Neural oscillations and top-down connectivity are modulated by object-scene congruency

Ye Gu[1], Alexandra Krugliak[2] & Alex Clarke[1,3]

[1] Department of Psychology, University of Cambridge
[2] Cognition and Brain Sciences Unit, University of Cambridge
[3] Department of Psychology, University of Warwick

Abstract

The knowledge we have about how the world is structured is known to influence object recognition. One way this is demonstrated is through a congruency effect, where object recognition is faster and more accurate if items are presented in expected scene contexts. However, our understanding of the dynamic neural mechanisms that underlie congruency effects are under-explored. Using MEG, we examine how the congruency between an object and a prior scene results in changes in the oscillatory activity in the brain, which regions underpin this effect, and whether congruency results arise from top-down or bottom-up modulations of connectivity. We observed that prior scene information impacts the processing of visual objects in behaviour, neural activity and connectivity. Processing objects that were incongruent with the prior scene resulted in slower reaction times, increased low frequency activity in the ventral visual pathway, and increased top-down connectivity from the anterior temporal lobe to the posterior ventral temporal cortex. Our results reveal that the recurrent dynamics within the ventral visual pathway are modulated by the prior knowledge imbued by our surrounding environment, suggesting that the way we recognise objects is fundamentally linked to their context.

Keywords: object recognition, effective connectivity, anterior temporal lobe, semantics, context

Introduction

The objects that we see and recognise are not experienced in isolation. It is well established that the process of object recognition is influenced by the scene an object is found within. This is demonstrated by faster and more accurate recognition rates for objects within congruent scenes compared to incongruent scenes (Bar, 2004; Davenport & Potter, 2004; Greene et al., 2015; Oliva & Torralba, 2007; Palmer, 1975), whilst visual scenes can also aid the disambiguation of ambiguous objects (Brandman & Peelen, 2017). However, we currently lack details about the neural underpinnings and mechanisms that are modulated by the relationship between an object and where it is located, which would give us a better understanding of object recognition under more naturalistic circumstances.

Object-scene congruency effects are typically explored in paradigms where an object is either embedded or overlaid on the visual scene, with most studies testing neural activity using EEG. Such research has widely found that neural responses are modulated approximately 250 ms after seeing the image, with various EEG components showing a sensitivity to object-scene congruency including the N300, N400 and P600, which have been linked to perceptual, semantic and structural processing of the object-scene relationship (Chen et al., 2022; Draschkow et al., 2018; Ganis & Kutas, 2003; Kumar et al., 2021; Lauer et al., 2018; Mudrik et al., 2014; Võ & Wolfe, 2013). While the modulation of ERPs is the dominant source of evidence for congruency effects, we have limited knowledge about how oscillatory processes might be modulated when it comes to object and scene congruence. Studies using congruency manipulations of word pairs or during sentence reading suggest that theta activity increases when there is a semantic violation or incongruency between the context and a new item (Bastiaansen et al., 2005; Packard et al., 2020). Increases in theta could reflect a modulation of retrieval processes during the semantic processing of the stimulus (Klimesch et al., 2001), with prior work also linking theta activity patterns to the semantic processing of objects (Clarke, 2020; Clarke et al., 2018). However, strikingly little research exists concerning the oscillatory nature of object-scene congruency effects.

A further issue is that we have little evidence for which neural regions are modulated by the recognition of objects in congruent or incongruent scenes. While N400 effects during reading and listening to words have been localised to language sensitive regions in the inferior frontal gyrus, posterior middle temporal gyrus, and the anterior temporal lobe (Ghosh Hajra et al., 2018; Halgren et al., 2002; Lau et al., 2008; Maess et al., 2006; Nobre & McCarthy, 1995), equivalent effects for object-scene congruency would be predicted in regions along the ventral temporal lobe that do display object-scene congruency modulations in fMRI (Li et al., 2023; Rémy et al., 2014) and are associated with the visual perception and integration of objects and scenes (Bar & Aminoff, 2003; Brandman & Peelen, 2017). Going beyond this, an understanding of how scene knowledge influences object recognition processes requires that we test how connectivity is modulated, and whether top-down or bottom-up connections are impacted.

We explored these issues using MEG, where objects were preceded by a visual scene that was either congruent or not with the object (Figure 1a). Through analysing behaviour, MEG sensor activity, time frequency responses, source localised activity, and effective connectivity, we asked if the prior scene context influenced the processing of visual objects through changes in oscillatory activity, which neural regions underpin this effect, and finally, how connectivity dynamics are modulated within this neural system by congruency.

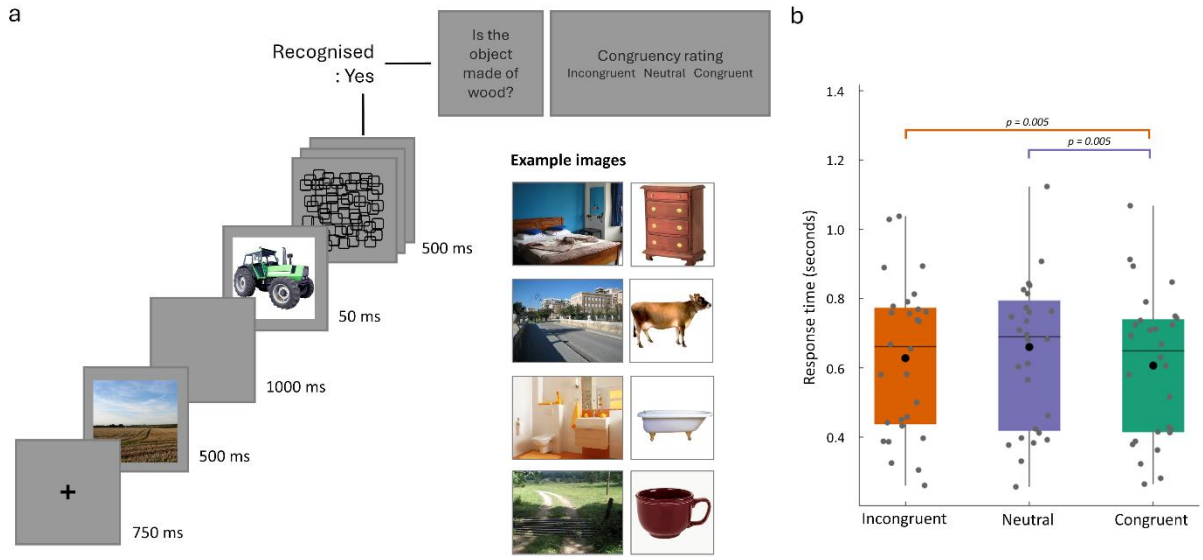

Figure 1. Experimental paradigm and behavioural performance. a) On each trial, a scene is presented followed by a backward masked visual object. Participants make a button press if they saw and knew what the object was. A catch question is used to ensure they recognised the object, and participants are then asked to rate if the prior scene was congruent with the object on a 3-point scale. Example images show the range of congruencies between the scenes and the objects. b) Behavioural response times for recognised objects, classified according to the congruency ratings. Congruent object-scene trials were significantly faster than both incongruent and neutral rated trials.

