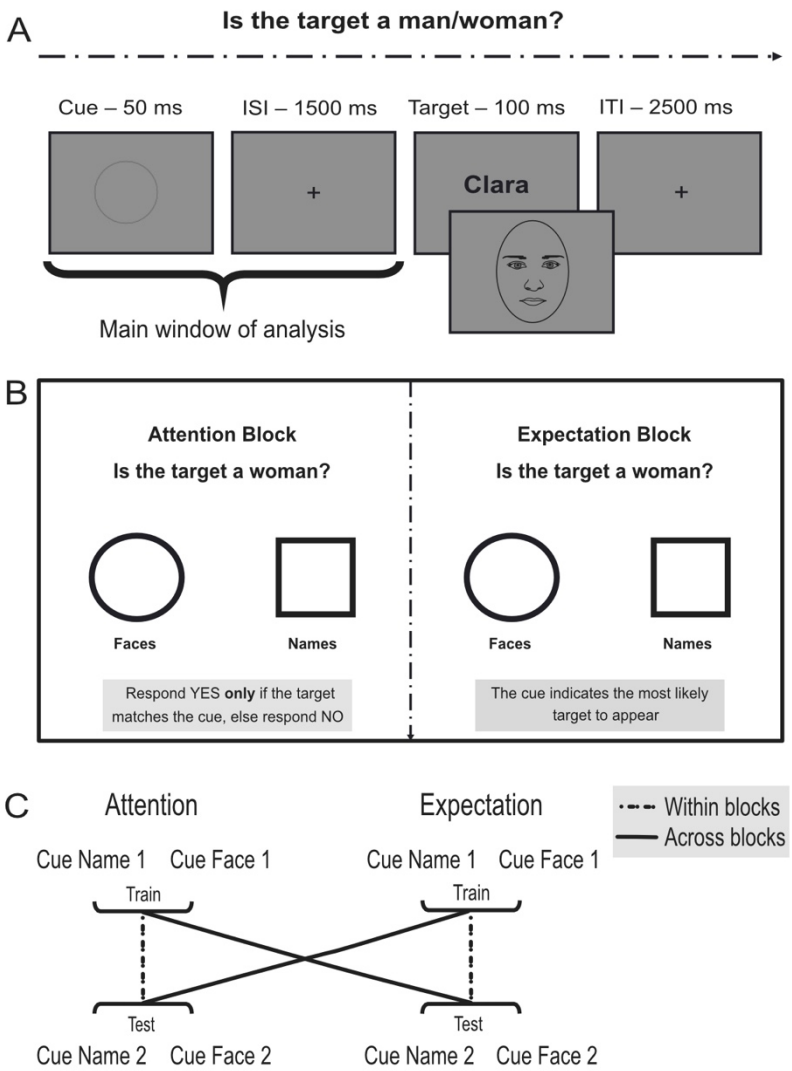

Figure 1. Behavioral task and design (A) Behavioral task: example trial. Participants were cued about an incoming target stimulus (a face or a name) with which they performed a sex/gender classification task. **(B)** Block starting screen. In Attention blocks, the cue indicated the relevant stimulus. Participants should only perform the task (respond to the question on top of the screen) if the stimulus matched the cue. In Attention blocks the cues were not predictive of the probability of the target (50% faces vs. names). In Expectation blocks the cue indicated the probability of the stimulus category, which appeared 75% of times. Participants had to perform the task regardless of the cue (thus predicted and unpredicted targets were equally task-relevant). **(C)** MVPA classification rationale. In cue-locked within-block analyses, the classifier was trained with one pair of cues and then tested on the other, in either attention or expectation blocks. Across-blocks, the classifier was trained with one pair of cues in one condition (e.g., attention) and then tested on the other pair of cues of the other condition (e.g., expectation).

specifically, we hypothesized that specific oscillatory patterns would induce differences in the classifiers' performance. Finally, we leveraged a Canonical Template Tracking approach[50,51] to observe the extent to which preparation induces the reinstatement of specific perceptual information in each condition.

Results

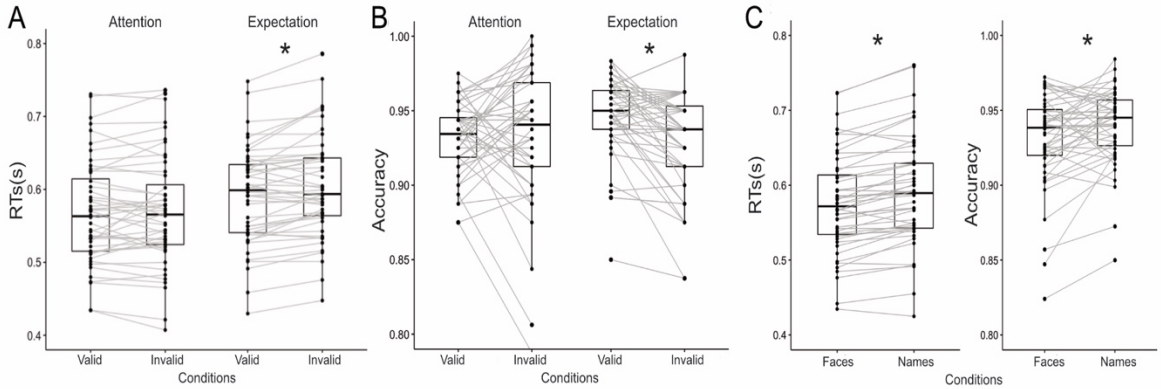

Figure 2. Behavioral results. (A) Reaction times (in milliseconds) in Attention and Expectation blocks, for cued and uncued trials. (B) Accuracy in Attention and Expectation blocks, for cued and uncued trials. (C) RTs and accuracy for face and name targets, averaged across conditions. Dots represent individual subject (n=48) scores per experimental condition. Grey lines connect each participant's score in the two conditions of each block. The line inside boxes represents the median, while the limits of the box indicate the first and third quartile. Whiskers indicate the 1.5 inter quartile range for the upper and lower quartiles.

1. Behavioral

Forty-eight participants completed a cue-target paradigm where, depending on the block, cues carried information about either the *relevance* (Attention) or the *probability* (Expectation) of upcoming face or word targets (Figure 1A). Analysis of participants' behavioral results showed that preparation for the incoming target stimuli affected performance. A three-way repeated measures ANOVA on reaction times (RTs) showed main effects of Block ($F_{47,1}$=56.45, p<0.001, ηp2 = 0.54), Cueing ($F_{47,1}$=5.59, p=0.022, ηp2 = 0.11) and Category ($F_{47,1}$=50.52, p<0.001, ηp2 = 0.52). Overall, responses were faster in Attention (M = 569 ms, SD = 0.01) than Expectation (M = 598 ms, SD = 0.01) blocks. Cueing affected RTs differently depending on the block (Block*Cueing, $F_{47,1}$=9.07, p=0.004, ηp2 = 0.16; Figure 2A). Post-hoc tests showed no effect of Cueing for Attention (t<1) whereas there was an effect for Expectation ($t_{47,1}$=3.6, p<0.001, Cohen's d = 0.52). Expected (cued) targets elicited faster responses (M = 592 ms, SD = 0.06) than unexpected (uncued) targets (M = 603 ms, SD = 0.07). Stimulus Category also induced significant differences. Post-hoc tests showed faster responses for faces (M = 576 ms, SD = 0.7) than for names (M = 591 ms, SD = 0.75), $t_{47,1}$=7.51, p<0.001, Cohen's d = 1.08; Figure 2C.

The ANOVA on behavioral accuracy showed results in the same direction albeit less prominent. Both Cueing ($F_{47,1}$=4.33, p=0.043, ηp2 = 0.01) and Category ($F_{47,1}$=7.04, p=0.011, ηp2 = 0.011) were

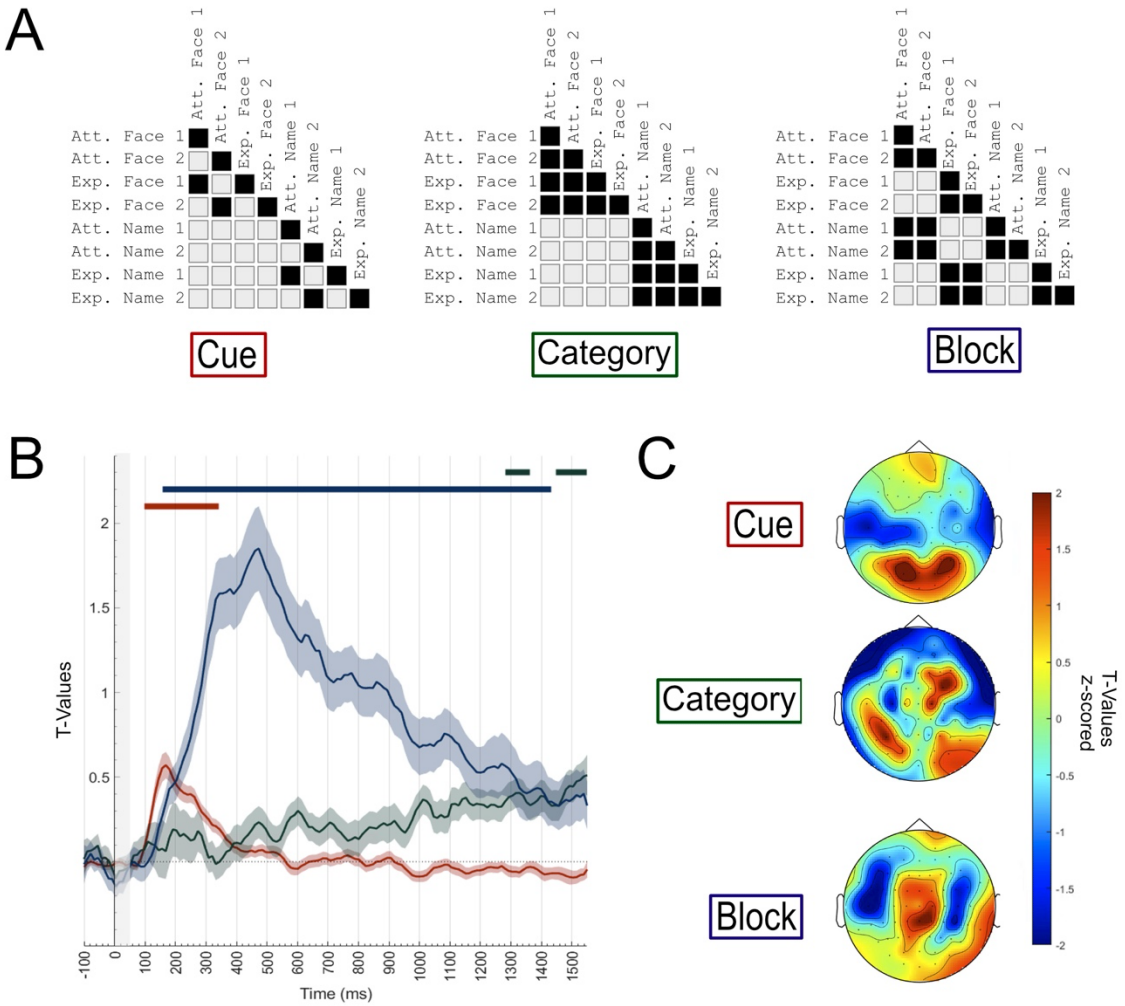

Figure 3. (A) RSA theoretical matrices and (B) results. In (A), black squares represent high similarity between conditions and grey squares represent low similarity. In (B), each line depicts t-values from a linear regression fit for the three theoretical dissimilarity models. The colored straight lines above indicate significance clusters after a cluster-based permutation analysis. Grey shading indicates cue presence onscreen. The cue model (red) indicates dissimilarity between similar cue shapes, regardless of the block or prediction; the category model (green) indicates similarity between cues predicting the same stimulus category (faces or names); the Block model (blue) shows preparatory differences between Attention and Expectation. (C) Topographies of each model after using time points as features and repeating the analysis on each EEG channel.

