Recurrent connectivity supports higher-level visual and semantic object representations in the brain

Jacqueline von Seth[2], Victoria I. Nicholls[1], Lorraine K. Tyler[1,3] & Alex Clarke[1*]

[1]Department of Psychology, University of Cambridge, UK

[2]MRC Cognition and Brain Sciences Unit, University of Cambridge, UK

[3]Cambridge Centre for Ageing and Neuroscience (Cam-CAN), University of Cambridge and MRC Cognition and Brain Sciences Unit, Cambridge, UK

*Corresponding author: Alex Clarke, Department of Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK. Email: ac584@cam.ac.uk

Acknowledgements: This research was funded in whole, or in part, by the Wellcome Trust [Grant number 211200/Z/18/Z to AC], and the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ ERC Grant agreement no 249640 to LKT. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) research was supported by the Biotechnology and Biological Sciences Research Council (grant number BB/H008217/1). We thank the Cam-CAN respondents and their primary care teams in Cambridge for their participation in this study. We also thank colleagues at the MRC Cognition and Brain Sciences Unit MEG and MRI facilities for their assistance. 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.

Abstract

Visual object recognition is a highly dynamic process by which we extract meaningful information about the things we see. However, the functional relevance and informational properties of feedforward and feedback signals remains largely unspecified. Additionally, it remains unclear whether computational models of vision alone can accurately capture object-specific representations and the evolving spatiotemporal neural dynamics. Here, we probe these dynamics using a combination of representational similarity and connectivity analyses of fMRI and MEG data recorded during the recognition of familiar, unambiguous objects from a wide range of categories. Modelling the visual and semantic properties of our stimuli using an artificial neural network as well as a semantic feature model, we find that unique aspects of the neural architecture and connectivity dynamics relate to visual and semantic object properties. Critically, we show that recurrent processing between anterior and posterior ventral temporal cortex relates to higher-level visual properties prior to semantic object properties, in addition to semantic-related feedback from the frontal lobe to the ventral temporal lobe between 250 and 500ms after stimulus onset. These results demonstrate the distinct contributions semantic object properties make in explaining neural activity and connectivity, highlighting it is a core part of object recognition.

Introduction

Successful visual object recognition enables us to see and understand the world around us. It is well established that the ventral visual pathway (VVP) critically supports this process (Clarke & Tyler, 2015; DiCarlo et al., 2012; Kravitz et al., 2013; Ungerleider & Mishkin, 1982), with activity within the first few hundred milliseconds supporting the recognition of the visual input (Bankson et al., 2018; T. Carlson et al., 2013; Chen et al., 2016; Cichy et al., 2014, 2016; Clarke et al., 2013, 2015, 2018; Giari et al., 2020; Isik et al., 2014; Rupp et al., 2017). Our increasingly detailed computational, cognitive and translational accounts of the spatiotemporal processes that underly recognition allow us to form an increasingly specific understanding of what kinds of information are linked to the evolving object representations, and how they are spatially and temporally distributed. Critical to advancing this endeavour, is detailing how semantic object processing relates to, and is distinct from, processing the physical visual attributes of the image. Here, we look to probe the evolving nature of object representations in a critical new domain, that of examining how dynamic representational connectivity patterns relate to different visual and semantic aspects of objects, thus revealing the core properties that feedforward and feedback connectivity can support.

Traditional models argued that visual object recognition is underpinned by a largely feedforward process along the ventral visual pathway (DiCarlo et al., 2012; Isik et al., 2014; Serre et al., 2007), while feedback is reserved for more complex situations such as occlusion or visually degraded images (Ganis et al., 2007; Rajaei et al., 2019; Wyatte et al., 2014), meaning it might not be initiated for unambiguous familiar images. However, it has also been argued that object recognition generally involves recurrent connectivity (Clarke et al., 2011; Hegdé, 2008; Hochstein & Ahissar, 2002; Kar et al., 2019; Kietzmann et al., 2019; Spoerer et al., 2017; Wyatte et al., 2014) - although the functional consequence of feedforward and feedback signals, in terms of what kinds of representational details they support, is largely unspecified. One argument put forward to explain when recurrent processes might be more involved, is that when detailed semantic representations are required (such as knowing something is a tiger compared to knowing it is a living thing), long-range recurrent connectivity between anterior and posterior ventral temporal cortex increases (Clarke et al., 2011). However, it remains unclear exactly what aspects of visual and semantic object information relate to feedforward and feedback signals when recognising unambiguous familiar objects. To address this, we can use a recently developed method, representational connectivity analysis (RCA; Karimi-

Rouzbahani et al., 2021), which can show how activity in one region contributes to information-specific activity in another region at a later point in time. This follows the recent development of multivariate informational connectivity approaches (Anzellotti & Coutanche, 2018; Basti et al., 2020; Goddard et al., 2016; Rahimi, Jackson, et al., 2022), which seek to provide information about the timing and direction of between region connectivity, alongside the representational content of such connectivity.

A further important issue concerns what kind of object properties relate to the representational content of visual objects, and whether representations can be sufficiently captured by models of visual information alone. One of the dominant contemporary approaches to understanding the computational properties of the VVP, is to assess the relationship between convolutional neural network models of vision and the brain. While artificial and deep neural networks have shown an impressive ability to capture something about the neural properties of the human VVP (Cichy et al., 2016; Clarke et al., 2018; Devereux et al., 2018; Güçlü & Gerven, 2015; Seeliger et al., 2017), it remains in question whether such models have the potential to fully capture object recognition, without considering an additional, qualitatively different, form of information - that of semantics. If the goal of visual object recognition is to understand what the object being perceived *is*, then this requires accessing the meaning of the item, not simply discerning a label we can attach to the input. While numerous studies have shown effects in the VVP attributed to semantics during object recognition (Bonner & Epstein, 2021; T. A. Carlson et al., 2013; Clarke & Tyler, 2014; Davis et al., 2021; Devereux et al., 2018; Jozwik et al., 2016; Leonardelli & Fairhall, 2022; Naspi et al., 2021; Rupp et al., 2017; Wurm & Caramazza, 2022), there remain questions over whether these effects are attributable to object semantics, or if they actually relate to higher-level visual object properties.

It is undeniable that accessing semantic memory for an object is a clearly dissociable form of information from visual processing, as demonstrated through specific deficits (Hillis & Caramazza, 1995; Warrington, 1975), studies controlling visual input while manipulating semantics (e.g. Clarke et al., 2011; Martin et al., 2018; Mollo et al., 2017; Tyler et al., 2004). This highlights the critical issue of understanding the relationships between visual and semantic processes, the timing of such informational types, and how connectivity patterns may differentially relate to different properties. In this regard, image-based neural network models offer a useful tool to investigate visual processes, and allow us to determine what, if anything, models of semantics can explain beyond the current state-of-the-art neural network models.

Here, we address two overarching issues relating to, first, the role of feedforward and feedback signals during objects recognition, and second the relationship between visual and semantic object properties. To achieve this, we will use a combination of neuroimaging approaches with RSA and RCA applied to the recognition of visual objects from a large variety of object categories (Figure 1). Building across complimentary analyses we will, (1) use fMRI searchlight RSA to reveal the cortical architecture related to semantic object properties and how they relate to those explained by a computational model of vision; (2) explore the relative temporal dynamics and connectivity of visual and semantic measures at the level of MEG sensor arrays using both RSA and RCA; and (3) examine the spatio-temporal distribution of semantic effects, beyond those explained by a computation model of vision, using searchlight RSA of MEG source localised neural patterns and how connectivity within the resulting network related to semantics and visual object properties.