Results

During MEG recording, participants saw a visual scene depicting an indoor setting, an outdoor urban environment, or outdoor rural environment, that was followed by a briefly presented isolated object. All 150 different objects were presented three times, once following a more congruent scene (e.g. a bathtub in a bathroom), once following a more incongruent scene (e.g. a cow in a city street) and once following a scrambled image. Participants were required to press a button if they knew what the object was, and provide a rating for the congruency between the object and the prior scene (Figure 1a). Based on these participant congruency ratings, we sorted the trial into congruent, incongruent and neutral conditions. Through analysing behaviour and MEG responses, we sought to determine if congruency impacted object recognition, when in time the processing of visual objects was impacted, and if neural oscillations were also modulated by congruency. To go beyond effects at the level of MEG sensors, we also examined the neural localisation of congruency effects and how congruency impacted connectivity.

Behaviour

Analysis of the behavioural responses show that participants were able to recognize the briefly presented objects with a high degree of accuracy (mean = 96%, minimum 79%, maximum 100%). From the trials that were successfully recognised, we tested for an effect of congruency on response times using the condition labels given by individual participants (Figure 1b). A repeated measures ANOVA with factor congruency (congruent, neutral, incongruent) showed a significant effect of congruency ($F(1.18,31.8)=7.4$, $p=0.008$). Posthoc comparisons showed that congruent trials were responded to fastest compared to both incongruent ($t(27)=3.44$, $p=0.005$) and neutral

trials (t(27)=3.48, p=0.005), while no significant differences were observed between incongruent and neutral trials (t(27)=1.82, p=0.08). These results align with past research where objects are recognized faster when they are related to a congruent scene.

Table 1. Behaviour performance.

			95% Confidence Interval			
	Mean	Std error	Upper	Lower	Mean	Std error
	Response time (sec)				*Trials*	
Incongruent	0.627	0.041	0.713	0.542	123.3	4.17
Neutral	0.660	0.047	0.756	0.563	38.8	4.68
Congruent	0.606	0.040	0.688	0.524	117.5	2.84

MEG sensor-level analysis

Before testing our main research question, we first established that our manipulation of object-scene congruency evoked effects similar to N300/N400 effects widely seen in EEG studies, despite our temporal separation of the scene from the object. To ensure the effects related to each individual participants view of congruency, we created conditions based on the subjective judgements of each participant.

Our MEG analysis focused on the time period after the object was presented, testing for differences between the congruent and incongruent trials across MEG sensors and time points. A cluster-based permutation test revealed a significant effect where incongruent trials showed larger negative field strength compared to congruent trials over the left anterior sensors between approximately 500 and 700 ms after object onset (cluster p = 0.036; Figure 2a,b,c). Whilst the significant cluster might look slightly later than many N300/N400 ERP effects, examination of the time-series suggests the effect might be temporally more extended, from around 200 to 1000 ms (Figure 2c). This shows that our paradigm and data illicit congruency effects during the processing and integration of objects with the prior scene.

One of our main research questions was how are neural oscillations modulated by object-scene congruency. Whilst modulations of activity seen though ERPs and ERFs have been observed, far less attention has been paid to modulations of oscillations. To address this, we calculated the power at each frequency between 3 and 40 Hz and for time points between 0 and 1250 ms for each trial and at every MEG sensor. This allowed us to test if total power at different frequencies was modulated by the congruency between the object and prior scene, and over which sensors any effect was located.

We used a cluster-based permutation test for an effect of congruency across MEG sensors, time and frequency. This showed a significant modulation of low frequency activity, where more power was seen for incongruent trials compared to congruent trials, with the effect peaking around 8 Hz, over left posterior sensors, and between approximately 500 to 900 ms after object onset (cluster p = 0.029; Figure 2d,e,f). Again, this appears part of a wider divergence between the conditions from near 200 ms to over 1 second, and across frequencies near 6-12 Hz in the high theta and alpha range (Figure 2f).