significant, and there was no main effect of Block ($p>0.05$), with both conditions presenting high accuracy overall (Attention: M = 0.92, SD = 0.07; Expectation: M = 0.93, SD = 0.07). Again, we found that cueing affected each block differently (Block*Validity, $F_{47,1}=9.67$, $p=0.003$, $\eta p2 = 0.15$; Figure 2B). Although there were no differences in Attention ($t<1$), there was an effect of expectations ($t_{47,1}=3.59$, $p=0.003$, Cohen's d = 0.52), with cued trials eliciting more accurate responses (M = 0.947, SD = 0.026) than uncued ones (0.929, SD = 0.035). Post-hoc tests showed more accurate responses for less accurate responses to faces (M = 0.93, SD = 0.4) than to names (M = 0.94, SD = 0.37; Figure 2C), $t_{47,1}=2.65$, $p=0.011$, Cohen's d = 0.38; Figure 2C. Overall, these results indicate that, as instructed, cues were used effectively and also differently across blocks.

2. Time-resolved profile of preparation

Our first aim was to assess the emergence of specific coding patterns linked to different information content during preparation in Attention vs. Expectation contexts. The perceptual features of the cue should be processed, and their contextual meaning extracted to anticipate general target-category information. This anticipated target category should be activated and maintained in working memory to later provide an efficient response. To assess the contribution of each of these sources of information to preparatory activity we employed model-based RSA [46].

Results for the three RSA models built (Cue shape, Category and Block, see Methods section and Figure 3A) revealed first a peak of the Cue model 160 ms after cue presentation and a fast decay afterwards (Figure 3B, red line). In contrast, the coding of the specific incoming target category increased progressively along the interval, reaching its peak right before the presentation of the actual target (Figure 3B, green line). In addition, block information was significantly active after cue onset and during most of the preparation window (Figure 3B., blue line).

3. Time-resolved classification shows increasing AUC throughout the preparation interval.

We then studied the coding of the anticipated stimulus categories during the preparation interval. Using the two cues of each category together for training and testing returned a classification weighted on the perceptual features of the cues (see Figure S5 for a detailed description of the result), similar to the Cue model in Figure 3 (red line). Afterwards, we employed a cross-classification analysis between different cues to avoid cue perceptual confounds in the classification (Figure 4B). Since two differently shaped cues coded for each type of category, we trained the classifier in one pair and tested it on the other, repeated the process in the opposite direction and averaged both. Similar to the Category model in Figure 3, time resolved cross-classification showed that the accuracy in the decoding of the anticipated category increased as the target onset approached. We employed the same subtraction method as before, but again, we found no differences in classification between Attention and Expectation (all ps>0.05). We repeated a regression analysis to predict the cross-

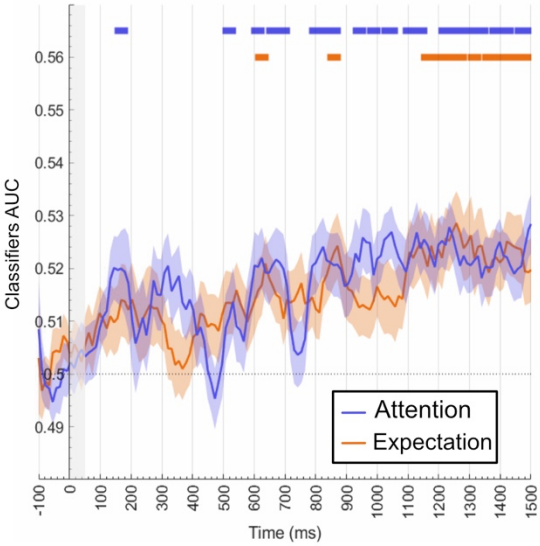

Figure 4. Raw voltage time-resolved cue decoding results. Result of the time-resolved classification of the category (faces vs. names) to be selected (blue) vs. expected (orange). Horizontal colored lines indicate statistical significance against chance within each block. Grey shading indicates cue presence onscreen. Figure shows the result of training and testing as described in Figure 1c (dotted line). For the results using all the cues together for training and testing see Figure S5.

classification result of each participant based on each RSA model. As expected, the category model (Figure 3, M = 0.35, SD = 2.37) explained the results better than any other. T-values were nominally higher than the cue model (M = 1.04, SD = 1.42) and significantly higher than the block model ($p<0.01$).

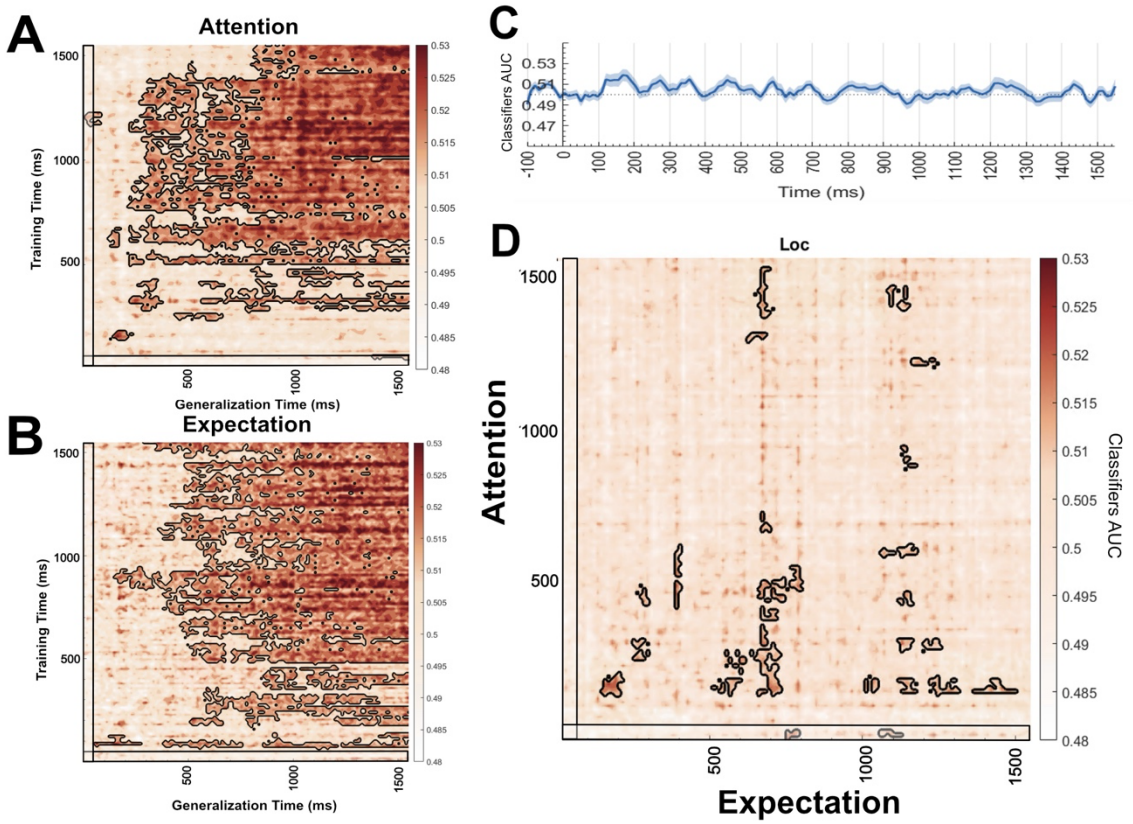

Figure 5. Raw voltage temporal generalization results. (**A-B**) Temporal generalization results using the cross-cue decoding scheme, within (**A**) Attention and (**B**) Expectation conditions. Black outlines indicate statistical significance against chance. Black boxes indicate cue presence on screen. **(C-D)** Cross-condition classification results. Classification using the scheme described in Fig 1C, depicting the average of the two train and test directions. **(C)** Time-resolved classification after training and testing the classifier in different conditions yielded no significant decoding. **(D)** Temporal generalization matrix. For visualization purposes, we averaged training and testing in the two directions (train attention and train expectation). The horizontal axis shows the times for training and testing in Expectation, and the vertical axis that of training and testing in Attention. That is, we inverted the axis of one of the directions before averaging. We chose this rationale because activity that is better represented in only one of the conditions should appear only on one side of the diagonal (e.g., above) when training and the opposite when testing (e.g., below). Averaging this way, we can show information that would not be visible doing a standard average. The non-averaged results for both directions can be found in the Supplementary Section (Figure S1).

4. Relevance- and probability-driven preactivations are stable

While the previous results provide an initial characterization of neural coding of specific category information during the preparation interval, they do not allow to explore the extent to which relevant representations are stable during this time window, since significant decoding on different time points could be driven in principle by different mechanisms. To investigate this, we employed a cross-time

decoding approach (King & Dehaene, 2014) to compare different patterns of brain activity across the preparation interval. For this, we trained a classifier in one time point and then tested it on all the points of the interval (Figure 5A, B). Results showed clear signs of generalization during the preparation interval in both Attention and Expectation conditions (black outlines). The clusters of activity grew increasingly larger up until the target's onset, indicating that the underlying patterns remain relatively stable during the preparation period. We then compared the results of both analyses following the same rationale as above: we subtracted the result matrices for both conditions and performed a one-side T-test against 0. This analysis did not yield any significant results, indicating that the accuracy of anticipatory category decoding was not different between conditions. Altogether, these results suggest that preparation in both Attention and Expectation leads to a similar level of discriminability of the anticipated (relevant or probable) category. However, this analysis is agnostic regarding potential similarities in how anticipated relevant vs. probable information is coded, as different underlying mechanisms could lead to similar accuracy results.