Results

Utilising the CORnet-S artificial neural network (ANN; Kubilius et al., 2019) of visual processing and a cognitive model of semantic object knowledge, we probed the spatial, temporal and connectivity

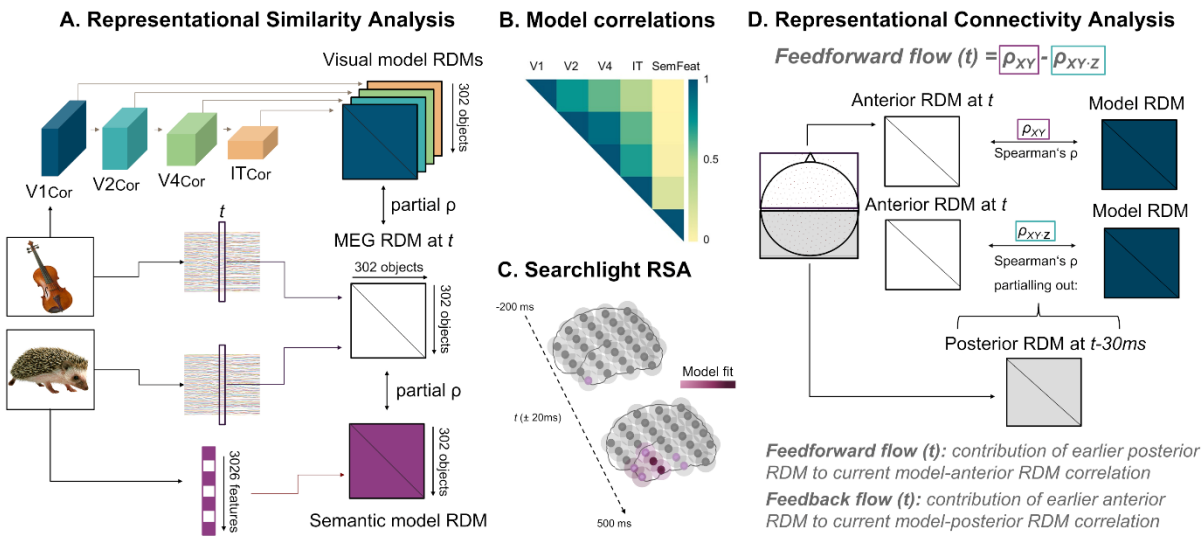

Figure 1. Schematic illustration of MEG analyses at the individual subject level. A) Visual RDMs are created from nodal activations extracted from 4 layers of the CORnet-s ANN. For the sensor-level MEG RSA analysis, RDMs for each timepoint are created from object-specific spatio-temporal patterns for each time-point extracted from MEG sensors. The semantic model is created using semantic feature vectors for each object. Vectorised Model RDMs are compared to neural RDMs using Spearman's correlations. B) Correlations between visual and semantic feature models show a high degree of correlation between the visual RDMs and limited correlation between the visual and semantic measure. C) Temporally resolved MEG RSA searchlight analysis for the semantic model. Vertices illustrated with grey dots with shaded searchlight spheres, and the degree of model fit indicated by the purple colouration. D) Representational connectivity analysis illustrated by the calculation of feedforward information flow (feedforward connectivity at each timepoint) between anterior and posterior regions at the MEG sensor-level, using RCA as introduced by Karimi-Rouzbahani et al. (2021).

characteristics of visual object recognition using RSA in fMRI and MEG, and Representational connectivity analysis (RCA) of MEG signals.

Semantic brain networks beyond ANNs

fMRI RSA searchlight

We first looked to establish which regions were associated with the layers of the CORnet ANN, and if a model of semantics could uniquely explain voxel patterns over and above CORnet. The four convolutional network layers of CORnet all showed significant RSA model fit to the VVP including occipital and posterior temporal lobes. This included primary visual cortex, lateral occipital cortex, and the posterior fusiform (Figure 2a; Table S1). Many of these regions also significantly related to the semantic feature model, which showed additional significant model fit to more anterior regions of the fusiform and parahippocampal gyrus, and ventral and medial aspects of the anterior temporal lobe (Figure 2a,b).

Critically, the semantic feature model could uniquely explain voxel PS over and above that which could be explained by CORnet, in bilateral pVTC, lateral occipital cortex, posterior medial cortex and medial ATL (Figure 2c). While there was some spatial overlap with CORnet models in pVTC, only the semantic model significantly related to voxel patterns in anterior temporal, posterior medial and anterior parts of the posterior ventral temporal cortex, replicating previous comparisons of visual and semantic effects (Clarke & Tyler, 2014; Devereux et al., 2018).

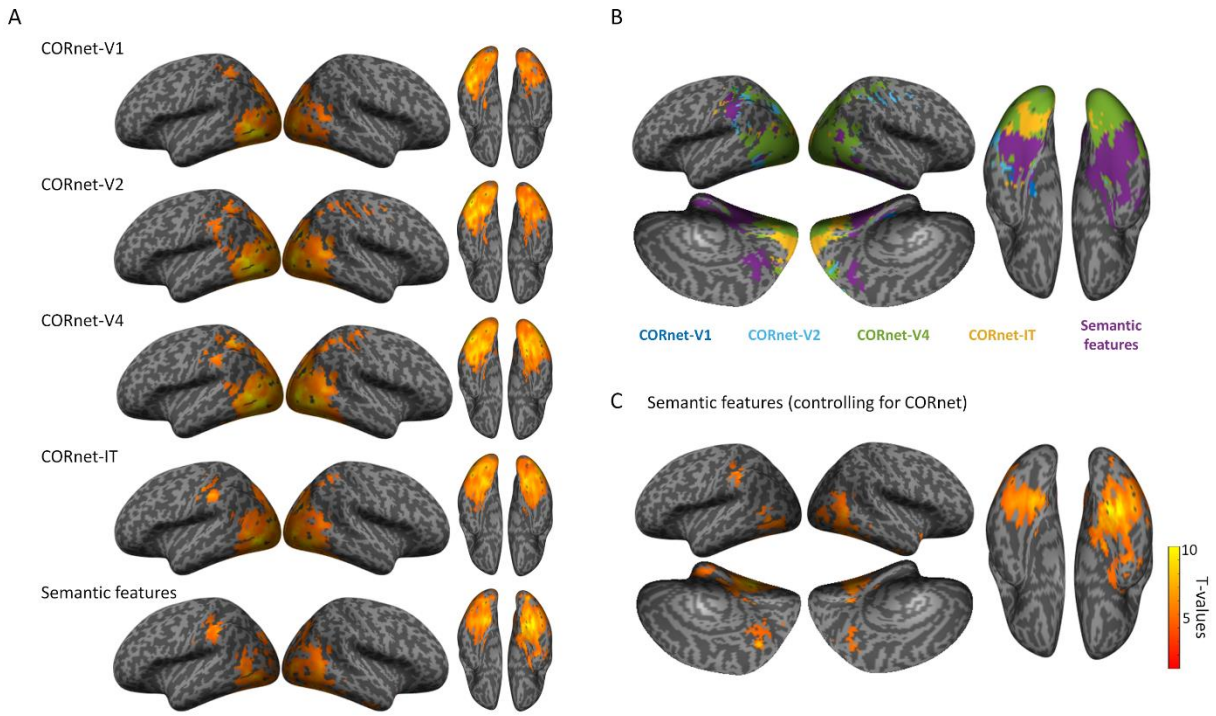

Figure 2. fMRI searchlight results. A) Maps show significant relationship between each model RDM and voxel patterns, voxelwise $p < 0.001$, cluster $p < 0.05$. B) Maps showing which model RDM, from panel A, has strongest effect size at each searchlight centre voxel. C) RSA effects of the semantic feature RDM partialling out effects of all CORnet layers.