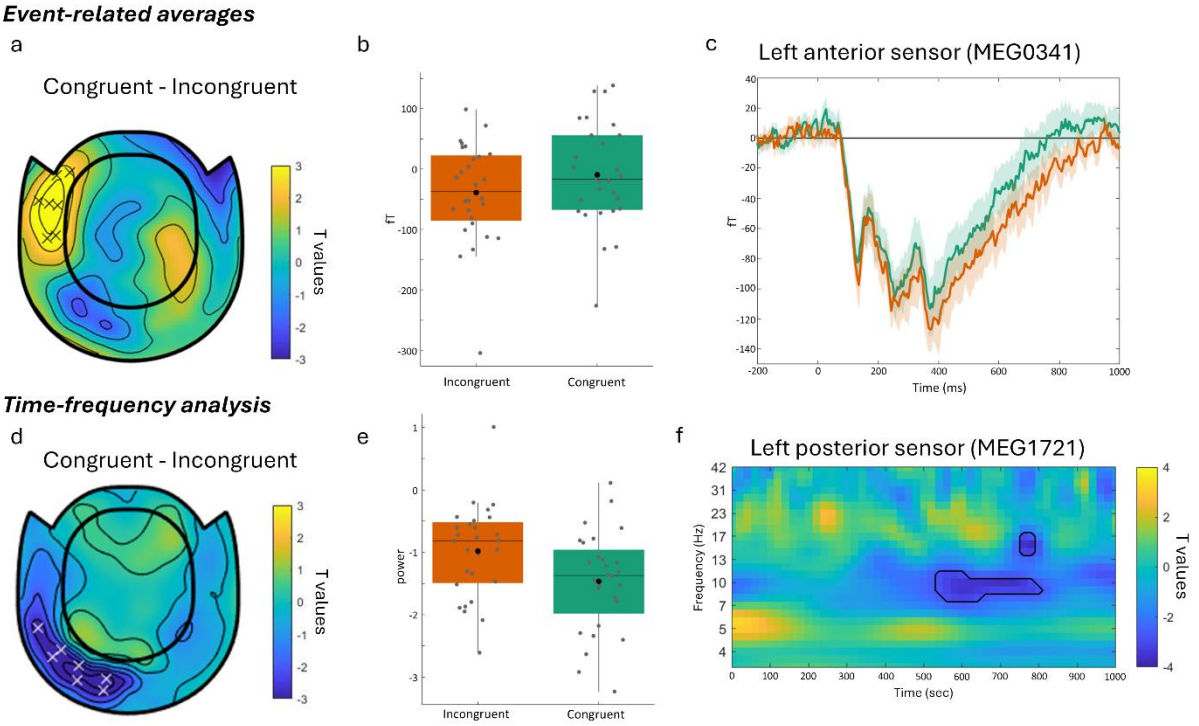

Figure 2. Sensor-level congruency effects. a) Event-related average analysis showing representative topography of the congruency effect. Data shows effect at 600 ms, with 'x' indicating sensors showing a significant effect. b) Data averaged across the significant cluster, plotted for congruent and incongruent trials. Each data-point is one participant. c) Event-related field activity at one sensor for the congruent and incongruent conditions. d) Time-frequency analysis showing representative topography of the congruency effect, plotted at 8.6 Hz and 560 ms. 'x' indicates sensors showing a significant effect. e) Data extracted from the significant cluster plotted for congruent and incongruent trials. Each data-point is one participant. f) Time frequency representation for a posterior sensor showing t-statistics for the contrast of congruent – incongruent trials. Black line shows the extent of the significant cluster at this sensor.

Neural network modulation by congruency

In order to understand which neural regions were underpinning the low frequency modulation by congruency, we performed a source localisation procedure incorporating the individuals anatomical MRI data. We used an LCMV beamformer to determine how activity was distributed across the brain for the congruent and incongruent conditions for each participant, with activity estimated for the time window between 200 and 1000 ms, and between 6 to 12 Hz. This revealed that objects that are incongruent with the preceding scene were associated with significantly increased low frequency activity in the left ventral visual pathway, from the posterior ventral temporal cortex (VTC) through to the anterior temporal lobe (ATL) (voxelwise $p < 0.01$, cluster $p < 0.05$; Figure 3).

Finally, we asked if connectivity in this network across the anterior and posterior ventral temporal lobe was modulated by congruency. We used dynamic causal modelling (DCM) of cross-spectral densities to determine the most likely directional-network model that explains our low-frequency modulations according to congruency. Our network consisted of two regions in the left ventral visual pathway, defined by peaks in our source analysis above – the left VTC and left ATL. We

defined this network as having feedforward and feedback anatomical connections between them, and we tested if these were modulated by congruency or not.

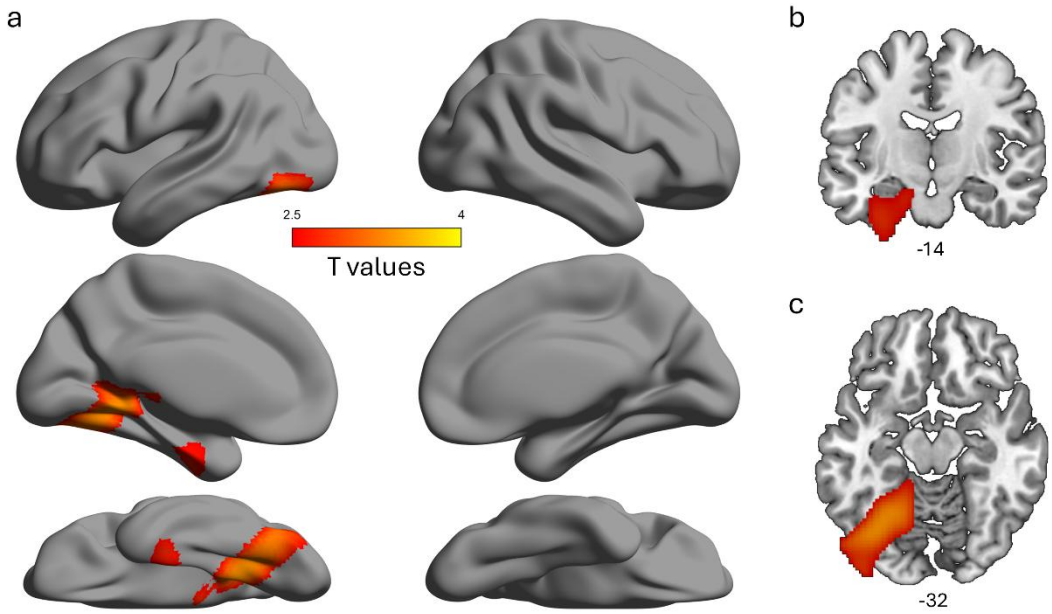

Figure 3. Source localisation of the low frequency congruency effect showing increased activity for the incongruent condition compared to the congruent condition, shown a) on an inflated surface, b) coronal view and c) axial view. Thresholded at $p < 0.01$, cluster $p < 0.05$

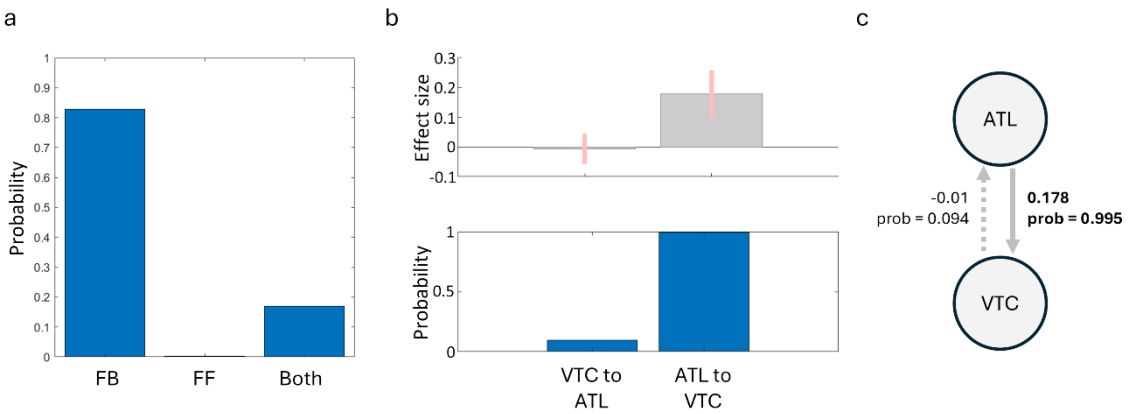

Figure 4. Congruency modulates feedback connectivity in the ventral visual pathway. a) Model probabilities when a given modulatory parameter is included. b) Bayesian model average parameters weighted by the model probabilities, showing the modulatory effect of the feedforward and feedback connection in explaining the congruency effect. Error bars show 95% confidence intervals. c) Summary of the modulatory effect of congruency, showing increased feedback from the ATL for the incongruent condition compared to the (baseline) congruent condition. Bold connections have probability greater than 99%.