5. Attention and Expectation induce distinct patterns of preparatory activity

To examine the mechanisms supporting the classification results, we used multivariate cross-classification by training the classifier in one condition and then testing it on the other. Similar patterns of brain activity should generalize between conditions, while differences should provide chance-level classification. The number of observations employed for this, as well as the inter-block temporal distance, were also matched. The averaged results (training in attention and training in expectation; see Figure S1 for results split by block) for raw voltage cross-classification are shown in Figures 5C, D (see Figure S1 for the results split by train and test direction). Common coding between anticipating relevance vs. probability of stimulus categories was scarce. Significant small clusters appear scattered through the temporal generalization interval. This result complements the previous analysis by suggesting that different neural mechanisms support the classification results. It is unlikely that this result was due to a block effect (training and testing in different blocks), since the majority of the cues used for training in within condition classification already belonged to different blocks. However, we repeated the analysis using only pairs of cues from different blocks. The results (see Figure S2) matched the within-condition analyses (although with reduced power, likely due to the loss of trials).

6. Oscillatory activity in Attention and Expectation

We have shown that although Attention and Expectation lead to similar degrees of anticipatory classification, their underlying neural patterns are partially different. We hypothesized that such differences could arise through changes in the oscillatory profile of representations. To test this, we performed MVPA and cross-classification with time generalization as before, but now using average time-resolved power values in each of four discrete frequency bands: Theta (4-8 Hz), Alpha (8-12 Hz), Beta (12-30 Hz) and Gamma (30-49 Hz).

6.1. Theta

255

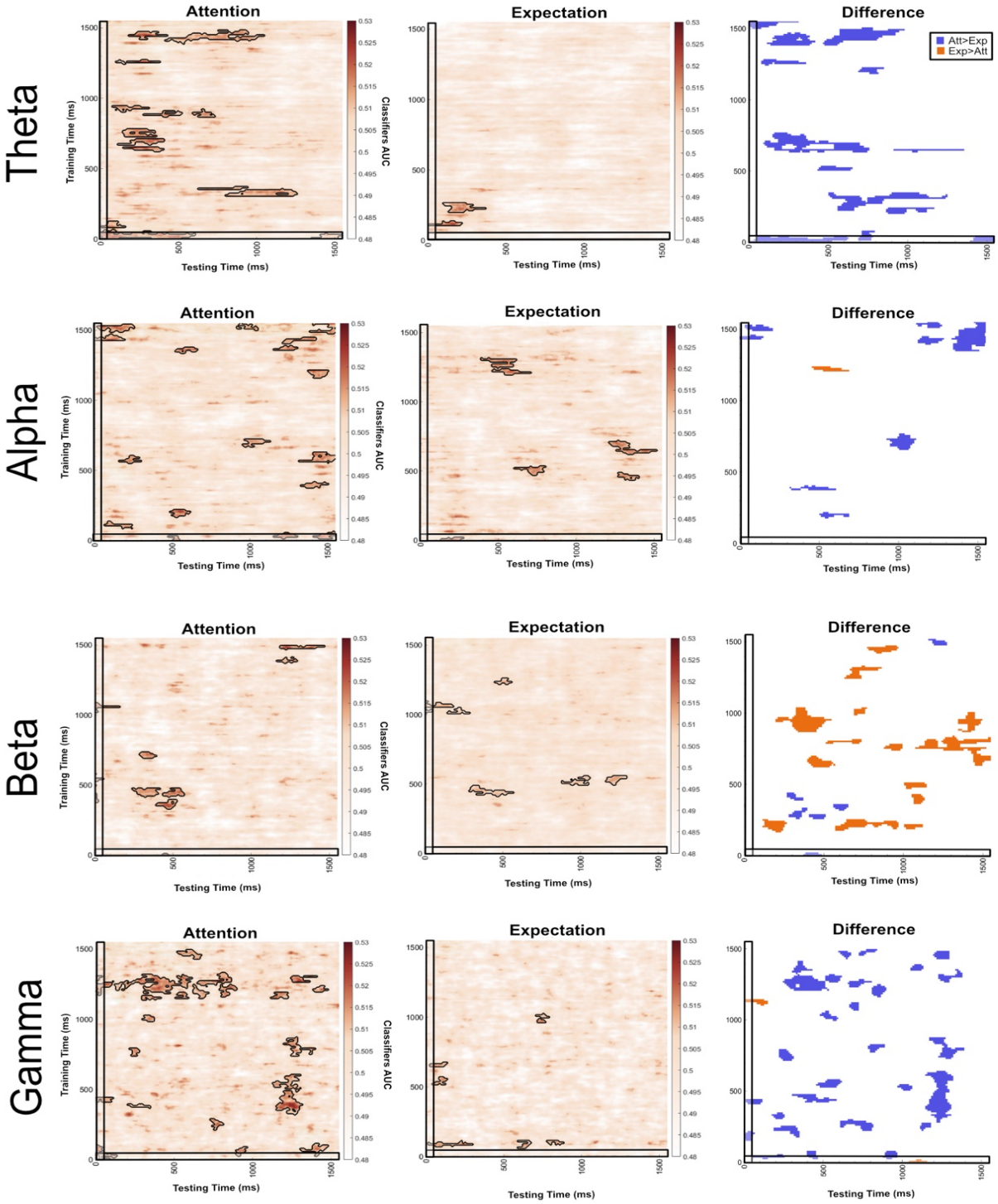

Figure 6. Oscillatory-based classification results within conditions. We used average power values for each of four discrete frequency bands: Theta (4-8 Hz), Alpha (8-12 Hz), Beta (12-30 Hz) and Gamma (30-50 Hz). For each frequency, figures match the ones in Figure 5. First column: classification within Attention blocks. Second column: classification within Expectation blocks. Third column: significantly different clusters between Attention and Expectation. Blue depicts higher classification values in Attention, and orange depicts higher values in Expectation.

for both

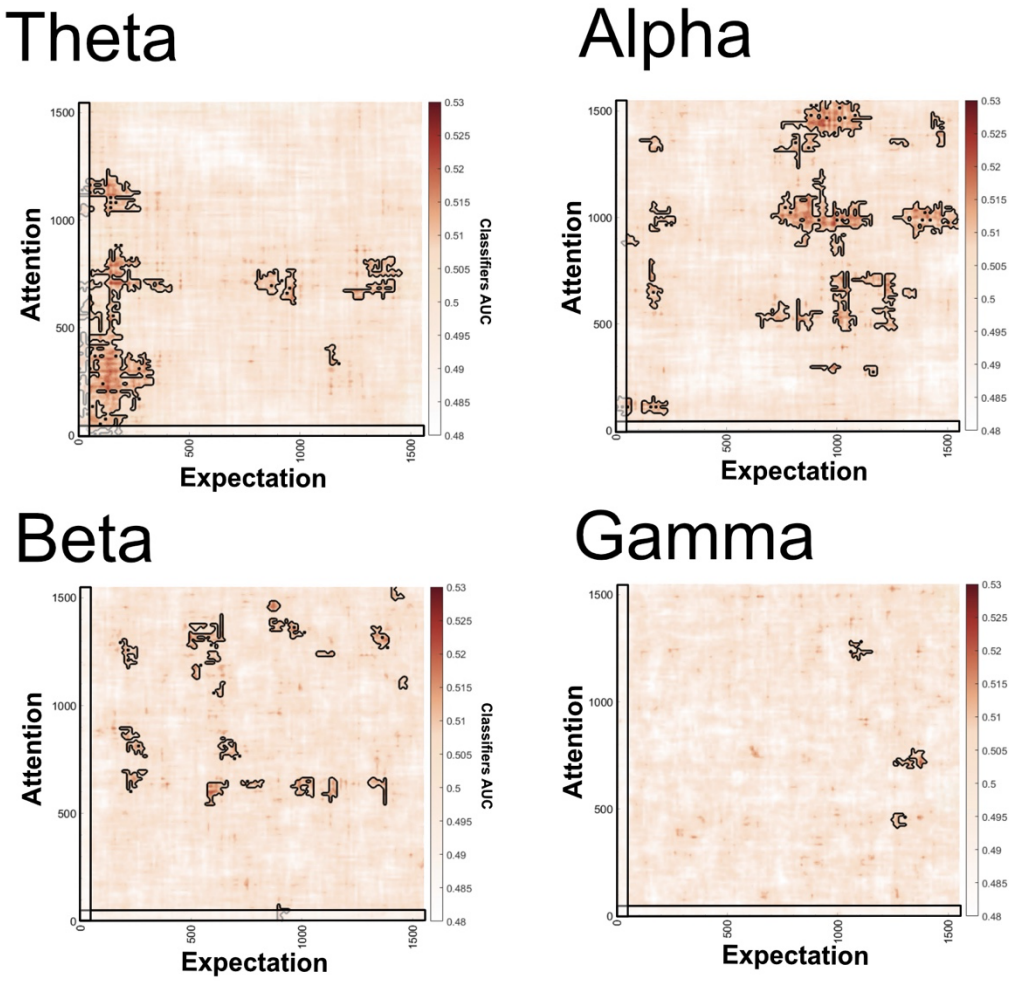

Figure 7. Cross-classification results across different frequency bands. The results are averaged in the two training and testing directions (see Figure 5D). See Figure S3 for results split by training and testing direction.

conditions (Figure 6 first and second column, black outlines). There were disperse significant clusters in several windows of Attention blocks, mostly in the first half of the matrix. For the expectation condition, the AUC only reached significance at the beginning of the interval. Results comparing the accuracy of Theta-classification in Attention and Expectation showed higher indices for Attention throughout most of the preparation window (Figure 6, third column).

A Theta cross-classification analysis (Figure 7, also see S3 for the results split by training and testing direction) showed Theta-shared patterns between anticipation in Attention and Expectation. The first 400 ms of the preparatory interval generalized through most of the selection window, until 1200 ms.

6.2. Alpha

There was little significant classification when employing Alpha in either condition. Only at the end of the interval in Attention we found two clusters of Alpha-driven decoding, at 1300 ms. This coding was also significantly higher in Attention than in Expectation. Interestingly, and similar to what

happened in Theta, Alpha-based cross-classification (Figure 7) resulted in higher generalization scores. Evidence of cross-condition generalization appeared at the end of the interval, at 1250 ms.

6.3. Beta

Beta classification was scarce in both conditions (Figure 6). Importantly, classification was higher in Expectation during a major part of the interval. This was probably due to Attention under chance values in certain windows of the interval, showing a perpendicular pattern [53]. Therefore, classification seems to be constant, with particular inversions at 500 and 1000 ms. In contrast to what we found in Theta and Alpha, there was little generalization between the two conditions, with only small and disperse clusters reaching significance (Figures 7 and S3).