Distinct dynamics of low and higher-level visual and semantic processing

MEG Sensor-level RSA

We next examined the temporal dynamics of object recognition with RSA of MEG sensor-level signals in relation to the same models of visual and semantic properties (Figure 3; Table S2). Looking at the unique effects of each CORnet model layer and the semantic model, all three of the lowest layers of CORnet showed rapid but transient significant model fits and peaking near 100 ms (CORnet-V1, cluster $p = 0.046$; cornet-V2, cluster $p = 0.011$; V4, cluster $p = 0.003$), with secondary effects of CORnet-V1 near 400 ms (cluster $p = 0.016$). The higher-level visual information captured through CORnet-IT showed a later model fit from around 130 ms, peaking near 250 ms, and remaining significant until approximately 600 ms (cluster $p < 0.0001$). The semantic model showed a distinct temporal morphology compared to the visual layers, becoming significant from around 250 ms with peaks near 350 ms and 600 ms (cluster1 $p = 0.001$; cluster2 $p = 0.023$; cluster3 $p = 0.0002$).

An analysis of the latency of the peak RSA effect sizes was conducted through calculating the 95% confidence intervals of the differences in RSA peak latencies between model RDMs (Figure 3B). This further revealed that the semantic feature model peaked significantly later compared to the low and mid-level CORnet layers (V1, V2 and V4) but not CORnet-IT, while CORnet-IT peaked significantly later compared to the low and mid-level CORnet layers (V1, V2 and V4). This suggests that, in addition to the spatial hierarchy seen in fMRI, we see a temporal evolution of object properties from low, to high-level visual properties, before semantic effects emerge later in time.

MEG sensor-level connectivity of visual and semantic effects

While these results might point to parallel spatial and temporal progressions of information, it reveals little of how regional interactivity might underly these effects. To address this, we utilised RCA of the sensor data to establish whether feedforward and/or feedback connectivity was related to the different visual and

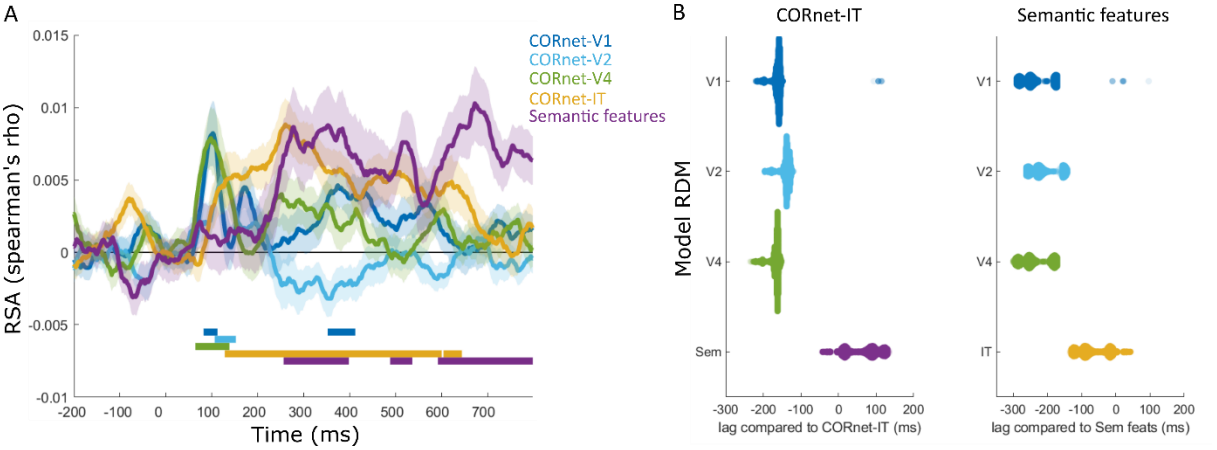

Figure 3. RSA results for the MEG sensor array. A) Partial correlation RSA analysis showing the unique effects of each model RDM over time. Shaded areas show standard error of the mean. Solid bars show time periods of significant effects. B) Swarmplots showing the differences in peak latency between model RDMs. Distributions display bootstrapped resamples of the data which were used to generate 95% CIs for the differences in peak latencies.

semantic properties. To do this, we split the MEG sensors into posterior and anterior regions and assessed RCA between them to provide a global measure of feedforward and feedback informational connectivity.

RCA revealed feedforward connectivity associated with all layers of CORnet from ~80 ms (cornet-V1, cluster $p = 0.003$; cornet-V2, cluster $p = 0.0003$; cornet-V4, cluster $p = 0.003$; cornet-IT, cluster $p < 0.0001$), with the first three layers peaking near 110-140 ms, and CORnet-IT peaking near 250 ms. Feedforward processing of semantic properties was significant later in time peaking before 300 ms (cluster $p < 0.0001$; Figure 4A; Table S3). Further, the latency of the peak feedforward effect of semantics was significantly delayed compared to all CORnet layers, and CORnet-IT was delayed compared to all other CORnet layers (Figure 4C). This shows that all types of object properties were associated with feedforward connectivity, with broad timings similar to the RSA effects presented above. Interestingly, only two types of object properties were associated with feedback connectivity, where patterns in anterior regions influenced later patterns in posterior regions. Feedback associated with CORnet-IT was seen around 250 to 400 ms (cluster $p = 0.0003$), and feedback associated with semantics was seen peaking near 400 ms (cluster $p = 0.0276$), and again significant around 600 ms (cluster $p < 0.0001$; Figure 4B; Table S3). The timing of these feedback effects coincided with feedforward effects for the same model RDMs, suggesting a period of dynamic recurrent connectivity (both feedforward and feedback) associated with higher-level visual and semantic object properties between 250 and 500 ms. In addition, the latency of

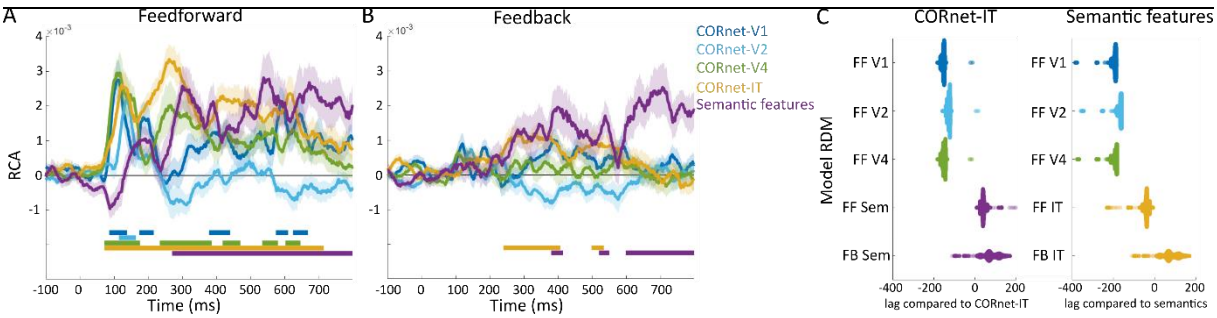

Figure 4. Representational connectivity analysis (RCA). A) RCA analysis applied to posterior and anterior sensor regions showing feedforward RCA effects of the model RDMs. B) Feedback RCA effects of the model RDMs. Shaded areas show standard error of the mean. Solid bars show time periods of significant effects. C) Swarmplots showing the differences in peak RCA latency between model RDMs. Distributions display bootstrapped resamples of the data which were used to generate 95% CIs for the differences in peak latencies.

the peak semantic feedback effects (near 400 ms) was significantly delayed compared to CORnet-IT (around 330 ms; Figure 4C). Finally, direct comparisons of feedforward and feedback peak effects showed feedforward peak effects were significantly earlier compared to feedback peaks for all model RDMs.

This shows that, at the global level of MEG sensor arrays, feedforward connectivity is associated with visual and semantic representations forming in anterior regions, in a temporal hierarchy of low, to mid, to high-level visual properties preceding semantic effects. Crucially, concomitant feedback signals were associated with higher-level visual and semantic properties pointing to the importance of recurrent activity during object recognition.