We used the second-level PEB framework to assess the probability that the feedforward or feedback connection was modulated by congruency. After estimating the connectivity parameters of the full model, we compared the full and reduced models using Bayesian Model Comparison and Reduction, where each of the parameters is switched on or off to determine the best network model of the data. This showed that the model that included a feedback modulatory connection had the highest posterior probability (83%) with low probability for models with either feedforward modulations only, or both feedforward and feedback modulations (Figure 4a).

To better understand the parameters, we performed Bayesian Model Averaging across the three model-types, where the parameters are weighted according to the probabilities of the model. This again indicates that while there is little evidence that feedforward connections were modulated by congruency (9.4% probability), there was strong evidence that feedback connectivity was increased for the incongruent condition compared to the congruent condition (99.5% probability; Figure 4b,c). This shows that when objects are incongruent with the prior scene, there is an increase in feedback connectivity from the ATL to the VTC (Figure 4c).

Discussion

We aimed to reveal how prior knowledge, established through a visual scene context, influenced subsequent object processing. Using MEG, we tested whether scene-object congruency impacted oscillatory activity, where this activity localised to in the brain, and if the connectivity dynamics of this system were modulated in a top-down or bottom-up fashion. We observed that prior scene information clearly impacts the processing of visual objects through behaviour, neural activity and connectivity, finding that processing objects that are incongruent with the prior scene leads to slowed reaction times, an increase in low frequency activity in the ventral visual pathway, and an increase in top-down connectivity from the ATL to the VTC. These results suggest that the neural processing and mechanisms underlying object recognition do not occur in isolation, but are sensitive to the environment visual objects are experienced within.

While models of object recognition are increasingly incorporating information about temporal dynamics, other research shows that the prior cognitive state of the brain influences subsequent visual processing (Aitken et al., 2020; de Lange et al., 2013; Dehaghani & Zarei, 2025; Mayer et al., 2016). For instance, when an auditory cue is predictive of an upcoming visual stimulus, the early sensory processing of that stimulus is modulated by the prediction cue (Aitken et al., 2020). In our paradigm, any predictions or expectations generated by the initial visual scene could lead to modulations of object processing, as seen through congruency between the scene and the object. We observed that top-down connectivity from the ATL to the VTC increased when the object was incongruent with the prior scene. Using DCM for cross-spectral densities, we observed that a model where feedback connectivity increased for incongruent compared to congruent conditions was the best explanation for the observed changes in low frequency activity in the ventral visual pathway. Top-down connectivity to the VTC during object recognition has been observed under challenging recognition conditions such as visual occlusion, degraded images and manipulations of consciousness (Bar et al., 2006; Ganis et al., 2007; Mohsenzadeh et al., 2018; Rajaei et al., 2019; Schendan & Ganis, 2015; Schendan & Stern, 2008; Wyatte et al., 2014). Our results further show that our prior expectations about what object might appear also modulates top-down effects, here within the ventral visual pathway, and we suggest that this

reflects the semantic processing of the objects which either do, or do not benefit from prior scene information. While it might be expected that top-down predictions about possible objects given a scene context come from frontal regions (Bar, 2004; Trapp & Bar, 2015), it is worth keeping in mind that our visual scenes afford predictions to occur in both our congruent and incongruent conditions, meaning that within temporal lobe connectivity itself could be biased by schema-consistent predictions driven by frontal connectivity upon seeing the scene. Furthermore, the task in our paradigm required participants to understand what the object was, not just detect a visual image, which is a semantic task, and so changes in the recognition process will more likely engage the anterior temporal lobe as a result.

Whilst our research does not shed light on whether congruency effects are ultimately the result of changes to the way objects are recognized, or how objects are integrated with the scene context, our results add mechanistic detail to how our prior knowledge influences the ongoing processing of objects. One of our principal findings is an increase in feedback activity from the ATL to the VTC when objects are preceded by an incongruent scene. An increasing body of work acknowledges that feedback within the ventral visual pathway increases under more challenging object recognition conditions (Campo et al., 2013; Chan et al., 2011; Clarke et al., 2011; Poch et al., 2015; Rajaei et al., 2019; Schendan & Ganis, 2012; Wyatte et al., 2014). Feedback from the ATL to the VTC is linked to more demanding semantic processing of objects (Clarke et al., 2011), a shift from visual to semantic processes (Clarke et al., 2018), and linked to recognition performance (von Seth et al., 2023). TMS evidence further suggests that top-down effects from the ATL support the semantic recognition of objects (Chiou & Lambon Ralph, 2016). We argue that the modulation observed here, within the ventral visual pathway, is likely to be concerned with the semantic processing of the visual objects.

If the increase in feedback we see here has a role in the semantic processing of the object given the prior context, then this would predict that we should observe changes in the timing of when semantic representations are accessed based on congruency (Truman & Mudrik, 2018). There is abundant evidence that recognition reaction times are slower for objects that are incongruent with the scene, but this does not address the potential differences in the dynamics of when semantic representations become accessed following a congruent or incongruent scene. Recent EEG evidence does speak to this issue, where it was shown that semantic representations are seemingly evoked at similar times for all objects, both in congruent and incongruent scene contexts, but that semantic representations persisted for a longer duration when the object was incongruent with the scene (Krugliak et al., 2023). Overall, this suggests that any recurrent dynamics in the ventral visual pathway during object recognition are further increased when objects do not benefit from prior knowledge indicated by the scene.

Our results also highlight that changes in activity in the high theta and alpha frequencies reflect the relationship of the object and scene. Theta/alpha power was higher for incongruent trials compared to congruent trials, which is in line with other recent evidence showing increased power for more unexpected objects that were situated in a real-world environment (Nicholls et al., 2025). However, most research on semantic congruency effects and oscillations comes from the language domain, where it is consistently shown that theta activity increases in response to an unexpected word or a semantic violation (Bastiaansen et al., 2005, 2008; Hald et al., 2006; Packard et al., 2020; Wang et al., 2012). The effects we see here are consistent with these findings, albeit in a visual object recognition domain and with effects extending into alpha frequencies. Further, we show that these low frequency modulations localise to the ventral

temporal lobe. Low frequency effects in the ventral visual pathway have been shown to reflect the higher-level visual or lexico-semantic properties of objects (Clarke, 2020; Clarke et al., 2018; Reddy et al., 2021), whilst studies across several domains have linked increases in low frequency activity with memory retrieval processes (Freunberger et al., 2008; Herweg et al., 2016; Klimesch et al., 2001). Together, this argues that the increase in low frequency activity we see reflects the additional semantic retrieval demands needed when the object recognition process does not benefit from information in the prior scene context.