6.4. Gamma

Gamma-driven classification was only present in the Attention condition (Figure 6). We found no significant clusters of over-chance classification in the Expectation condition throughout the entire preparatory window. However, Attention showed significant accuracy in the two axes of the matrix at 1200-1300 ms. Again, this perpendicular pattern suggests that Gamma coding is present in most of the Attention preparatory interval, but it is more prominent shortly before stimulus presentation. Cross-classification barely yielded any significant cluster (Figure 7).

Altogether, these frequency-band results suggest that representations during preparation in attention and expectation work through different but partially overlapping mechanisms.

7. Tracking perceptual patterns of brain activity

Other possible and non-exclusive source of variability between Attention and Expectation is the extent to which anticipatory representations function as perceptual reinstatements of the prepared target categories. In this case, the preparatory patterns that allow to classify anticipated faces and names should be similar to those triggered during actual stimulus perception. To assess such reinstatement of target representations during preparation, we obtained canonical template patterns (CTP) of brain activity associated with faces and names (as described in the Methods section). We then fit a linear regression with the two CTPs as regressors to measure the extent to which these explained variance in preparatory activity across conditions, in each time point (see Figure 8).

First, to complement the behavioral results regarding face and name processing, we applied the CTP to EEG activity locked to target onset in Attention and Expectation (Figure S4B). The face CTP explained better the perception of target faces, while same was true for the name pattern on name perception. Moreover, face processing was more pronounced, but decreased more rapidly. Name processing, on the other hand, reached a lower peak but it decreased more slowly over time. As expected, these results further suggest that faces and names are processed differently.

Then, we compared how these canonical templates overlapped with preparatory activity in Attention vs. Expectation contexts. Since both CTPs similarly explained face and name predicting cues, here we show the average result of the two types of cues separately in Attention and Expectation blocks. Crucially, both perceptual templates were better predictors of category anticipation in the Attention compared to the Expectation condition (see Figure 8). Overall, these results indicate that for both faces and names, regardless of their intrinsic differences in processing, there is an overall stronger perceptual reinstatement for Attention than for Expectation.

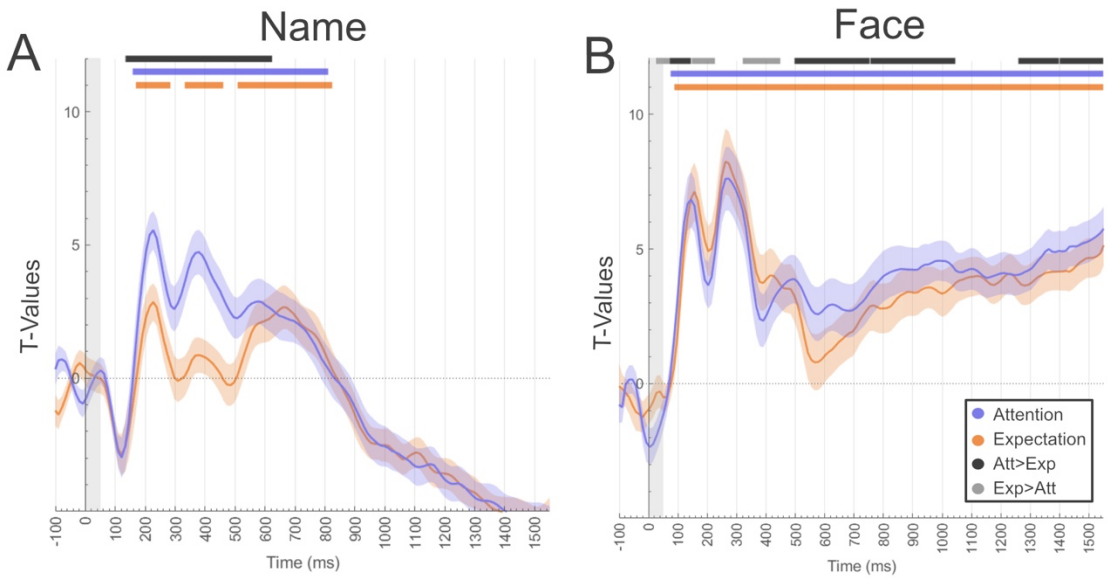

Figure 8. Canonical Template Tracking results. Grey shading represents cue on screen. **(A-B)** CTP results for separate cues for Attention and Expectation. Average of face and name cues **(A)** Name CTP model for cues. **(B)** Face CTP for cues. Horizontal lines depict significance. Black horizontal lines indicate points where the results for Attention were significantly higher than for Expectation. Grey horizontal lines show higher perceptual reinstatement for Expectation.

Discussion

In the present study we studied and contrasted, for the first time, the top-down neural mechanisms engaged during category-specific preparation in two major information-processing contexts: Selective Attention and Perceptual Expectations. Our results reveal that the anticipated stimulus category is active during preparation, regardless of the specific demands. Crucially, we demonstrate that relevance and probability influence the preparatory patterns of brain activity in a unique manner, potentially via different oscillatory profiles. Overall, these results have important implications for theoretical models that explain how the brain anticipates relevant vs. probable forthcoming information.

Our behavioral results show that the paradigm manipulated Attention (relevance) and Expectation (probability) in an effective manner and as expected[6,54]. Cued trials in Attention required further processing than those uncued, as participants were required to evaluate the sex/gender of the cued

stimulus category and skip this judgment in uncued ones[55–58]. Importantly, the probability of either category was equal (50%), keeping expectation constant in relevant and irrelevant trials. The high accuracy in Attention blocks indicates that the cues were used effectively to select when needed and to respond. On the other hand, both expected and unexpected trials required selecting the forthcoming target, and thus equated relevance while leading to predicted and unpredicted stimulus categories. In line with validity effects repeatedly reported in the literature in expectation contexts, the efficiency of behavior increased in cued trials[44,59].

We employed RSA[46,60] to study the temporal profile of information coding throughout the preparatory interval. This showed that anticipatory cues triggered several time-overlapping, yet distinguishable, effects in time (Figure 3B) and spatial arrangement (Figure 3C). The perceptual characteristics of the cue were coded early in the time interval. Immediately after the peak where this perceptual coding explained signal variability, and during the majority of the interval, mechanisms pertaining to either selective attention or expectations were deployed, as revealed by the Block model temporal profile. Evidence of anticipatory coding of different task sets has been found before[50,61,62]. Interestingly, this result cannot be explained by motor preparation, since cues did not predict motor responses; instead, it is most likely due to the control context in each scenario. Finally, the influence of the Category model on anticipatory activity increased steadily and peaked before target onset. This pattern of emerging, ramping-up category representations could reflect two non-exclusive anticipatory mechanisms. Once the meaning of the cue is extracted and stored in working memory, the actual anticipation of a stimulus could be facilitated by top-down preactivation of perceptual regions[63,64]. Simultaneously, the effect of temporal expectations (as the preparatory interval was fixed) could induce an increasing preactivation of perceptual regions as the stimulus onset approaches[28,65]. In either case, these results suggest that preparation engages a series of mechanisms that act both sequentially and in parallel, including the processing of bottom-up signals that are transformed into top-down category anticipation in a context of relevance or probability.

The block model in the RSA suggested that Attention and Expectation led to different coding patterns. One mechanistic explanation for this result is that the patterns for the anticipated category are more robustly coded in one condition over the other. Hence, we employed an MVPA approach[47,66] to classify the prepared category (faces or names). Results matched the RSA model, pointing to a robust effect of category anticipation. Furthermore, this was stable and increased towards the end of the interval. Again, this resonates with literature on temporal anticipation[65,67,68]. Moreover, prominent theories of attention[69,70] pose that information can be held in WM by strengthening the synaptic weights between neurons, allowing for shifts (in this case, steady increases) in decoding results during the delay period without losing information about the maintained stimulus. The results revealed no significant differences in AUC between the conditions in either time-resolved or temporal generalization matrices, and suggest that anticipated targets are coded with similar fidelity during Expectation and Attention. Importantly, these results were obtained cross-classifying different pairs of cues, thus avoiding perceptual confounds. However, a similar degree of accuracy classification does not warrant identical underlying mechanisms. To test this, we performed multivariate cross-classification between Attention and Expectation. Surprisingly, despite well above-chance decoding within each type of block, there was little cross-classification between conditions. Importantly, this could not be due to an effect of block, since classification scores were similar when training and

testing in different blocks of the same condition (see Figure S2). This result suggests that whereas Attention and Expectation both lead to anticipatory category representations, their top-down mechanisms are partially different, providing further support to the dissociation of relevance and probability observed during target processing[17,71].

One candidate explanation of the lack of generalization concerns the oscillatory profile of Attention and Expectation. To test this, we repeated the classification splitting the data into discrete frequency bands[72]. Surprisingly, for two of the frequencies (Theta and Alpha) we found cross-condition classification with no such results with (1) raw data or (2) within each condition splitting by frequency-band. Importantly, classifiers use both signal and noise covariance associated with other factors[9]. Therefore, the first situation could be explained by considering that the noise related to other frequency bands is reduced, allowing the emergence of results that were possibly shaded in raw voltage classification. In the second case, the distribution of data and noise within one condition may distort classification. Noise in one condition might mask classification scores[9]. If there is signal that is common across both conditions (in terms of shared information), the classifier might underperform because of the variability of each condition separately, but obtain better performance using two different datasets that share signal but differ in noise.

Theta (4-8 Hz) has been related to predictive processing[6], being key to the coordination of top-down mechanisms such as spatial attention[23] and working memory[73]. Theta decoding was more prominent at the beginning of the interval for Expectation, but sparse throughout the interval for Attention. Crucially, Theta cross-classification led to better results that within-condition. Similarly, while Alpha barely yielded results within-condition, it was more successful across conditions, with similarities appearing at the end of the anticipation interval, matching the ramping-up effects. Late object representation has been shown before[74], related to top-down modulations of object coding. Moreover, the result could also be linked to temporal expectations. Last, Gamma-based classification was only significant on the Attention condition. It showed a perpendicular pattern of generalization (see Fig 6), which may be interpreted as a mechanism being constant but subtle during most of time window, and emerging more prominently at certain time points[53,75]. This frequency band has been associated with prediction error in the predictive coding literature[76]. More broadly, Gamma seems the main frequency band for the feedforward communication in superficial layers of the cortex[29,77–79]. As discussed later, Attention showed an increased perceptual reinstatement of the anticipated category. Given the relevance of the Gamma band on feedforward perceptual coding, the relevant category could be primarily coded in this frequency, which in turn could be more similar to the actual stimulus perception.