MEG searchlight analysis of semantic effects (after controlling for visual models)

A core issue we address next is how semantic information is accessed through a dynamic neural system at a more detailed level than is provided by the relatively global sensor analysis. Knowing this can address what regions reflect semantic information over time, and how connectivity aids this evolution of meaning. To this end, we conducted a searchlight RSA analysis to reveal the regions and time-points sensitive to semantic structure, while partialling out any influence of the ANN CORnet layers (Figure 5). This resulted in four significant spatio-temporal clusters that were sensitive to semantic feature information (Figure 5B) – one covering the right posterior ventral temporal cortex, lateral temporal and extending toward the ATL, lasting between approximately 100 and 450 ms (cluster time-window = 104-442 ms, peak effect = 282 ms, $p < 0.0001$). A second cluster was spatially centred in left pVTC between approximately 230 and 400 ms (cluster time-window = 234-406 ms, peak effect = 306 ms, $p < 0.0001$). A third cluster covered the right ATL including lateral and medial aspects between approximately 350 and 450 ms (cluster time-window = 346-456 ms, peak effect = 420 ms, $p = 0.008$), and a fourth cluster covered left lateral ventral PFC an lateral ATL between approximately 290 and 470 ms (cluster time-window = 386-468 ms, peak effect = 450 ms, $p = 0.01$).

Considering the pattern of semantic effects over space and time, we see significant model fit beginning in right pVTC after ~100 ms and spreading along bilateral pVTC and the right ATL within 250 ms. Semantic effects remained in the right ATL past 400 ms, alongside an additional cluster spreading across the left ventral PFC and lateral ATL (Figure 5A). This shows that accessing semantic information about objects engages a network of bilateral pVTC, ATL and the left ventral PFC within the first 600 ms of an object appearing. These effects for semantics are seen over and above those explained by the CORnet ANN of

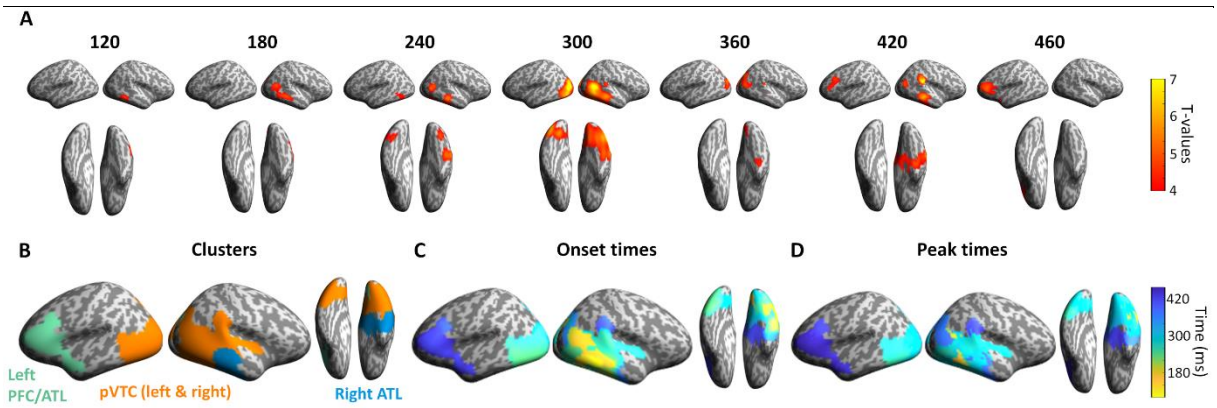

Figure 5. Searchlight RSA on source-localised MEG signals. A) Maps showing significant effects of semantic features (partialling our all effects of the ANN layers) over time. B) The effects were seen in four spatio-temporal clusters across bilateral pVTC, right ATL and left frontal/ATL. C) Onset time of the semantic effects at each vertex and D) time of peak RSA effect size at each vertex.

vision, with the spatial distribution of semantic effects closely overlapping with those revealed through fMRI RSA searchlight results.

Multivariate RCA shows feedback from anterior to posterior regions related to semantics

We next asked how connectivity within this network related to semantic information, and in addition, visual object properties more generally. To do this we used an extended version of RCA, to show how connectivity between a pair of regions related to semantic information, while controlling for the influence of all ANN layers and the other regions of the network showing semantic effects. We created our network to consist of three ROIs based on the significant RSA clusters, being bilateral pVTC (clusters 1 and 2), right ATL (cluster 3), and left vPFC/ATL (cluster 4). Any areas of spatial overlap between the clusters were excluded from all ROIs.

Relating specifically to semantic object properties, RCA revealed feedforward connectivity between pVTC and right ATL from ~150ms and continuing throughout the epoch (cluster $p < 0.001$), and feedback connectivity from right ATL to pVTC near 400 ms (cluster $p = 0.0058$; Figure 6A, Table S4). We also saw feedforward connectivity between the pVTC and left PFC/ATL starting at a similar time, around 150 ms, and continuing to 400ms (cluster $p < 0.001$), plus a second phase of feedforward connectivity from around 600 ms (cluster 1 $p = 0.009$; cluster2 $p = 0.001$). Small clusters of feedback from left PFC/ATL to pVTC relating to semantics were seen between 400 and 500 ms (cluster $p = 0.007$), and between 600 and 700 ms (cluster $p = 0.009$;Figure 6B, Table S4). Finally, RCA revealed no significant feedforward connectivity effects between the right ATL and left PFC/ATL region, but we did observe a small cluster of feedback from left PFC/ATL to right ATL between 400 and 500 ms (cluster $p = 0.0078$; Figure 6C, Table S4). Comparing the latencies of the peak effects, we saw that feedforward peaks occurred significantly earlier than feedback peaks between pVTC and right ATL (Figure 6A), except for in the case of pVTC and left PFC/ATL (Figure 6B). While this points to prominent feedforward semantic effects from the pVTC to both ATL and frontal regions from ~150 ms, we also found evidence that feedback between all regions in the network were significant beyond around 400 ms.

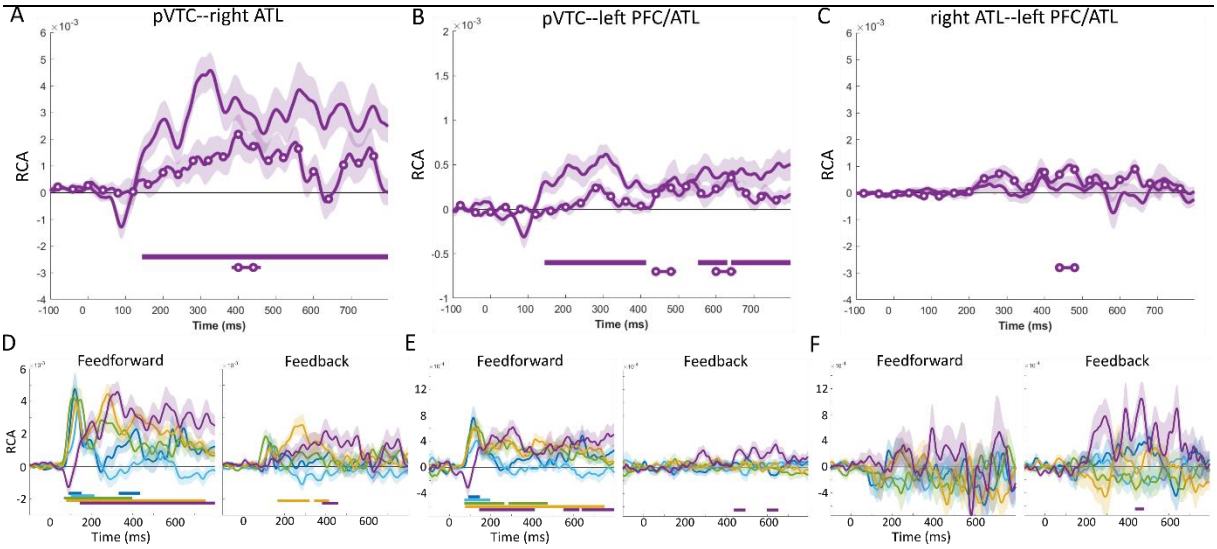

Figure 6. RCA of source localised MEG signals. RCA analysis applied to ROIs showing feedforward and feedback RCA effects of the semantic model RDM. A) Effects between the pVTC and right ATL. Solid line shows feedforward effects and dotted line shows feedback effects. B) RCA effects between pVTC and the left PFC/ATL. C) RCA effects between right ATL and left PFC/ATL. Shaded areas show standard error of the mean. Solid bars show time periods of significant effects. D,E,F) RCA effects for visual and semantic models for the three connections showing feedforward RCA effects (left) and feedback RCA effects (right).