Low frequency neural signals are hypothesised to reflect long-range communication between regions (Colgin, 2013; Herweg et al., 2016; Sauseng et al., 2010). Alpha activity in particular is associated with top-down signalling during visual perception (Bar et al., 2006; Chen et al., 2023; Kerkoerle et al., 2014; Mayer et al., 2016; Michalareas et al., 2016; Stecher et al., 2025), which might align with our connectivity data showing both an increase in feedback from the ATL, and higher low frequency power during incongruent trials. An alternative, but not incompatible view of the low frequency modulations, is that low frequency activity decreases for objects that are congruent with the prior scene, with the decreased activity reflecting the successful access to semantic representations. Further studies linking specific frequencies to the representational content of neural signals across the conditions could seek to uncover this.

Where we are in the world clearly shapes the cognitive processes associated with recognizing an object, with our research highlighting the role low frequency oscillations play, and that top-down connectivity functions to support object processing when those objects are increasingly unexpected.

Methods

Participants

Thirty-one participants (12 males, age range 18-35 years) took part in the study. All participants were right-handed and had normal or corrected-to-normal vision. Three participants were excluded due to poor behavioural performance leading to a final sample size of 28. All participants gave written informed consent. The study was approved by the Cambridge University Psychology Research Ethics Committee.

Stimuli

A total of 450 colour images were used in the study, including 150 scene images, 150 scrambled scene images, and 150 object images. The scenes images were obtained from Lauer et al. (2018) the SUN397 scene image database (Xiao et al., 2010), and internet searches using Google Image Search. Objects were shown in colour, isolated on a white background, from the Hemera photo object image set or Clarke et al., (2018).

Each scene image was paired with two object images, one that was congruent with the scene, and one that was incongruent. Objects only appeared once with a congruent scene and once with an incongruent scene, meaning that each scene was paired with two objects, and each object was paired with two scenes. The object-scene pairings were determined with a pretest using a separate group of 37 participants and a larger range of scene and object images. The pairings were initially composed by the research team prior to the pretest. Participants rated each

object-scene pairing based on how likely they were to encounter the object within the scene using a 5-point scale (1 = very unlikely, 2 = unlikely, 3 = neutral, 4 = likely, 5 = very likely). Congruent pairings were defined as those with a mean rating above 3.5 and incongruent pairings with a mean rating of less than 1.5. If a pairing did not have a clear congruent or incongruent rating, the images were adjusted and rated again, resulting in 150 congruent pairs and 150 incongruent pairs where each scene and object appeared once in each list. Importantly, the objects did not appear in the scenes they were paired with.

The scrambled scene image that was paired with each object was created from the two scene images paired with that object. The two scene images were divided into small squares and a new image was created by combining half the squares from one scene with half from the other scene. The squares were flipped and randomly shuffled to create a scrambled image. The condition with scrambled images is not analysed in this study.

Paradigm

Participants performed an object recognition task, where an object image followed a scene that was either congruent or incongruent with the object, or a scrambled scene image. Each trial began with a fixation cross lasting 750 ms before a scene image was shown for 500 ms. This was followed by a blank grey screen lasting 1 second. An object was then displayed in the centre of the screen for 50 ms followed by a backwards mask lasting 200 ms. The mask was made of dynamically changing overlapping squares with each pattern lasting 50 ms (Figure 1a).

Participants pressed a button using their right index finger if they recognized the object and knew what the object was (not just that they saw something), or using their right middle finger if they did not recognize the object. For objects that were recognized, they next answered a yes/no question about the object's characteristics (e.g., ''Is the object edible?'', or ''Can you buy the object in a supermarket?''), followed by a congruency rating for the object and scene pair. Congruency ratings were made using the left hand and with 3 response options (unlikely, neutral, likely) to the question 'How likely would you expect to encounter this object in this context?'.

If the participant did not recognize the object, no following questions were asked. Each object was shown three times, on each occasion being preceded by the congruent, incongruent or scrambled scene image, with a total of 450 trials. Each participant saw the 450 trials in a randomized order. The experiment was controlled using MATLAB (MathWorks, Natick, MA) and Psychtoolbox (Brainard, 1997).

MEG/MRI acquisition

Continuous MEG data were recorded using a whole-head 306 channel (102 magnetometers, 204 planar gradiometers) TRIUX neo system (MEGIN, Espoo, Finland) at the MRC Cognition and Brain Sciences Unit, Cambridge. Five head-position indicator (HPI) coils were used to record the head position (every 200 ms) within the MEG helmet and eye movements and blinks were monitored using the electrooculogram (EOG) recorded bipolarly through electrodes placed above and below the right eye (vertical) and at the outer canthi (horizontal). The participants' head shape was digitally recorded using a 3D digitizer (Fastrak Polhemus, Inc., Colchester, VA, USA) along with the positions of the EOG electrodes, HPI coils, and fiducial points (nasion, left and right

periarticular). MEG signals were recorded at a sampling rate of 1000 Hz, with a high-pass filter of 0.03 Hz.

To facilitate source reconstruction, high-resolution structural T1-weighted MPRAGE scans were acquired during a separate session with a Siemens 3T Tim Trio scanner (Siemens Medical Solutions, Camberley, UK) located at the MRC Cognition and Brain Sciences Unit, using a 3D MPRAGE sequence (field-of-view 256 mm × 240 mm × 160 mm, matrix dimensions 256 × 240 × 160, 1 mm isotropic resolution, TR = 2250 msec, TI = 900 msec, TE = 2.99 msec, flip angle 9°).

MEG processing

The raw data were Maxfiltered using MNE-python (https://mne.tools/). Static bad channels were detected and subsequently reconstructed by interpolating neighbouring channels, as were bad channels containing long periods of high amplitude or noisy signals. The temporal extension of the signal-space separation (SSS) technique (Taulu et al., 2005) was applied to the data every 4s to segregate the signals originating from within the participants' heads from those generated by external sources of noise, along with head movement compensation and transformation of the head position to a common position.