Additionally, we studied the degree of preparatory perceptual reinstatement[73,80–84] in Attention and Expectation by employing a Canonical Template Tracking procedure[50,51] to compare perceptual reinstatement in attention and expectation. We obtained canonical representations of the two target categories (faces and names) from an independent localizer, and then estimated the variance explained by these perceptual specific patterns during the anticipatory window. First, we applied the extracted canonical templates of the actual target processing, with results supporting that faces are processed differently from words. Importantly, when applied to anticipatory activity, we found that the canonical templates related differently to the preparation interval, regardless of the predicted

category. Our results showed that the perceptual reinstatement for both faces and names was higher in Attention than in Expectation. Note that this cannot be due to a higher difficulty of the attention blocks, as RT showed the opposite pattern (faster responses in Attention). Instead, this higher reinstatement may be due to Attention directing resources to further activate the relevant category, making it more relevant than the ignored one. This would be in line with the Biased Competition model[1,12,85], that states that during target processing more resources are allocated to relevant stimuli, to enhance their cortical representation. Crucially, our results extrapolate this finding to the previous top-down phase, and show that this preparatory reinstatement is larger for relevant than for probable targets. This higher reinstatement in attention could be related to the oscillatory results in the decoding analysis that showed that Gamma drove the classification of the anticipated category in Attention, but not in Expectation, as Gamma activity has consistently been associated with perception and pyramidal cell communication in middle layers of the cortex[29,55,86]. A further strength of the scope of this finding is that the task presented a large pool of faces and words instead of repeating a small set of stimuli, and thus we induced proper category preparation, instead of reinstatements of concrete stimuli.

How Attention and Expectation affect perception is an ongoing debate. Several studies have used paradigms that combine both processes to study how they interact. Attention has been suggested to sharpen the differences between expected and unexpected stimuli[49], possibly changing the oscillatory profile of relevant categories[36] while reversing repetition suppression[48]. Relatedly, Attention boosts signal-to-noise processing by suppressing noise, while Expectation increases baseline activity in perceptual regions (Wyart et al., 2012). Predictive coding accounts propose that Attention increases prediction error of selected stimuli by suppressing noise of unattended categories, while expectation increases global sensitivity through prediction signals. Crucially, our results extend this literature by showing anticipatory differences between conditions, which cannot be accounted for by prediction error differences as targets have not been processed yet. Speculatively, attention could bias anticipatory neuronal sensitivity by increasing perceptual differences between relevant categories, coding these changes at least partially in the gamma band perhaps through anticipatory biasing of error processing units. Expectation could increase sensitivity to probable categories by increasing excitability of, perhaps, perceptual units.

Although overall our results are a crucial first step to characterize mechanisms across relevance and probability anticipation, they should be complemented by further studies. First, although the results are statistically significant after robust cluster-based correction, accuracy values are lower than those obtained, for example, using target-locked data (e.g., Figure S4A). Importantly, decoding accuracies do not equal effect sizes[9]. The values we obtained are within typical classification ranges when studying subtle neural patterns[9,81] (also see [87]) employing non-invasive human neuroimaging. Future studies may obtain higher accuracies by increasing stimulus repetition and thus reinforcing specific neural traces of the information anticipated. Second, our scope was limited to the temporal domain, and questions arise regarding potential differences between brain regions. Internal predictions have been generally associated with the hippocampus[43,44,88,89], which has a location that challenges EEG sensitivity. Further studies should employ more spatially sensitive techniques. Additionally, we focused on category-based preparation of faces and words, to facilitate having the same task across categories (sex/gender judgments). These two types of stimuli, although frequently combined in the

literature[81,90–93] have a different spatial layout, which may have generated the anticipation of different spatial templates. Although this is not a confound in our task, as such difference is constant, it may have added a spatial component to the preparation.

Altogether, our results show that anticipatory representations are category-specific and also task dependent, and highlight the role of proactive processes that precede stimulus perception. These findings have important implications for current models of brain functioning. We show that instead of single, unitary top-down phenomena, the brain implements distinct modes of content-specific anticipation, which are tailored to task context, operate within partially separate frequency bands and involve different levels of perceptual reinstatement. Predictive coding and attention models that differentiate these processes during target processing should be extended to the preparatory interval, acknowledging the specificity and complexity of these top-down anticipatory phenomena.

Acknowledgments

This research was supported by the Spanish Ministry of Universities under the PID2019-111187GB-100 grant to MR and by the Spanish Ministry of Economy and Business under the TEC2015-64718-R to JG. JMGP is supported by a scholarship from the Spanish Ministry of Universities (FPU18/01853). CGG was funded by the Spanish Ministry of Science and Innovation (Grants Ref.: IJC2019-040208-I and PID2020-116342GA-I00).

Author Contributions

Conceptualization, JMGP and MR; Data Collection, JMGP, DLG, and BAL; Analyses, JMGP and DLG; Resources, JG and MR; Writing – Original Draft, JMGP; Writing – Review & Editing, CGG and MR; Supervision, CGG and MR; Project Administration, MR; Funding Acquisition, MR

Declaration of Interests

The authors declare no competing interests.

Methods

Methods are reported in accordance with the Committee on Best Practices in Data Analysis and Sharing (COBIDAS) M/EEG (Pernet et al., 2019).

Data and code availability

All original code has been deposited at Github and is publicly available as of the date of submission (github.com/ChemaGP-UGR/AttExpLoc_EEG). Results have been deposited at the OSF framework: https://osf.io/2rhjn/?view_only=1a729284b9d549b082ef725bc5081f3d. Raw and preprocessed data are not available online due to storage size limitations, but are available under request.

1. Participants

Forty-eight participants (mean age = 22.06, range = 18-31; 29 women, 18 men, 1 non-binary) from the University of Granada were recruited and received from 20 to 25 euros, depending on their performance. Two additional participants were discarded due to low behavioral accuracy (less than 80%) or excessive noise in the EEG (more than 20% discarded trials). They were all native Spanish speakers, right-handed with normal or corrected vision, and signed informed consent prior to participation. Besides, to comply with COVID-19 guidelines, the temperature of participants was measured upon arrival (always <37ºC), they confirmed to have had no illness symptoms in the days prior to the experiment and wore a face mask during the whole session. Sample size was calculated to achieve a statistical power of 80% for an estimated small effect size (Cohen's D = 0.3) and three independent variables (block x category x cueing). Using PANGEA (Power ANalysis for GEneral ANOVA designs[95] we obtained a minimum of 32 participants to be able to detect the block x cueing interaction in reaction times and behavioral accuracy, our main behavioral prediction. To fit the counterbalancing scheme, we tested 48 participants. This sample size provides an estimated power of 94% under the described parameters.

2. Apparatus, stimuli, and procedure

Stimulus presentation and behavioral data collection were done with The Psychophysics Toolbox 3[96] on MATLAB (v.2020a) in a Microsoft PC. Stimuli were presented on an LCD screen (Benq, 1920x1080 resolution, 60 Hz refresh rate) over a grey background. We employed 160 male and female faces (50% each, with ~6ºx9º visual angle, extracted from The Chicago Face Database[97]) plus 160 unique Spanish male and female names (50% each, with ~8ºx2º visual angle). Four different geometrical shapes (circle, square, rain-drop and diamond with thin black outlines, unfilled, ~2ºx2º visual angle) were used as cues in the main task. The sound stimuli employed in the localizer blocks consisted on four different tones (250, 300, 350 and 400 Hz).

The main task was a cue-target paradigm where, depending on the block, cues carried information about either the *relevance* (Attention) or the *probability* (Expectation) of upcoming face or word targets. Each trial started with the presentation of this visual cue. For each participant, and to avoid perceptual confounds, two cue shapes (counterbalanced across participants) were associated with faces and two with names. The task was to indicate the sex/gender of this target (male or female). Participants pressed one of two keys ("a", "l", counterbalanced across participants) to respond whether or not the target belonged to the gender stated at the beginning of each block. Half of the blocks belonged to the Attention condition, and the other half to the Expectation condition. Participants were verbally instructed to use the cues in the two blocks to respond as fast as possible while avoiding mistakes. At the beginning of each block (Figure 1B), they were informed about the block (Attention or Expectation), the target sex/gender ("Is the target male/female?"), and the two cues (one for faces and one for names). Importantly, and since Attention and Expectation are involved in almost any act of visual perception, we aimed at manipulating one process while keeping the other constant. In Attention blocks, the cue indicated the *relevant* stimulus category to select (either faces or names). Only if the stimulus belonged to the relevant category (50% trials, cued), the

participant had to perform the gender discrimination task on the target. Otherwise, participants had to answer '*no*' regardless of the stimulus sex/gender (non-relevant category, uncued). Note that this manipulation of relevance, where further processing has to be applied only to selected stimuli, is similar to that employed in previous literature (e.g.[55–58]). Importantly, both relevant and non-relevant targets were matched in expectation, as by design they appeared with a 50% probability after each attention cue. On the other hand, in Expectation blocks the cue indicated the *probable* category of the target, with a 75% likelihood (e.g.[6,18] for similar manipulations). Here, participants had to perform the gender discrimination task in all trials, whether or not the target was cued. This way, both the expected and unexpected targets were equally relevant.

In every trial of the main task, the sequence of events was as follows: a 50 ms cue was followed by a fixed Cue-Target Interval (CTI) of 1500 ms and then the target appeared for 100 ms. Trials were separated by 2500 ms intervals. Auditory (tone) and visual feedback (words "Attention" or "Expectation", depending on the block) was presented for 500 ms in case of a wrong answer 1.3 seconds after target presentation, and lasted for 500ms, without altering the trials duration. Each trial lasted 4.15 seconds and each block 1.23 minutes. This main task was composed by 32 blocks of 20 trials each, or 640 trials in total.

In addition, Attention and Expectation blocks were interspersed with Localizer ones, used to measure target perceptual processing without overt motor activity (adapted from Egner et al., 2010). In these, the same stimuli as in the main task were presented. They were preceded by auditory cues that predicted either faces or names with 75% validity. Similar to the main task, there were two tones predicting faces, and two tones predicting names. One of each type was used per block. Participants had to press a key *only* if the stimulus appeared upside down (rotated 180º, 10% of trials equated in stimulus category and sex/gender), regardless of cue validity. At the beginning of each block, a screen indicated the type of block and the two tones for the block. There were 16 localizer blocks of 40 trials each, 640 trials in total. The cue lasted 200 ms, the visual stimuli appeared after an ISI of 1500 ms and stayed onscreen for 100 ms. Trials were separated by 1500 ms intervals. Overall, each localizer trial lasted 3.3 seconds and each localizer block lasted 2.12 minutes.

The three types of blocks (Attention, Expectation and Localizer) appeared in a fully counterbalanced order, so that that they preceded and followed each other an equal number of times. Cues and target stimuli were also fully counterbalanced across participants. In total, the whole experimental session lasted approximately 80 minutes, with additional practice and EEG preparation time.

3. EEG data acquisition and preprocessing

3.1. Acquisition

High-density EEG was recorded with 64 channels mounted on an elastic cap (actiCap Slim, BrainVision) at the Mind, Brain and Behavior Research Center (CIMCYC) of the University of Granada. Impedances were kept below 10 kΩ as recommended by the amplifiers' manufacturers. EEG activity was referenced online to FCz and recorded at a sampling rate of 1000 Hz.