Our analyses show that semantic processing is associated with feedforward effects from pVTC, and feedback effects to pVTC, with both the right ATL and left PFC/ATL. Direct comparisons of the RCA effect sizes between these connections revealed feedforward effects of semantics were significantly stronger for the pVTC-right ATL connection compared to the pVTC-left PFC/ATL connection between approximately 150 and 800 ms (cluster1: $p = 0.005$; cluster2: 249-800 ms, $p < 0.001$). In addition, feedback signals to pVTC were significantly stronger from right ATL compared to being from left PFC/ATL around 400 ms (cluster: 385-456 ms, $p = 0.0069$). While our RCA results point to feedforward and feedback effects associated with both connections, this shows that connectivity effects are stronger in the case of the pVTC and right ATL connection for semantic processing.

Finally, we extended the RCA analysis to probe how these semantic effects temporally relate to visual ANN connectivity effects. This allows us to situate the semantic effects in the context of low and higher-level visual properties. While feedforward connectivity effects between the pVTC and both anterior regions related to rapid effects of all ANN layers, the higher-level layers and the semantic model also related to later feedforward connectivity (Figure 6D-F, Table S5). Most interestingly, there were limited effects of feedback relating to the visual ANN, where we only saw feedback connectivity related to the IT-layer prior to feedback relating to semantics from the right ATL to the pVTC (Figure 6D, Table S5). Testing the latencies of the semantic and Cornet-IT peak effects revealed that feedback was significantly earlier for CORnet-IT compared to semantics. This highlights that while feedback activity across this network is linked to semantic information, higher-level visual properties also relate to feedback connectivity, but at an earlier point in time compared to semantic effects. Together with the sensor analyses, our results support the notion that feedback activity to the posterior ventral temporal lobe is important for object recognition, and feedback within the temporal lobe is most relevant to the processing of higher-level visual and semantic object properties between 200 and 500 ms.

Conclusions

Visual object recognition is a dynamic and interactive process. However, our lack of understanding of the functional role that the interplay between feedforward and feedback connectivity play, and the unique contributions visual and semantic object properties make to this, means our cognitive and computational models of visual object recognition have remained limited. Recurrent processing within artificial neural network models is claimed to provide a better account of object processing compared to feedforward neural network models (Kar et al., 2019; Kietzmann et al., 2019; O'Reilly et al., 2013; Rajaei et al., 2019; Spoerer et al., 2017), while brain activity within the ventral visual pathway has been shown to display long-range recurrent dynamics (Ahlfors et al., 2015; Clarke et al., 2011, 2018; Kietzmann et al., 2019; Lamme, 2003; Mohsenzadeh et al., 2018). But despite highlighting that recurrent activity might play an important role in models and the brain, a critical question is what the functional relevance of recurrent activity in the brain is, with regards to the object properties and representations that are embedded in our models. Here, our research makes two critical advances to enable us to address this vital question. First, using RSA and RCA combined with an ANN of vision and a semantic feature model, we showed that different visual and semantic properties of objects related to time-resolved feedforward and feedback signals in the VVP and frontal lobe. Importantly, we linked recurrent connectivity to specific cognitive measures of higher-level vision and semantics. Second, we demonstrated the importance of semantic object information to explain neural activity and connectivity in the VVP and frontal lobe, illlustrating the contributions semantics makes beyond that explained by ANNs. In this manner, we showed that recurrent activity supports object recognition in the brain, and critically showed that this relates to higher-level visual and semantic properties between ~200 and 500 ms, driven predominantly by interactions between the anterior and posterior ventral temporal cortex.

Many traditional models of visual object recognition have stressed that 'core' aspects of object recognition are supported through feedforward processes along the VVP (DiCarlo et al., 2012; Serre et al., 2007).

More recently, long-range feedback and recurrent activity has been acknowledged as being important for object recognition (Bar et al., 2006; Clarke & Tyler, 2015; Hegdé, 2008; Kar et al., 2019; Kietzmann et al., 2019; Schendan & Ganis, 2012; Spoerer et al., 2017; Wyatte et al., 2014), although the situations when recurrent activity is more prevalent, and critically, what its functional cognitive role is, have been less clear. Our results make a significant contribution to this issue – namely the role of recurrent activity. Using RSA, RCA and quantifiable models of visual and semantic object properties, we were able to track the spatio-temporal evolution of object properties, and how different properties related to feedforward and feedback connectivity both within the VVP, and between temporal and frontal regions.

We showed that higher-level visual properties of the ANN (CORnet layer IT) and semantic feature measures both related to feedback connectivity from the ATL to the pVTC, with higher-level visual feedback preceding semantic effects. These feedback effects overlapped in time with feedforward connectivity effects of the same object properties, suggesting a period of dynamic interactions between the pVTC and the ATL in support of higher-level visual and semantic representations. In addition, we saw more transient effects of feedback from left PFC/ATL to right ATL, and left PFC/ATL to the pVTC beyond 400 ms. This suggests that recurrent activity in the VVP, with support from the PFC, enables higher-level visual and semantic object representations, and does so even for highly familiar, unambiguous objects like those used here.

Our connectivity analysis established a distinct later period of feedback captured uniquely by the ANN layer-IT and the semantic feature model, predominantly beyond around 200ms. This is in line with the view that recurrent processing is crucial for the transition from lower-level visual to higher-level visual and object-specific semantic representations (e.g. Campo et al., 2013; Chiou & Lambon Ralph, 2016; Clarke et al., 2011, 2018; Hochstein & Ahissar, 2002; Poch et al., 2015). According to this view, variations of which were posited by multiple-state interactive (MUSI) theory (Schendan & Ganis, 2012; Schendan & Maher, 2009) or the top-down-facilitation model (Ahlfors et al., 2015; Bar, 2003; Bar et al., 2006), an initial feedforward sweep activates visual and coarse categorical representations, before an interplay of higher-level feedforward and feedback signals serves to build object-specific semantic representations at 200-500ms (Clarke, 2019). Previous connectivity studies have shown that feedback signals become increasingly active between 200-400ms (Ahlfors et al., 2015) and are modulated by semantic task demands (Clarke et al., 2011; Poch et al., 2015). Further, sematic properties have also been related to feedforward, feedback and lateral interactions within the VVP (Clarke et al., 2018). Here, we revealed that high-level visual and semantic information are uniquely related to these dynamic recurrent interactions. Furthermore, whilst being staggered in time, feedforward and feedback flows overlapped, which may be in line with stochastic evidence accumulator models of perceptual decision making, with decisions made on continually evolving representations through recurrence, and recognition triggered when a threshold has been reached (Gwilliams & King, 2020; Mack & Palmeri, 2011; O'Connell et al., 2012).