The resulting MEG data were processed using SPM12 (Wellcome Institute of Imaging Neuroscience, London, UK, www.fil.ion.ucl.ac.uk/spm), and were downsampled to 250 Hz, low-pass filtered at 100 Hz, stop-pass filtered between 48 and 52 Hz, and high-pass filtered at 0.1 Hz in forward and reverse directions using a fifth-order bidirectional Butterworth digital filter. The data were then epoched from -2500 to 2000 ms relative to the onset of the object (or -1000 to 3500 ms relative to the onset of the scene). Independent components analysis (ICA) was implemented using EEGLab (Delorme & Makeig, 2004), and components whose time series showed a Pearson's correlation greater than 0.4 with horizontal or vertical EOG channels were labelled for rejection. Additional manual inspections identified components involving apparent rhythmic cardiac noise, before MEG signals were reconstructed using the montage function to retain the non-identified components. Epoch were baseline corrected using the 200 ms period prior to the object, and trials that were not recognized were rejected (mean = 18, min 0 trials, max 74 trials, SD = 19.2 trials).

Our statistical analysis of the data used Fieldtrip (Oostenveld et al., 2010). The SPM files were converted into fieldtrip format (spm2fieldtrip.m), trials were re-labelled according to the participant ratings and event-related averages calculated for each participant. Our analyses only used the congruent and incongruent trials as there were fewer trials rated as neutral. To test for congruency effects across sensors and time-points, we used a cluster-based permutation test, focusing on time points between 0 and 1250 ms with a cluster-forming alpha of 0.01, 1000 randomisations, and a minimum number of cluster-forming channels set to 2.

We also used an analysis of time frequency representations to test for congruency effects. Time frequency representations were calculated for each trial using Morlet wavelets of length 5 cycles at 40 logarithmically spaced frequencies between 2 and 100 Hz. A baseline correction was applied to convert the data into dB using data between -500 and -200 ms prior to the onset of the object, and the data was averaged across trials. To test for congruency effects across sensors, frequencies and time-points, we used a cluster-based permutation test, focusing on time points between 0 and 1250 ms and frequencies between 3 and 40 Hz. We used a cluster-forming alpha of 0.01, 1000 randomisations, and a minimum number of cluster-forming channels set to 2.

Source localization

To localize the time frequency effects we used a beamforming approach in SPM12. Individual participants MRI images were segmented and spatially normalised to an MNI template brain consisting of 20,484 vertices, which was inverse normalised to the individuals specific MRI space. MEG sensors were co-registered to the MRI image using the three fiducial points and the additional headpoints. A single shell forward model was used. A linearly constrained minimum variance (LCMV) beamformer was implemented using the DaiSS toolbox. Only the magnetometers were included in the inversion, with a frequency range set between 6 and 12 Hz, a time window of 200 to 1000 ms, and a grid size of 5 mm. Voxel-based images showing the source activity for each participant and the two conditions of interest, congruent and incongruent, were created in MNI space, and smoothed using a 10 mm FWHM kernel. Source activity values were log transformed. These images were entered into a paired-samples T-test, assessed using a voxel-wise threshold of $p < 0.01$, and a cluster threshold of $p(FWE) < 0.05$. Images are displayed on an inflated surface using BrainNet viewer (Xia et al., 2013).

Effective connectivity

We used DCM for cross-spectral density (Friston et al., 2012) to assess the directed effective connectivity between regions identified in the beamformer analysis outlined above. Using the peak coordinates from the beamformer analysis, we specified a two node network consisting of the left posterior temporal cortex (MNI coordinate -18, -50, -6) and left anterior temporal lobe (MNI coordinate –32, -10, -34). The network connections set out by the A matrix included bi-direction connections, which were both modulated in the B matrix. The frequency range was set to 6-12 Hz, and the time window between 0 and 1000 ms after the object appeared, with the congruent condition treated as a baseline. To make inferences about how feedforward or feedback connectivity was modulated by congruency, we assessed the group-level modulations using the PEB framework, using Bayesian model reduction (BMR) and Bayesian model averaging (BMA) (Zeidman et al., 2019).

Behavioural stats

For each participant, we collected object recognition reaction times and the rated congruency between the object and the preceding scene. Reaction times for successful recognition trials were classified into congruent, incongruent and neutral conditions based on the participants own ratings. Repeated measures ANOVA and follow-up tests were run using JASP 0.18.3.0. Greenhouse-giesser correction to the degrees of freedom was used for violations of sphericity, and multiple follow-up t-tests were corrected for multiple comparisons using the Holm-Bonferroni correction.

Acknowledgments

This research was funded in whole, or in part, by the Wellcome Trust (grant no. 211200/Z/18/Z to AC) and the Isaac Newton Trust. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

Aitken, F., Turner, G., & Kok, P. (2020). Prior Expectations of Motion Direction Modulate Early Sensory Processing. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *40*(33), 6389–6397. https://doi.org/10.1523/JNEUROSCI.0537-20.2020

Bar, M. (2004). Visual objects in context. *Nat Rev Neurosci*, *5*(8), Article 8. https://doi.org/10.1038/nrn1476

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, *38*(2), Article 2.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hamalainen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), Article 2.

Bastiaansen, M. C. M., Oostenveld, R., Jensen, O., & Hagoort, P. (2008). I see what you mean: Theta power increases are involved in the retrieval of lexical semantic information. *Brain and Language*, *106*(1), 15–28. https://doi.org/10.1016/j.bandl.2007.10.006

Bastiaansen, M. C. M., van der Linden, M., ter Keurs, M., Dijkstra, T., & Hagoort, P. (2005). Theta responses are involved in lexical-semantic retrieval during language processing. *Journal of Cognitive Neuroscience*, *17*(3), Article 3.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897x00357

Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing revealed by human fMRI and MEG decoding. *Journal of Neuroscience*, 0582–17. https://doi.org/10.1523/JNEUROSCI.0582-17.2017

Campo, P., Poch, C., Toledano, R., Igoa, J. M., Belinchon, M., Garcia-Morales, I., & Gil-Nagel, A. (2013). Anterobasal temporal lobe lesions alter recurrent functional connectivity within the ventral pathway during naming. *The Journal of Neuroscience*, *33*(31), Article 31.

Chan, A. M., Baker, J. M., Eskandar, E., Schomer, D., Ulbert, I., Marinkovic, K., Cash, S. S., & Halgren, E. (2011). First-pass selectivity for semantic categories in human anteroventral temporal cortex. *Journal of Neuroscience*, *31*(49), Article 49.