3.2. Preprocessing

Data were preprocessed using EEGLAB[98] and custom MATLAB scripts[99]. EEG recordings downsampled to 256Hz, digitally low-pass filtered using FIR filter at 120 Hz and high-pass filtered at 0.1 Hz. A notch bandpass was applied at 50 Hz and 100 Hz to remove line noise and its harmonics. Channels with excessive noise were identified by visual inspection and removed from the data (1.85% of channels on average, range 0-5). We used different epochs for cue and target stimuli and in each we split the data into 3 seconds epochs (-1 to 2 seconds after the onset of each stimulus). Then, Independent Component Analysis (ICA) was carried out to remove eye artifacts (i.e. blinks and lateral eye movements) using the *runica* algorithm from EEGLAB. Component rejection was guided by visual inspection of scalp maps, raw activity and power spectrum. ICLabel was used for further confirmation. In average 1.85 components were removed per participant (range 2-4). Next, we performed automatic trial rejection to prune the data from non-stereotypical artifacts. It was based on three factors: 1) abnormal spectra: trials in which the spectrum deviated from baseline by ±50 dB in the 0–2 Hz frequency window (sensitive to remaining eye artifacts) or deviated by −100 dB or +25 dB in 20–40Hz (sensitive to muscle activity); 2) improbable data: the probability of occurrence of each trial was computed by determining the probability distribution of values across trials, with a rejection threshold established at ±6 SD; (3) extreme values: all trials with amplitudes in any electrode out of a ±150μV range were automatically rejected (see [99–101] for similar preprocessing routines). The three methods in sum yielded an average of 8% of rejected trials per participant (range 1.8%-19%). Afterwards, the removed channels were repaired by spherical interpolation. We then applied common average to re-reference the data, given its widespread use and its optimal adaptation to high-density recordings[94]. Finally, trials were baseline corrected in the -200 to 0 ms prior to stimulus onset.

3.3. Frequency band extraction

We performed frequency extraction separately for each trial and electrode. The data was convolved with complex Morlet wavelets (logarithmically spaced in 36 frequencies from 4 to 49 Hz) and the wavelet's length was calculated separately for each frequency following [102]. The power spectrum was estimated with a fast Fourier transform, and then multiplied by the power spectrum of the frequency specific Morlet wavelet. Then, with an inverse Fast Fourier Transform we obtained the time-resolved analytic signal. The squared absolute value of this signal returned the power values for each channel and trial, which we later normalized with respect to pre-stimulus baseline (-200 to 0). Finally, we averaged the 36 frequencies into four bands, which were used separately in posterior analyses (see [72] for a similar approach): Theta (4-7 Hz, 10 out of the 36 frequencies), Alpha (8-12 Hz, 6 frequencies), Beta (13-30 Hz, 12 frequencies) and Gamma (31-49 Hz, 8 frequencies).

4. Analyses

4.1. Behavioral

The main task design had three within-subject factors: Block type (Attention vs. Expectation), Cueing (Cued vs. Uncued) and Stimulus category (Faces vs. Names). We calculated three-way repeated

640 measures ANOVA separately for behavioral accuracy and reaction times (RTs) employing JASP [103].
641 For each participant, trials with longer or shorter RTs than the average ± 2 SDs were discarded (11.54
642 % on average). Behavioral results in the localizer were not relevant and were just considered as
643 exclusion criteria in case of poor performance.
644
645 4.2. Representational Similarity Analysis
646

647 Representational Similarity Analysis (RSA) allows relating empirical multivariate measures of brain
648 activity to theoretical models[104]. We performed RSA using voltage values on a subject-by-subject
649 basis. Prior to the analyses data were normalized by z-scoring the values across all trials, regardless
650 of the condition. We then constructed empirical Representational Dissimilarity Matrices (RDMs)
651 every three time points, which measure the geometrical distances between all experimental conditions
652 (see MVPA section) and, finally, estimated the relationship between empirical RDMs and theoretical
653 models.

654 RDMs were built with data from eight conditions, yielding 8x8 symmetrical matrices. These
655 conditions were all the possible combinations between the design variables: *cue prediction* (faces and
656 names); *cue shape* (shapes 1 and 2 for faces, shapes 3 and 4 for names) and *block* (attention and
657 expectation). We employed a Linear Discriminant Contrast (LDC, also known as Crossvalidated
658 Mahalanobis Distance, see[105] as measure of distance between conditions, for the following reasons:
659 it is a continuous measure, so it is highly reliable, informative, and lacks a ceiling effect; it includes
660 a crossvalidation loop, which makes it less prone to biases; and it is centered around 0 when the true
661 distance is 0 and therefore it is easier to interpret and more generalizable [106]. We calculated LDC as
662 described in[107]. For every time point and each pair of conditions, we calculated the mean of each
663 channel. This was done in two different datasets (train and test), to perform two-fold crossvalidation.
664 The distance between the two conditions in the two folds was multiplied by the pseudo inverse
665 covariance matrix between the residuals of the first and the second conditions in the training set, and
666 the distance values were then averaged across the two folds.

667 Theoretical RDMs (Figure 3a) were built based on the expected distances (assigning values of 0 or
668 1) between conditions, according to different hypotheses. We built three such model RDMs based
669 on: (1) cue shape (increased similarity between cues with same shape, regardless of the block and
670 predicted target); (2) category (increased similarity between cues that predict the same stimulus
671 category); and (3) block (increased similarity between cues belonging to the same block). The next
672 step was to estimate the share of variance that each of these three model RDMs explained. To do so,
673 at each time point we fitted a linear regression with the model RDMs as regressors and the empirical
674 matrix as dependent variable. As a result, we obtained a t-value for each model, time point, and
675 participant.

676 We used a non-parametric cluster-based permutation method to infer statistical significance at the
677 group level, against empirical chance levels. First, at the single-subject level we randomly permuted
678 the values of the theoretical matrices to generate 100 RSA maps per participant. Then, one map was
679 randomly drawn per subject and the selected maps were averaged. This was repeated 10^5 times to
680 obtain 10^5 permuted group RSA maps. For each time point, the empirical chance distribution was
681 estimated. As expected, this yielded a t-value distribution centered around 0. The above and below

thresholds were estimated so that they included the 95% of the distribution. Clusters containing time points with values outside the previously calculated thresholds were measured generating the null distribution of cluster sizes. Finally, we corrected for multiple comparisons using a False Discovery Rate (FDR) approach to determine the smallest cluster size deemed significant for $\alpha = 0.05$ [99]. Note that here permuted maps were obtained by randomly exchanging the values of the theoretical matrices.

Besides, to estimate the topography of each model we repeated the analyses for each channel separately, estimating the empirical matrix using all time points. That is, at each channel we fitted a linear regression with the models as regressors.

4.3. Time resolved Multivariate Pattern Analysis (MVPA)

We used time-resolved MVPA[47] to study stimulus category-specific preparation by classifying faces vs. names. This was done *before* stimulus presentation with cue-locked EEG, where faces and names were anticipated as relevant or probable by the cue (targets were not relevant for this work and thus not analyzed). The steps of the classification were equal for all analyses unless otherwise specified and were performed with both voltage and time-frequency data. Classification was performed using MVPAlab[99] running on MATLAB. To maximize observations while reducing computational costs, we performed an MVPA classification every three time points. This way features used for classification were trials by channels matrices of raw voltage/frequency in single time points t_n.

We applied two strategies to increase signal-to-noise ratio. First, the trials used were the result of averaging sets of three trials (randomly selected) in each condition[47]. Then, we employed a smoothing method, based on a moving average filter with a length of 3 time points. For every t_n the features of the previous and the following data points were averaged, so that $t_n = (t_{n-1} + t_n + t_{n+1})/3$. Then, classes were balanced by subsampling the class with more trials so that the number of trials from the two conditions fed to the algorithm stayed the same [47]. A five-fold stratified cross-validation loop was implemented [108], which ensures that the proportion of each class stays balanced across folds, thus increasing the classification applicability to unknown data. Data were divided into five parts, which are enough to obtain unbiased results with a relatively small computational cost[47,109]. The classification algorithm was trained with the first four divisions (training set) and tested on the remaining one (test set). To improve the performance of the classifiers and the generalizability of multivariate analyses results[110], we normalized the data. Normalization was carried out within the cross-validation loop[52]. Within each fold, we calculated the mean and standard deviation of each electrode across the training trials. The train set (X_{train}) and testing set (X_{test}) were normalized as follows:

$$X_{train} = \frac{X_{train} - \mu_{train}}{\sigma_{train}} \qquad X_{test} = \frac{X_{test} - \mu_{train}}{\sigma_{train}}$$

Where μ_{train} and σ_{train} are the mean and standard deviation of the training set.

We used Linear Discriminant Analysis as classification algorithm, given its good fit to typical EEG variability and higher sensitivity than similar methods (e.g. Support Vector Machines, see [22,47]. Time

721 ranges were from -100 ms to 1550 ms. Classification results were estimated with an empirical receiver
722 operative curve (ROC) analysis and reported as the area under the curve (AUC). This method works
723 as an estimate of the true positive rate as a function of the false positive rate. The AUC is a non-
724 parametric criterion-free method, so it does not involve assumptions about of the true distribution of
725 the data[52]. Besides, it is less susceptible to systematic biases and it is especially sensitive to two-class
726 differences. AUC results can be interpreted similarly to classification accuracy, with 0.5 indicating
727 equal probability of true positives and false positives and 1 accounting for perfect discriminability
728 between classes (King et al., 2013). Statistical significance was calculated, as previously, by
729 performing cluster-based permutation analyses. Note that here we randomly permuted the labels of
730 each class, which yielded a distribution centered around chance level (0.5).

731

732 #### 4.3.1. Temporal Generalization Analyses

733

734 To characterize the changes of the signal throughout the temporal window, we employed a temporal
735 generalization approach[111]. On each time point we trained a classifier following the process described
736 above. Then, we tested it on all time points of the preparation time window. This rendered a Temporal
737 Generalization Matrix (TGM) representing thep AUC values for each train-test pair. Statistical
738 significance was then calculated for each TGM following the same rationale as with time-resolved
739 analyses. The only difference was setting the minimum statistically significant threshold for cluster
740 sizes to $p<0.001$, to avoid small clusters.