Our research helps expand our understanding of the situations when recurrent activity occurs during object recognition. It had been suggested that recurrent activity in the VVP increases when the task requires a more semantically detailed response (Clarke et al., 2011), while other research suggests that top-down effects within the VVP, or from frontal regions, increases under challenging object recognition conditions such as occlusion and degraded stimuli (Ganis et al., 2007; Rajaei et al., 2019; Schendan & Maher, 2009; Wyatte et al., 2014). Given the lack of challenging conditions here, this suggests that while recurrent activity supports higher-level visual and sematic processing in unchallenging circumstances, these object properties, and indeed lower-level ones, may be facilitated to a greater extent by feedback activity when the situation dictates.

Our searchlight RSA analysis of source-localised MEG signals pointed to a dynamic network underpinning semantic representations involving bilateral pVTC, the ATL and left frontal lobe. In addition to feedback connectivity linked to semantics between the ATL and pVTC, significant periods of feedback connectivity beyond 400 ms from the frontal lobe were also linked to semantic object properties. Between 400 and 500 ms, feedback connectivity from the left frontal lobe was seen to both the right ATL and pVTC, with

an additional feedback cluster from left frontal to pVTC beyond 600 ms. We also saw evidence of a feedforward effect of semantics from the pVTC to both the ATL and frontal lobes beginning before 200 ms, suggesting that semantic representations are supported by early feedforward activity from pVTC to the ATL and frontal lobe, and later feedback activity from the frontal and ATL, to the pVTC, with an additional period of feedback from the frontal to the ATL.

While it is undeniable that semantic responses to visual objects are a dissociable form of information compared to image-based visual object properties, there has been some suggestions that semantic effects seen in neuroimaging might be partially explained by visual aspects of the stimuli. While a clear distinction between higher-level visual and semantic object properties can be problematic and is a potentially futile endeavour in some regards (e.g. is the knowledge that something has stripes purely distinct from cells tuned to detect striped textures), there are certainly semantic properties which transcend vision. Here, we show that visual object properties, captured through an ANN, and semantic properties, display different spatial, temporal and morphological RSA and RCA effects across both the MEG and fMRI datasets. Both our fMRI RSA searchlight and the MEG-based searchlight showed spatially similar patterns where semantic properties related to activity patterns in bilateral pVTC and the ATL beyond that which the visual ANN could explain. Our MEG results further showed that semantic effects tended to lag behind higher-level visual effects, as expected considering they reflected a more abstracted type of object information. We would suggest that a visual to semantic transition, in terms of the representational state and how it relates to cognitive properties, occurs gradually within the first 400 ms, with previous MEG connectivity evidence suggesting recurrent connectivity supports this translation, including feedback from the ATL to the pVTC (Clarke et al., 2018). Together with previous connectivity, fMRI, and neuropsychological studies, our results underpin the role of feedback from the ATL to the pVTC in supporting the semantic aspects of visual object recognition (Campo et al., 2013; Chiou & Lambon Ralph, 2016; Clarke et al., 2018; Mummery et al., 1999), with additional feedback contributions from the frontal lobe (Bar et al., 2006; Kar & DiCarlo, 2021; Poch et al., 2015; Rahimi, Farahibozorg, et al., 2022; Riès et al., 2017). While current accounts might point to differential effects of feedback either within the temporal lobe, or between frontal and temporal regions (Chiou et al., 2018; Ries et al., 2020; Riès et al., 2017; Schendan & Ganis, 2012), the approaches we outline here could be used to further understand such relative contributions to semantics and behaviour in finer detail.

Finally, consistent with many accounts of 'core' object recognition (DiCarlo et al., 2012; Riesenhuber & Poggio, 1999; Serre et al., 2007), earlier layers of the visual neural network model related to rapid feedforward connectivity effects between posterior VTC and both the ATL and frontal lobe. In relation to higher-level visual and semantic effects, this points to feedforward connectivity supporting a representational shift through low to high-level visual features and semantics, prior to recurrent effects of higher-level visual and semantic object properties.

Visual object recognition must be supported by dynamic and interactive neural mechanisms, which requires techniques suited to tracking neural activity and connectivity at a fine-temporal resolution, combined with cognitive models of our stimuli. Here, we show that long-range recurrent activity and connectivity underpins object recognition in the VVP, and outside. Through specifying models of both the visual and semantic aspects of our stimuli, we saw that recurrent activity within the VVP related to higher-level visual properties prior to semantic object properties, with addition feedback connectivity from the frontal lobe relating to semantics. These results highlight the important contributions of both visual and semantic object properties to the evolving object representations, and that visual models alone can not account for all aspects of visual object recognition.

Methods

Here we combined participants from two MEG datasets that were previously reported by Clarke et al., (2015, 2018) and Bruffaerts et al. (2019), both using the same stimuli, task and acquisition parameters.

We also report a reanalysis of a previously published fMRI dataset (Clarke & Tyler, 2014). Therefore, only the main aspects of data acquisition and pre-processing are reported here.

MEG Participants and Procedure

The study used a total of 36 participants, 15 from Clarke et al. (2015, 2018) and 21 from the young group in Bruffaerts et al. (2019). Data from Bruffaerts et al. was part of the Cambridge Centre for Ageing and Neuroscience (CamCAN) study, and the data used in the preparation of this work can be obtained from the CamCAN repository (available from https://www.cam-can.org) (Shafto et al., 2014; Taylor et al., 2017). Informed consent was obtained from all participants and ethical approval for the studies were obtained from the Cambridgeshire Research Ethics Committee. All experiments were performed in accordance with relevant guidelines and regulations.

All participants performed a basic-level naming task during which they were asked to identify 302 common objects from 11 superordinate categories, comprising of both living and non-living things. All objects were shown to participants in a pseudo-randomised order and presented in colour on a white background. A trial began with a fixation cross for 500 ms, before the object was presented for 500ms, after which followed a blank screen for a random interval between 2400ms and 2700ms. Incorrectly named trials, where participants could not name or incorrectly named the object, were excluded from all further analyses.

MEG recording/MRI acquisition

All MEG data were collected at the MRC Cognition and Brain Sciences Unit, Cambridge, UK, using a whole-head 306 channel (102 magnetometers, 204 planar gradiometers) Vector-view system (Elekta Neuromag, Helsinki, Finland). Blinks and eye movements were recorded using electro-oculogram (EOG) electrodes and the head position was recorded using five Head-Position Indicator (HPI) coils. Participants' head shapes, and the positions of EOG electrodes, HPI coils and fiducial points (nasion, left and right periauricular) were digitally recorded with a 3D digitiser (Fastrak Polhemus, Inc., Colchester, VA, USA). MEG signals were recorded at a sampling rate of 1000Hz and a high-pass filter of 0.03 Hz. To enhance source localization, T1-weighted MP-RAGE scans with 1 mm isotropic resolution were acquired for each subject using Siemens 3-T Tim Trio.

MEG pre-processing

Initial pre-processing of the raw MEG data used MaxFilter (version 2.2) to apply temporal signal space separation (tSSS) and head-motion correction. The MEG signals were then low-pass filtered at 200 Hz using a fifth-order Butterworth filter, and high-pass filtered at 0.1Hz using a fourth-order Butterworth filter in SPM12 (Wellcome Institute of Imaging Neuroscience, London, UK). The data were then epoched between -1000ms to 1000ms and down-sampled to 500Hz. Artefact removal was performed using Independent Component Analysis (ICA) implemented with RUNICA in EEGlab (Delorme & Makeig, 2004), supplemented with SASICA (Chaumon et al., 2015) to identify muscle and speech-related artefacts. Finally, baseline correction was applied using the pre-stimulus time window of -200 to 0ms. Source localisation of MEG signals used a minimum-norm procedure based on both magnetometer and gradiometer sensors. Individual participants MRI images were segmented and spatially normalised to an MNI template brain consisting of 5124 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.

MEG Representational Similarity Analysis (RSA)

Representational Similarity Analysis (Kriegeskorte, 2008) was used to test model representational dissimilarity matrices (RDMs) capturing visual and semantic information against neural activity patterns as measured by MEG, at both the sensor-level and for a spatiotemporal searchlight analysis (Su et al., 2012).