Chen, L., Cichy, R. M., & Kaiser, D. (2022). Semantic scene-object consistency modulates N300/400 EEG components, but does not automatically facilitate object representations. *Cerebral Cortex*, *32*(16), 3553–3567. https://doi.org/10.1093/cercor/bhab433

Chen, L., Cichy, R. M., & Kaiser, D. (2023). Alpha-frequency feedback to early visual cortex orchestrates coherent naturalistic vision. *Science Advances*, *9*(45), eadi2321. https://doi.org/10.1126/sciadv.adi2321

Chiou, R., & Lambon Ralph, M. A. (2016). The anterior temporal cortex is a primary semantic source of top-down influences on object recognition. *Cortex*, *79*, 75–86. https://doi.org/10.1016/j.cortex.2016.03.007

Clarke, A. (2020). Dynamic activity patterns in the anterior temporal lobe represents object semantics. *Cognitive Neuroscience*, *11*(3), Article 3. https://doi.org/10.1080/17588928.2020.1742678

Clarke, A., Devereux, B. J., & Tyler, L. K. (2018). Oscillatory Dynamics of Perceptual to Conceptual Transformations in the Ventral Visual Pathway. *Journal of Cognitive Neuroscience*, *30*(11), Article 11. https://doi.org/10.1162/jocn_a_01325

Clarke, A., Taylor, K. I., & Tyler, L. K. (2011). The evolution of meaning: Spatiotemporal dynamics of visual object recognition. *Journal of Cognitive Neuroscience*, *23*(8), Article 8.

Colgin, L. L. (2013). Mechanisms and functions of theta rhythms. *Annual Review of Neuroscience*, *36*, 295–312. https://doi.org/10.1146/annurev-neuro-062012-170330

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), Article 8. https://doi.org/10.1111/j.0956-7976.2004.00719.x

de Lange, F. P., Rahnev, D. A., Donner, T. H., & Lau, H. (2013). Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(4), Article 4. https://doi.org/10.1523/JNEUROSCI.1094-12.2013

Dehaghani, N. S., & Zarei, M. (2025). Pre-stimulus activities affect subsequent visual processing: Empirical evidence and potential neural mechanisms. *Brain and Behavior*, *15*(2), e3654. https://doi.org/10.1002/brb3.3654

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

Draschkow, D., Heikel, E., Võ, M. L.-H., Fiebach, C. J., & Sassenhagen, J. (2018). No evidence from MVPA for different processes underlying the N300 and N400 incongruity effects in object-scene processing. *Neuropsychologia*, *120*, 9–17. https://doi.org/10.1016/j.neuropsychologia.2018.09.016

Freunberger, R., Klimesch, W., Griesmayr, B., Sauseng, P., & Gruber, W. (2008). Alpha phase coupling reflects object recognition. *NeuroImage*, *42*(2), 928–935. https://doi.org/10.1016/j.neuroimage.2008.05.020

Friston, K. J., Bastos, A., Litvak, V., Stephan, K. E., Fries, P., & Moran, R. J. (2012). DCM for complex-valued data: Cross-spectra, coherence and phase-delays. *NeuroImage*, *59*(1), 439–455. https://doi.org/10.1016/j.neuroimage.2011.07.048

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, *16*(2), Article 2. https://doi.org/10.1016/S0926-6410(02)00244-6

Ganis, G., Schendan, H. E., & Kosslyn, S. M. (2007). Neuroimaging evidence for object model verification theory: Role of prefrontal control in visual object categorization. *Neuroimage*, *34*, 384–398.

Ghosh Hajra, S., Liu, C. C., Song, X., Fickling, S. D., Cheung, T. P. L., & D'Arcy, R. C. N. (2018). Multimodal characterization of the semantic N400 response within a rapid evaluation brain vital sign framework. *Journal of Translational Medicine*, *16*(1), 151. https://doi.org/10.1186/s12967-018-1527-2

Greene, M. R., Botros, A. P., Beck, D. M., & Fei-Fei, L. (2015). What you see is what you expect: Rapid scene understanding benefits from prior experience. *Attention, Perception, & Psychophysics*, *77*(4), 1239–1251. https://doi.org/10.3758/s13414-015-0859-8

Hald, L. A., Bastiaansen, M. C. M., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language*, *96*(1), 90–105. https://doi.org/10.1016/j.bandl.2005.06.007

Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., & Dale, A. M. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage*, *17*(3), 1101–1116. https://doi.org/10.1006/nimg.2002.1268

Herweg, N. A., Apitz, T., Leicht, G., Mulert, C., Fuentemilla, L., & Bunzeck, N. (2016). Theta-Alpha Oscillations Bind the Hippocampus, Prefrontal Cortex, and Striatum during Recollection: Evidence from Simultaneous EEG–fMRI. *Journal of Neuroscience*, *36*(12), 3579–3587. https://doi.org/10.1523/JNEUROSCI.3629-15.2016

Kerkoerle, T. van, Self, M. W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., Togt, C. van der, & Roelfsema, P. R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proceedings of the National Academy of Sciences*, *111*(40), Article 40. https://doi.org/10.1073/pnas.1402773111

Klimesch, W., Doppelmayr, M., Stadler, W., Pöllhuber, D., Sauseng, P., & Röhm, D. (2001). Episodic retrieval is reflected by a process specific increase in human electroencephalographic theta activity. *Neuroscience Letters*, *302*(1), Article 1. https://doi.org/10.1016/S0304-3940(01)01656-1

Krugliak, A., Draschkow, D., Võ, M. L.-H., & Clarke, A. (2023). Semantic object processing is modulated by prior scene context. *Language, Cognition and Neuroscience*, *0*(0), 1–10. https://doi.org/10.1080/23273798.2023.2279083

Kumar, M., Federmeier, K. D., & Beck, D. M. (2021). The N300: An Index for Predictive Coding of Complex Visual Objects and Scenes. *Cerebral Cortex Communications*, *2*(2), Article 2. https://doi.org/10.1093/texcom/tgab030

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews. Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532

Lauer, T., Cornelissen, T. H. W., Draschkow, D., Willenbockel, V., & Võ, M. L.-H. (2018). The role of scene summary statistics in object recognition. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-32991-1

Li, C., Ficco, L., Trapp, S., Rostalski, S.-M., Korn, L., & Kovács, G. (2023). The effect of context congruency on fMRI repetition suppression for objects. *Neuropsychologia*, *188*, 108603. https://doi.org/10.1016/j.neuropsychologia.2023.108603

Maess, B., Herrmann, C. S., Hahne, A., Nakamura, A., & Friederici, A. D. (2006). Localizing the distributed language network responsible for the N400 measured by MEG during auditory sentence processing. *Brain Research*, *1096*(1), 163–172. https://doi.org/10.1016/j.brainres.2006.04.037