741

742 #### 4.3.2. Category-specific anticipation within attention and expectation contexts

743 The process described was applied to the cue-locked EEG separately for Attention and Expectation
744 blocks, training classifiers to tell apart data from trials in which the cue anticipated faces vs. names.
745 In a first approximation, we used trials of all cue shapes in each category (Figure 1c). Next, to ensure
746 that classification results were not biased by perceptual differences between the geometrical shapes
747 used as cues, we implemented a classification approach across cues. We first trained the classifier
748 with data from only one cue from each category pair (e.g., classifying between squares predicting
749 faces and diamonds anticipating names) and then tested it on the remaining pair (e.g., testing with
750 circles that predicted faces and drops that predicted names). This cross-classification ensures that only
751 the common differences between the two classified pairs will be decoded from the results [75], thus
752 removing perceptual confounds. We averaged the results of both directions and their permutation
753 maps to obtain a greater signal-to-noise ratio and to reduce biases due to the classification of specific
754 perceptual features. The averaged results were then fed to the same statistical algorithm used
755 previously to obtain cluster-based thresholds of statistical significance.

756 Once the results of these classifications were obtained, we compared the scores in Attention vs.
757 Expectation blocks. To do so, we subtracted the empirical results and the permuted maps of the two
758 conditions (Attention – Expectation). We used a two-tailed one sample t-test to evaluate differences
759 in the two directions (Attention > or < Expectation), following the same cluster-based permutation
760 implementation described above. Note that in this analysis the permuted distribution is centered
761 around zero.

762

4.3.3. Cross-classification between Attention and Expectation

To estimate the degree to which patterns of brain activity are shared for preparation across Attention and Expectation, we employed a cross-classification approach. We trained a classifier with data from one condition and then tested it on the other. We first did this using data from all trials in each condition, incorporating all four types of cues. Again, to rule out perceptual confounds we repeated the analyses following the same rationale described in the previous section (see Figure 1c for a summary of the condition selection strategies).

4.4. Canonical Template Tracking

Finally, we compared the sustained patterns that arose during the preparatory interval with the actual perception of face and name stimuli. To do so, we obtained Canonical Template Patterns[51,112] of brain activity generated by faces and names in the independent localizer blocks. First, we performed an MVPA analysis in localizer trials following the same process as in previous analysis. Then, we selected the time window where the classifiers obtained higher AUC (i.e., when activity patterns were more dissimilar) across participants (S4a), which was 100 – 300 ms after stimulus onset. Then, and separately for face and name localizer trials, we averaged the raw information of every time-point and trial in the selected window for each channel and category. This resulted in a vector of 64 channel activity values for faces and another one for names in each participant[73]. Next, these CTP of faces and names were used as regressors in a linear regression where the dependent variable was the raw channel activity for each channel and condition during every time-point in the preparation or target window of the main experimental task. This rendered two t-values per time-point that accounted for the variance explained for each CTP (faces and names). We did this analysis separately in Attention and Expectation, for cues predicting faces and names and for face and name targets. To estimate statistical significance, we used the same cluster-based permutation analysis described above. Here, permutations were performed by shuffling the channel values separately for each CTP.

References

1. Battistoni, E., Stein, T. & Peelen, M. V. Preparatory attention in visual cortex. *Ann. N. Y. Acad. Sci.* **1396**, 92–107 (2017).

2. González-García, C., Mas-Herrero, E., de Diego-Balaguer, R. & Ruz, M. Task-specific preparatory neural activations in low-interference contexts. *Brain Struct. Funct.* **221**, 3997–4006 (2016).

3. Nobre, A. C. & Serences, J. T. Building on a Solid Baseline: Anticipatory Biases in Attention. *Trends Neurosci.* **41**, 120–122 (2018).

4. 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).

5. Aranda, C., Madrid, E., Tudela, P. & Ruz, M. Category expectations: A differential modulation of the N170 potential for faces and words. *Neuropsychologia* **48**, 4038–4045 (2010).

6. Kok, P., Mostert, P. & de Lange, F. P. Prior expectations induce prestimulus sensory templates. *Proc. Natl. Acad. Sci.* 201705652 (2017) doi:10.1073/pnas.1705652114.

7. Koshino, H., Minamoto, T., Yaoi, K., Osaka, M. & Osaka, N. Coactivation of the Default Mode Network regions and Working Memory Network regions during task preparation. *Sci. Rep.* **4**, 5954 (2015).

8. van Driel, J., Gunseli, E., Meeter, M. & Olivers, C. N. L. Local and interregional alpha EEG dynamics dissociate between memory for search and memory for recognition. *Neuroimage* **149**, 114–128 (2017).

9. Hebart, M. N. & Baker, C. I. Deconstructing multivariate decoding for the study of brain function. *Neuroimage* **180**, 4–18 (2018).

10. Baines, S., Ruz, M., Rao, A., Denison, R. & Nobre, A. C. Modulation of neural activity by motivational and spatial biases. *Neuropsychologia* **49**, 2489–2497 (2011).

11. Auksztulewicz, R. & Friston, K. Repetition suppression and its contextual determinants in predictive coding. *Cortex* **80**, 125–140 (2016).

12. Desimone, R. & Duncan, J. Neural Mechanisms of Selective Visual Attention. *Annu. Rev. Neurosci.* 193–222 (1995).

13. Schröger, E., Kotz, S. A. & SanMiguel, I. Bridging prediction and attention in current research on perception and action. *Brain Res.* **1626**, 1–13 (2015).

14. Stokes, M., Thompson, R., Nobre, A. C. & Duncan, J. Shape-specific preparatory activity mediates attention to targets in human visual cortex. *Proc. Natl. Acad. Sci.* **106**, 19569–19574 (2009).

15. Hong, X., Wang, Y., Sun, J., Li, C. & Tong, S. Segregating Top-Down Selective Attention from Response Inhibition in a Spatial Cueing Go NoGo Task: An ERP and Source Localization Study. *Sci. Rep.* **7**, 1–14 (2017).

16. Summerfield, C. & De Lange, F. P. Expectation in perceptual decision making: Neural and computational mechanisms. *Nat. Rev. Neurosci.* **15**, 745–756 (2014).

17. Wyart, V., Nobre, A. C. & Summerfield, C. Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 3593–3598 (2012).

18. de Lange, F. P., Rahnev, D. A., Donner, T. H. & Lau, H. Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. *J. Neurosci.* **33**, 1400–1410 (2013).

19. Ho, T. C., Brown, S., Abuyo, N. A., Ku, E. H. J. & Serences, J. T. Perceptual consequences of feature-based attentional enhancement and suppression. *J. Vis.* **12**, 1–17 (2012).

20. Stein, T. & Peelen, M. V. Specific Expectations Enhance Stimulus Detectability by Increasing Perceptual Sensitivity Content-Specific Expectations Enhance Stimulus Detectability by Increasing Perceptual Sensitivity. *J. Exp. Psychol. Gen.* (2015).

21. Landau, A. N., Schreyer, H. M., Van Pelt, S. & Fries, P. Distributed Attention Is Implemented through Theta-Rhythmic Gamma Modulation. *Curr. Biol.* **25**, 2332–2337 (2015).

22. Kerrén, C., Linde-Domingo, J., Hanslmayr, S. & Wimber, M. An Optimal Oscillatory Phase for Pattern Reactivation during Memory Retrieval. *Curr. Biol.* **28**, 3383-3392.e6 (2018).

23. Fiebelkorn, I. C. & Kastner, S. A Rhythmic Theory of Attention. *Trends Cogn. Sci.* **23**, 87–101 (2019).

24. Wöstmann, M., Alavash, M. & Obleser, J. Alpha oscillations in the human brain implement distractor suppression independent of target selection. *J. Neurosci.* **39**, 9797–9805 (2019).

25. Fries, P. A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends Cogn. Sci.* **9**, 474–480 (2005).

26. Bosman, C. A. *et al.* Attentional Stimulus Selection through Selective Synchronization between Monkey Visual Areas. *Neuron* **75**, 875–888 (2012).

27. Auksztulewicz, R. *et al.* Not all predictions are equal: 'What' and 'When' predictions modulate activity in auditory cortex through different mechanisms. *J. Neurosci.* **38**, 0369–18 (2018).

28. Rohenkohl, G., Cravo, A. M., Wyart, V. & Nobre, A. C. Temporal expectation improves the quality of sensory information. *J. Neurosci.* **32**, 8424–8428 (2012).

29. Michalareas, G. *et al.* Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas. *Neuron* **89**, 384–397 (2016).

30. Brunet, N. M. *et al.* Stimulus repetition modulates gamma-band synchronization in primate visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 3626–3631 (2014).

31. Schröger, E. A neural mechanism for involuntary attention shifts to changes in auditory stimulation. *J. Cogn. Neurosci.* **8**, 527–539 (1996).

32. Posner, M. I. Orienting of attention. *Q. J. Exp. Psychol.* **32**, 3–25 (1980).

33. Zuanazzi, A. & Noppeney, U. Distinct neural mechanisms of spatial attention and expectation guide perceptual inference in a multisensory world. *J. Neurosci.* **39**, 2873–18 (2019).

34. Wyart, V., Nobre, A. C. & Summerfield, C. Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 3593–3598 (2012).

35. Simon, A., Schachtner, J. & Gallen, C. L. Disentangling expectation from selective attention during perceptual decision-making. *SSRN Electron. J.* **5**, 1–19 (2019).

36. Auksztulewicz, R., Friston, K. J. & Nobre, A. C. Task relevance modulates the behavioural and neural effects of sensory predictions. *PLoS Biol.* **15**, 1–27 (2017).

37. Giesbrecht, B., Weissman, D. H., Woldorff, M. G. & Mangun, G. R. Pre-target activity in visual cortex predicts behavioral performance on spatial and feature attention tasks. *Brain Res.* **1080**, 63–72 (2006).

38. González-García, C., Flounders, M. W., Chang, R., Baria, A. T. & He, B. J. Content-specific activity in frontoparietal and default-mode networks during prior-guided visual perception. *Elife* **7**, 1–25 (2018).

39. Esterman, M. & Yantis, S. Perceptual expectation evokes category-selective cortical activity. *Cereb. Cortex* **20**, 1245–1253 (2010).

40. Peelen, M. V. & Kastner, S. A neural basis for real-world visual search in human occipitotemporal cortex. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 12125–12130 (2011).

41. Soon, C. S., Namburi, P. & Chee, M. W. L. Preparatory patterns of neural activity predict visual category search speed. *Neuroimage* **66**, 215–222 (2013).

42. Ekman, M., Kok, P. & De Lange, F. P. Time-compressed preplay of anticipated events in human primary visual cortex. *Nat. Commun.* **8**, 1–9 (2017).

43. Hindy, N. C., Ng, F. Y. & Turk-Browne, N. B. Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nat. Neurosci.* **19**, 665–667 (2016).

44. de Lange, F. P., Heilbron, M. & Kok, P. How Do Expectations Shape Perception? *Trends Cogn. Sci.* **22**, 764–779 (2018).

45. Egner, T., Monti, J. M. & Summerfield, C. Expectation and surprise determine neural population responses in the ventral visual stream. *J. Neurosci.* **30**, 16601–16608 (2010).