Construction of model RDMs

Visual RDMs. Visual model RDMs were created using CORnet-S, a four-layer recurrent DNN for core object recognition which was modelled on the anatomy of the primate ventral visual stream, encompassing areas V1, V2, V4 and IT (Kubilius et al., 2019). CORnet-s has been shown to perform well in predicting both neural responses recorded in macaque V4 and IT, as well as human behavioural similarity judgements (Kar et al., 2019; Kubilius et al., 2019). The nodal activations for network layers V1, V2, V4 and IT were extracted for each object using the THINGSVision toolbox in Python (Muttenthaler & Hebart, 2021). After extracting the activations for each of the 302 objects, an RDM was created for each layer by calculating the Pearson's correlation distance between all possible pairs of pictures resulting in a 302x302 symmetrical dissimilarity matrix per layer.

Semantic RDM. The degree of semantic similarity between objects was quantified based on semantic features, using the Centre for Speech, Language and the Brain property norms (Devereux et al., 2014), containing 836 different concepts and their association to 3026 different properties (excluding taxonomic features, which refer to superordinate categories and are not typically regarded as semantic features in studies of conceptual representation). Using these property norms, each object could be represented by a list of features that collectively define its conceptual content. A 3026-long vector was created for each object, consisting of zeros and ones indicating whether each specific feature was associated with that object. An RDM was then created by calculating the Cosine distance between all possible object pairs resulting in a symmetric 302x302 matrix.

MEG sensor RSA

RDMs were created in order to calculate the similarity between MEG sensor activity patterns associated with each object, for each participant and timepoint. RDMs were derived from spatiotemporal MEG sensor patterns extracted over a 40ms sliding window centred around each timepoint, *t*, recorded from 204 planar gradiometers. 1-Pearson correlations were calculated pairwise for all combinations of object pairs, resulting in a 302x302 dissimilarity matrix for each timepoint.

In order to assess if the MEG RDMs related to visual or semantic object properties, the MEG RDMs were correlated with each model RDM at every timepoint, resulting in an RSA time-course for every model RDM and for each participant. Partial Spearman's correlations were used to test between a single model RDM and the MEG signals, controlling for the influence of all other model RDMs in order to find the unique effects of each model RDM.

The group-level RSA analysis was assessed for statistical significance using one-sample t-tests against a baseline of zero (α=0.01) for each timepoint and each model RDM. Cluster-mass permutation tests were applied in order to control for multiple comparisons across the number of time-points tested (Maris & Oostenveld, 2007). Each cluster of timepoints for which the correlation between MEG RDM and model RDM was significant was assigned a cluster p-value based on the sum of the t-values within the cluster. To determine these cluster p-values, a null distribution was created from the permuted RSA time-courses, where the sign of each subject's RSA time-course was randomly flipped before conducting the one-sample t-test again and retaining the cluster size of the largest cluster, which was added to the null distribution. This was repeated for 10,000 random permutations of the data. The cluster p-value for each of the original clusters was defined as the proportion of the 10,000 permutations (plus the observed cluster) that were greater than or equal to the observed cluster-mass.

Peak latency analysis

In order to assess the statistical significance of the difference in peak latencies, we used a leave-n-out approach, with n set to 4 (12% of the data; highly similar results are seen at n=3 and n=5). For each of the possible 31,465 unique resamples of the data, group average RSA time-courses were calculated and the timepoint of the maximum effect was extracted. This created a distribution of peak times for each model RDM. The distribution of peak was then used to define 95% confidence intervals (CIs) of the peak, and distribution of peak differences to define 95% CIs of pairwise differences. The null hypothesis was rejected if the 95% CIs of the pairwise differences did not include zero (Mohsenzadeh et al., 2019).

Searchlight RSA

The relationship between the representational geometries of the cognitive RDMs and of the MEG signals was tested using spatiotemporal searchlight RSA (Su et al., 2012). At each vertex and time-point, *t*, single-trial MEG signals were extracted for all vertices within a 10 mm radius for timepoints within a 40 ms time window centred at *t*. RDMs were calculated as 1-Pearson's correlation between all pairs of single trial responses. This procedure was repeated for all vertices and timepoints between -200 and 500 ms (at 2 ms increments). Here we focus on the semantic model, therefore the MEG data RDMs for each point in space and time were compared against the dissimilarity values computed for the semantic model RDM using Partial Spearman's correlation, controlling for the CORnet-S layer model RDMs, resulting in an RSA timeseries for each vertex, timepoint and participant.

Random effects analysis testing for positive RSA effects was conducted for each timepoint and vertex using one-sample t-tests against zero (alpha 0.01). To correct for multiple statistical comparisons over time and space, a second level of statistical thresholding was applied using cluster-based permutation testing after clusters were formed by contiguous above threshold effects in space (defined by adjacent vertices on the cortical mesh) and time. The size of the observed clusters was calculated as the sum of the t-values for time/spatial points within the cluster. P-values were assigned to each cluster using a null distribution based on permuted MEG signals. For each of the 1000 permutations, the sign of the correlations was randomly flipped for each participant before one-sample t-tests of the permuted data at each vertex and timepoint, and the largest cluster (sum of above-threshold t-values) over time and space was retained for the null distribution. The cluster p-value for each of the observed clusters was defined as the proportion of the 1000 permutations (plus the observed cluster-mass) that were greater than or equal to the observed cluster-mass.

Representational Connectivity Analysis (RCA)

To assess whether information captured by our models of visual and semantic processing is transferred between regions during object recognition, we applied an RSA-based informational connectivity analysis developed by Karimi-Rouzbahani et al. (2021). Representational connectivity analysis (RCA) works by establishing the contribution of earlier (in time) neural representations from the source region to current informational content in the target region, thereby uniquely providing information about the latency, direction, as well as the representational content of information flow. This analysis was applied to both the sensor and source-localised MEG signals.

Sensor RCA

To approximate anterior and posterior brain regions, MEG gradiometer sensors were split into two equal sensor regions according to the sensor y-coordinate. MEG RDMs were then calculated for each region separately, at each timepoint and for each subject. However, in contrast to the RSA analysis, MEG RDMs were based on a single timepoint rather than a 40 ms time-window to make sure time-lagged posterior

and anterior RDMs did not share any of the same data, and the upper triangle of each RDM was converted into a representational dissimilarity vector (RDV). RCA between a source and target region was calculated for each timepoint, *t*, and each model RDV in two steps. First, we calculated the partial Spearman's correlation between the target MEG RDV and a single model RDV controlling for all other model RDVs. Second, we calculated a second partial Spearman's correlation between the target MEG RDV and a single model RDV, controlling for all other model RDVs (as in the first calculation), but additionally controlling for the influence of the source MEG RDVs from previous timepoints. The time range of the previous RDVs was set to include -30 ms to -2 ms, chosen based on previous studies investigating the timeframe of activation flow between sensory and frontal regions (Foxe & Simpson, 2002; Karimi-Rouzbahani et al., 2021). RCA was defined as the difference between the first and second calculation, with feedforward RCA being when the posterior region was the source and anterior region was the target, and feedback RCA being when the anterior region was the source and posterior region was the target.

Source-level RCA

RCA was also applied to source-level regions of interest. Regions were derived from the spatiotemporal clusters revealed from the searchlight RSA analysis. Each region could act as a source and target to the other regions, resulting in three feedforward and three feedback RCA time-courses for each model RDM and participant. In order to ensure bivariate effects could not be explained by a common effect from the third region, the third region was added to the partial correlation analyses. For example, to ensure the RCA effects between Region A and Region B could not be explained by past effects in Region C, the Region C MEG RDM was partialled out in both the first and second partial Spearman's correlations.