Mayer, A., Schwiedrzik, C. M., Wibral, M., Singer, W., & Melloni, L. (2016). Expecting to See a Letter: Alpha Oscillations as Carriers of Top-Down Sensory Predictions. *Cerebral Cortex*, *26*(7), 3146–3160. https://doi.org/10.1093/cercor/bhv146

Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J.-M., Kennedy, H., & Fries, P. (2016). Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas. *Neuron*, *89*(2), 384–397. https://doi.org/10.1016/j.neuron.2015.12.018

Mohsenzadeh, Y., Qin, S., Cichy, R. M., & Pantazis, D. (2018). Ultra-Rapid serial visual presentation reveals dynamics of feedforward and feedback processes in the ventral visual pathway. *eLife*, *7*, e36329. https://doi.org/10.7554/eLife.36329

Mudrik, L., Shalgi, S., Lamy, D., & Deouell, L. Y. (2014). Synchronous contextual irregularities affect early scene processing: Replication and extension. *Neuropsychologia*, *56*, 447–458. https://doi.org/10.1016/j.neuropsychologia.2014.02.020

Nicholls, V. I., Krugliak, A., Alsbury-Nealy, B., Gramann, K., & Clarke, A. (2025). *Contextual Expectations in the Real-World Modulate Low-Frequency Neural Oscillations* (p. 2024.05.30.596613). bioRxiv. https://doi.org/10.1101/2024.05.30.596613

Nobre, A. C., & McCarthy, G. (1995). Language-related field potentials in the anterior-medial temporal lobe: II. Effects of word type and semantic priming. *The Journal of Neuroscience*, *15*(2), 1090–1099.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), Article 12. https://doi.org/10.1016/j.tics.2007.09.009

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2010). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, e156869. https://doi.org/10.1155/2011/156869

Packard, P. A., Steiger, T. K., Fuentemilla, L., & Bunzeck, N. (2020). Neural oscillations and event-related potentials reveal how semantic congruence drives long-term memory in both young and older humans. *Scientific Reports*, *10*(1), 9116. https://doi.org/10.1038/s41598-020-65872-7

Palmer, tephen E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*(5), Article 5. https://doi.org/10.3758/BF03197524

Poch, C., Garrido, M. I., Igoa, J. M., Belinchón, M., García-Morales, I., & Campo, P. (2015). Time-Varying Effective Connectivity during Visual Object Naming as a Function of Semantic Demands. *The Journal of Neuroscience*, *35*(23), Article 23. https://doi.org/10.1523/JNEUROSCI.4888-14.2015

Rajaei, K., Mohsenzadeh, Y., Ebrahimpour, R., & Khaligh-Razavi, S.-M. (2019). Beyond core object recognition: Recurrent processes account for object recognition under occlusion. *PLOS Computational Biology*, *15*(5), Article 5. https://doi.org/10.1371/journal.pcbi.1007001

Reddy, L., Cichy, R. M., & VanRullen, R. (2021). Representational Content of Oscillatory Brain Activity during Object Recognition: Contrasting Cortical and Deep Neural Network Hierarchies. *eNeuro*, *8*(3). https://doi.org/10.1523/ENEURO.0362-20.2021

Rémy, F., Vayssière, N., Pins, D., Boucart, M., & Fabre-Thorpe, M. (2014). Incongruent object/context relationships in visual scenes: Where are they processed in the brain? *Brain and Cognition*, *84*(1), 34–43. https://doi.org/10.1016/j.bandc.2013.10.008

Sauseng, P., Griesmayr, B., Freunberger, R., & Klimesch, W. (2010). Control mechanisms in working memory: A possible function of EEG theta oscillations. *Neuroscience & Biobehavioral Reviews*, *34*(7), 1015–1022. https://doi.org/10.1016/j.neubiorev.2009.12.006

Schendan, H. E., & Ganis, G. (2012). Electrophysiological potentials reveal cortical mechanisms for mental imagery, mental simulation, and grounded (embodied) cognition. *Frontiers in Psychology*, *3*(329), Article 329.

Schendan, H. E., & Ganis, G. (2015). Top-down modulation of visual processing and knowledge after 250 ms supports object constancy of category decisions. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01289

Schendan, H. E., & Stern, C. E. (2008). Where vision meets memory: Prefrontal-posterior networks for visual object constancy during categorization and recognition. *Cerebral Cortex*, *18*, 1695–1711.

Stecher, R., Cichy, R. M., & Kaiser, D. (2025). Decoding the rhythmic representation and communication of visual contents. *Trends in Neurosciences*, *48*(3), 178–188. https://doi.org/10.1016/j.tins.2024.12.005

Taulu, S., Simola, J., & Kojola, M. (2005). Applications of the signal space separation method. *IEEE Transactions on Signal Processing*, *53*(9), Article 9.

Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition. *Annals of the New York Academy of Sciences*, *1339*(1), Article 1. https://doi.org/10.1111/nyas.12680

Truman, A., & Mudrik, L. (2018). Are incongruent objects harder to identify? The functional significance of the N300 component. *Neuropsychologia*, *117*, 222–232. https://doi.org/10.1016/j.neuropsychologia.2018.06.004

Võ, M. L.-H., & Wolfe, J. M. (2013). Differential Electrophysiological Signatures of Semantic and Syntactic Scene Processing. *Psychological Science*, *24*(9), Article 9. https://doi.org/10.1177/0956797613476955

von Seth, J., Nicholls, V. I., Tyler, L. K., & Clarke, A. (2023). *Recurrent connectivity supports higher-level visual and semantic object representations in the brain* (p. 2023.03.02.530768). bioRxiv. https://doi.org/10.1101/2023.03.02.530768

Wang, L., Zhu, Z., & Bastiaansen, M. (2012). Integration or Predictability? A Further Specification of the Functional Role of Gamma Oscillations in Language Comprehension. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00187

Wyatte, D., Jilk, D. J., & O'Reilly, R. C. (2014). Early recurrent feedback facilitates visual object recognition under challenging conditions. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00674

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3485–3492. https://doi.org/10.1109/CVPR.2010.5539970

Zeidman, P., Jafarian, A., Seghier, M. L., Litvak, V., Cagnan, H., Price, C. J., & Friston, K. J. (2019). A guide to group effective connectivity analysis, part 2: Second level analysis with PEB. *NeuroImage*, *200*, 12–25. https://doi.org/10.1016/j.neuroimage.2019.06.032