46. Kriegeskorte, N. Representational similarity analysis – connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 1–28 (2008).

47. Grootswagers, T., Wardel, 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).

48. Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C. & De Lange, F. P. Attention reverses the effect of prediction in silencing sensory signals. *Cereb. Cortex* **22**, 2197–2206 (2012).

49. Jiang, J., Summerfield, C. & Egner, T. Attention sharpens the distinction between expected and unexpected percepts in the visual brain. *J. Neurosci.* **33**, 18438–18447 (2013).

50. González-García, C., Formica, S., Wisniewski, D. & Brass, M. Frontoparietal action-oriented codes support novel instruction implementation. *Neuroimage* **226**, (2021).

51. Wimber, M., Alink, A., Charest, I., Kriegeskorte, N. & Anderson, M. C. Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nat. Neurosci.* **18**, 582–589 (2015).

52. King, J.-R. & Dehaene, S. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn. Sci.* **18**, 203–210 (2014).

53. King Dehaene, J. R. & S. Characterizing the dynamics of mental representations: The temporal generalization method. *Trends Cogn. Sci.* **18**, 203–210 (2014).

54. Baldauf, D. & Desimone, R. Neural mechanisms of object-based attention. *Science (80-.).* **344**, 424–427 (2014).

55. Womelsdorf, T., Fries, P., Mitra, P. P. & Desimone, R. Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature* **439**, 733–736 (2006).

56. Baldauf, D. & Desimone, R. Neural mechanisms of object-based attention. *Science (80-.).* **344**, 424–427 (2014).

57. Saenz, M., Buracas, G. T. & Boynton, G. M. Global effects of feature-based attention in human visual cortex. *Nat. Neurosci.* **5**, 631–632 (2002).

58. Summerfield, C. *et al.* Predictive Codes for Forthcoming Perception in the Frontal Cortex. *Science (80-.).* **314**, 1311–1314 (2006).

59. Sy, J. L., Guerin, S. A., Stegman, A. & Giesbrecht, B. Accurate expectancies diminish perceptual distraction during visual search. *Front. Hum. Neurosci.* **8**, 1–18 (2014).

60. Carlson, T. A., Grootswagers, T. & Robinson, A. K. An introduction to time-resolved decoding analysis for M/EEG. (2019).

61. Hebart, M. N., Bankson, B. B., Harel, A., Baker, C. I. & Cichy, R. M. The representational dynamics of task and object processing in humans. *Elife* **7**, (2018).

62. Palenciano, A. F., González-garcía, C., Arco, J. E. & Ruz, M. Transient and Sustained Control Mechanisms Supporting Novel Instructed Behavior. *Cereb. Cortex* 1–13 (2018) doi:10.1093/cercor/bhy273.

63. Auksztulewicz, R. & Friston, K. Repetition suppression and its contextual determinants in predictive coding. *Cortex* **80**, 125–140 (2016).

64. Trapp, S., Lepsien, J., Kotz, S. A. & Bar, M. Prior probability modulates anticipatory activity in category-specific areas. *Cogn. Affect. Behav. Neurosci.* **16**, 135–144 (2016).

65. Jin, W., Nobre, A. C. & van Ede, F. Temporal expectations prepare visual working memory for behavior. *J. Cogn. Neurosci.* **32**, 2320–2332 (2020).

66. Haxby, J. V., Connolly, A. C. & Guntupalli, J. S. Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. *Annu. Rev. Neurosci.* **37**, 435–456 (2014).

67. Ruz, M. & Nobre, A. C. Attention modulates initial stages of visual word processing. *J. Cogn. Neurosci.* **20**, 1727–1736 (2008).

68. Barbosa, J. *et al.* Interplay between persistent activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory. *Nat. Neurosci.* **23**, 1016–1024 (2020).

69. Trübutschek, D. *et al.* A theory of working memory without consciousness or sustained activity. *Elife* **6**, 1–29 (2017).

70. Mongillo, G., Barak, O. & Tsodyks, M. SynaptiC Theory of Working Memory. *Science (80-.).* **319**, 1543–1546 (2008).

71. Summerfield, C. & Egner, T. Expectation (and attention) in visual cognition. *Trends Cogn. Sci.* **13**, 403–409 (2009).

72. Xie, S., Kaiser, D. & Cichy, R. M. Visual Imagery and Perception Share Neural Representations in the Alpha Frequency Band. *Curr. Biol.* **30**, 2621-2627.e5 (2020).

73. Kerrén, C., Linde-Domingo, J., Hanslmayr, S. & Wimber, M. An Optimal Oscillatory Phase for Pattern Reactivation during Memory Retrieval. *Curr. Biol.* **28**, 3383-3392.e6 (2018).

74. Hebart, M. N., Bankson, B. B., Harel, A., Baker, C. I. & Cichy, R. M. The representational dynamics of task and object processing in humans. *Elife* **7**, 1–21 (2018).

75. Kaplan, J. T., Man, K. & Greening, S. G. Multivariate cross-classification: applying machine learning techniques to characterize abstraction in neural representations. *Front. Hum. Neurosci.* **9**, 1–12 (2015).

76. Auksztulewicz, R. *et al.* The cumulative effects of predictability on synaptic gain in the auditory processing stream. *J. Neurosci.* **37**, 6751–6760 (2017).

77. Fries, P. Neuronal Gamma-Band Synchronization as a Fundamental Process in Cortical Computation. (2009) doi:10.1146/annurev.neuro.051508.135603.

969 78. Scheeringa, R. & Fries, P. Cortical layers, rhythms and BOLD signals. *Neuroimage* **197**,
970 689–698 (2019).

971 79. Xing, D., Yeh, C. I., Burns, S. & Shapley, R. M. Laminar analysis of visually evoked activity
972 in the primary visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 13871–13876 (2012).

973 80. Vetter, P., Smith, F. W. & Muckli, L. Decoding sound and imagery content in early visual
974 cortex. *Curr. Biol.* **24**, 1256–1262 (2014).

975 81. Rose, N. S. *et al.* Reactivation of latent working memories with transcranial magnetic
976 stimulation. *Science (80-.).* **354**, 1136–1139 (2016).

977 82. Kilner, J. M., Friston, K. J. & Frith, C. D. Predictive coding: An account of the mirror
978 neuron system. *Cogn. Process.* **8**, 159–166 (2007).

979 83. Muckli, L. *et al.* Contextual Feedback to Superficial Layers of V1. *Curr. Biol.* **25**, 2690–
980 2695 (2015).

981 84. Smith, F. W. & Muckli, L. Nonstimulated early visual areas carry information about
982 surrounding context. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 20099–20103 (2010).

983 85. Feldman, H. & Friston, K. J. Attention, uncertainty, and free-energy. *Front. Hum.*
984 *Neurosci.* **4**, 1–23 (2010).

985 86. Scheeringa, R. E. & Fries, P. Cortical layers, rhythms and BOLD signals. *Neuroimage* 1–
986 10 (2017) doi:10.1016/j.neuroimage.2017.11.002.

987 87. Christophel, T. B., Cichy, R. M., Hebart, M. N. & Haynes, J. D. Parietal and early visual
988 cortices encode working memory content across mental transformations. *Neuroimage*
989 **106**, 198–206 (2015).

990 88. Aitken, F. & Kok, P. Hippocampal representations switch from errors to predictions
991 during acquisition of predictive associations. *Nat. Commun.* **13**, 1–13 (2022).

992 89. Stachenfeld, K. L., Botvinick, M. M. & Gershman, S. J. The hippocampus as a predictive
993 map. *Nat. Neurosci.* **20**, 1643–1653 (2017).

994 90. Dumas, J. A. What Is Normal Cognitive Aging? Evidence from Task-Based Functional
995 Neuroimaging. *Curr. Behav. Neurosci. Reports* **2**, 256–261 (2015).

996 91. Amado, C. *et al.* Neuroimaging results suggest the role of prediction in cross-domain
997 priming. *Sci. Rep.* **8**, 1–10 (2018).

998 92. Sperling, R. *et al.* Putting names to faces: Successful encoding of associative memories
999 activates the anterior hippocampal formation. *Neuroimage* **20**, 1400–1410 (2003).

1000 93. Alm, K. H., Rolheiser, T. & Olson, I. R. Inter-individual variation in fronto-temporal
1001 connectivity predicts the ability to learn different types of associations. *Neuroimage*
1002 **132**, 213–224 (2016).

1003 94. Pernet, C. *et al.* Best Practices in Data Analysis and Sharing in Neuroimaging using

MEEG. 1–64 (2018) doi:10.31219/osf.io/a8dhx.

95. Westfall, J. *PANGEA: Power ANalysis for GEneral Anova designs*. (2016).

96. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).

97. Ma, D. S., Correll, J. & Wittenbrink, B. The Chicago face database: A free stimulus set of faces and norming data. *Behav. Res. Methods* **47**, 1122–1135 (2015).

98. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).

99. López-García, D., Peñalver, J. M. G., Górriz, J. M. & Ruz, M. MVPAlab: A machine learning decoding toolbox for multidimensional electroencephalography data. *Comput. Methods Programs Biomed.* **214**, 106549 (2022).

100. López-García, D., Sobrado, A., Peñalver, J. M. G., Górriz, J. M. & Ruz, M. Multivariate Pattern Analysis Techniques for Electroencephalography Data to Study Flanker Interference Effects. *Int. J. Neural Syst.* **30**, (2020).

101. Keil, A. *et al.* Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology* **51**, 1–21 (2014).

102. Cohen, M. X. A better way to define and describe Morlet wavelets for time-frequency analysis. *Neuroimage* **199**, 81–86 (2019).

103. Love, J. *et al.* JASP: Graphical statistical software for common statistical designs. *J. Stat. Softw.* **88**, (2019).

104. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 1–28 (2008).

105. Walther, A. *et al.* Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage* **137**, 188–200 (2016).

106. Nili, H. *et al.* A Toolbox for Representational Similarity Analysis. *PLoS Comput. Biol.* **10**, (2014).

107. Bueno, F. D. & Cravo, A. M. Post-interval EEG activity is related to task-goals in temporal discrimination. *PLoS One* **16**, 1–19 (2021).

108. King, J. R. *et al.* Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *Neuroimage* **83**, 726–738 (2013).

109. Varoquaux, G. Cross-validation failure: Small sample sizes lead to large error bars. *Neuroimage* **180**, 68–77 (2018).

110. Singh, D. & Singh, B. Investigating the impact of data normalization on classification performance. *Appl. Soft Comput.* **97**, 105524 (2020).

111. King, J. & Dehaene, S. Characterizing the dynamics of mental representations : the temporal generalization method. **18**, 203–210 (2017).

112. González-García, C., Formica, S., Wisniewski, D. & Brass, M. Frontoparietal action-oriented codes support novel instruction implementation. *Neuroimage* **226**, (2021).