For both the senor and source-level RCA analyses, feedforward and feedback RCA time-courses for each model RDM and each participant were created, which were assessed for statistical significance using the same procedures outlined above for the senor RSA analysis and peak latency analyses.

fMRI Participants and Procedure

We re-analysed fMRI data, originally reported in Clarke and Tyler (2014), collected from 16 participants (10 female, 6 male). All participants gave informed consent and the study was approved by the Cambridge Research Ethics Committee. The data pre-processing and RSA analysis framework were the same as in Clarke and Tyler (2014) and so only the main aspects of experiment are reported here. Participants named 145 objects, where each object was depicted by a photograph of an isolated object on a white background. From the 145 objects, 131 were from one of six object categories (animals, fruit, vegetables, tools, vehicles, musical instruments), and the remaining 14 objects did not clearly belong to a category and were not included in the analyses. All objects were presented once in each of six fMRI scanning blocks. The un-normalised and non-smoothed EPI images were entered into a general linear model to obtain a single t-statistic image for each concept picture. These t-maps, masked by the grey matter mask from each subject, were used in subsequent representational similarity analysis.

In the whole brain searchlight mapping analysis each of the visual and semantic RDMs were tested across the brain (spherical searchlights, sphere radius of 7mm). For group random-effects analyses, the Spearman's correlation maps for each participant were Fisher-transformed, normalized to standard MNI space, and spatially smoothed with a 6mm FWHM Gaussian kernel. Maps were entered into a random effects analysis (RFX) in SPM12 where each voxel was tested using a one-sampled t-test against zero. Voxelwise multiple comparisons correction was applied through a voxelwise threshold of $p < 0.001$ and FWE-cluster $p < 0.05$.

References:

Ahlfors, S. P., Jones, S. R., Ahveninen, J., Hämäläinen, M. S., Belliveau, J. W., & Bar, M. (2015). Direction of magnetoencephalography sources associated with feedback and feedforward contributions in a visual object recognition task. *Neuroscience Letters*, *585*, 149–154. https://doi.org/10.1016/j.neulet.2014.11.029

Anzellotti, S., & Coutanche, M. N. (2018). Beyond Functional Connectivity: Investigating Networks of Multivariate Representations. *Trends in Cognitive Sciences*, *22*(3), 258–269. https://doi.org/10.1016/j.tics.2017.12.002

Bankson, B. B., Hebart, M. N., Groen, I. I. A., & Baker, C. I. (2018). The temporal evolution of conceptual object representations revealed through models of behavior, semantics and deep neural networks. *NeuroImage*, *178*, 172–182. https://doi.org/10.1016/j.neuroimage.2018.05.037

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609.

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), 449–454.

Basti, A., Nili, H., Hauk, O., Marzetti, L., & Henson, R. N. (2020). Multi-dimensional connectivity: A conceptual and mathematical review. *NeuroImage*, *221*, 117179. https://doi.org/10.1016/j.neuroimage.2020.117179

Bonner, M. F., & Epstein, R. A. (2021). Object representations in the human brain reflect the co-occurrence statistics of vision and language. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-24368-2

Bruffaerts, R., Tyler, L. K., Shafto, M., Tsvetanov, K. A., & Clarke, A. (2019). Perceptual and conceptual processing of visual objects across the adult lifespan. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-50254-5

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), 12679–12688.

Carlson, T. A., Simmons, R., Kriegeskorte, N., & Sleve, L. R. (2013). The emergence of semantic meaning in the ventral temporal pathway. *Journal of Cognitive Neuroscience*, *26*(1), 120–131.

Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, *13*(10). https://doi.org/10.1167/13.10.1

Chaumon, M., Bishop, D. V. M., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of Neuroscience Methods*, *250*, 47–63. https://doi.org/10.1016/j.jneumeth.2015.02.025

Chen, Y., Shimotake, A., Matsumoto, R., Kunieda, T., Kikuchi, T., Miyamoto, S., Fukuyama, H., Takahashi, R., Ikeda, A., & Lambon Ralph, M. A. (2016). The 'when' and 'where' of semantic coding in the anterior temporal lobe: Temporal representational similarity analysis of electrocorticogram data. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *79*, 1–13. https://doi.org/10.1016/j.cortex.2016.02.015

Chiou, R., Humphreys, G. F., Jung, J., & Lambon Ralph, M. A. (2018). Controlled semantic cognition relies upon dynamic and flexible interactions between the executive 'semantic control' and hub-and-spoke 'semantic representation' systems. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *103*, 100–116. https://doi.org/10.1016/j.cortex.2018.02.018

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

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, *6*, 27755. https://doi.org/10.1038/srep27755

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635

Clarke, A. (2019). Chapter Three—Neural dynamics of visual and semantic object processing. In K. D. Federmeier & D. M. Beck (Eds.), *Psychology of Learning and Motivation* (Vol. 70, pp. 71–95). Academic Press. https://doi.org/10.1016/bs.plm.2019.03.002

Clarke, A., Devereux, B. J., Randall, B., & Tyler, L. K. (2015). Predicting the Time Course of Individual Objects with MEG. *Cerebral Cortex*, *25*(10), 3602–3612. https://doi.org/10.1093/cercor/bhu203

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), 1590–1605. https://doi.org/10.1162/jocn_a_01325

Clarke, A., Taylor, K. I., Devereux, B., Randall, B., & Tyler, L. K. (2013). From Perception to Conception: How Meaningful Objects Are Processed over Time. *Cerebral Cortex*, *23*(1), 187–197. https://doi.org/10.1093/cercor/bhs002

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), 1887–1899.

Clarke, A., & Tyler, L. K. (2014). Object-specific semantic coding in human perirhinal cortex. *Journal of Neuroscience*, *34*(14), 4766–4775.

Clarke, A., & Tyler, L. K. (2015). Understanding What We See: How We Derive Meaning From Vision. *Trends In Cognitive Sciences*, *19*(11), 677–687.

Davis, S. W., Geib, B. R., Wing, E. A., Wang, W.-C., Hovhannisyan, M., Monge, Z. A., & Cabeza, R. (2021). Visual and Semantic Representations Predict Subsequent Memory in Perceptual and Conceptual Memory Tests. *Cerebral Cortex*, *31*(2), 974–992. https://doi.org/10.1093/cercor/bhaa269

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Devereux, B. J., Clarke, A., & Tyler, L. K. (2018). Integrated deep visual and semantic attractor neural networks predict fMRI pattern-information along the ventral object processing pathway. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-018-28865-1

Devereux, B. J., Tyler, L. K., Geertzen, J., & Randall, B. (2014). The Centre for Speech, Language and the Brain (CSLB) concept property norms. *Behavior Research Methods*, *46*(4), 1119–1127. https://doi.org/10.3758/s13428-013-0420-4

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434.

Foxe, J., & Simpson, G. (2002). Flow of activation from V1 to frontal cortex in humans. *Experimental Brain Research*, *142*(1), 139–150. https://doi.org/10.1007/s00221-001-0906-7

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.

Giari, G., Leonardelli, E., Tao, Y., Machado, M., & Fairhall, S. L. (2020). Spatiotemporal properties of the neural representation of conceptual content for words and pictures – an MEG study. *NeuroImage*, *219*, 116913. https://doi.org/10.1016/j.neuroimage.2020.116913

Goddard, E., Carlson, T. A., Dermody, N., & Woolgar, A. (2016). Representational dynamics of object recognition: Feedforward and feedback information flows. *NeuroImage*, *128*, 385–397. https://doi.org/10.1016/j.neuroimage.2016.01.006

Güçlü, U., & Gerven, M. A. J. van. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *Journal of Neuroscience*, *35*(27), 10005–10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015

Gwilliams, L., & King, J.-R. (2020). Recurrent processes support a cascade of hierarchical decisions. *ELife*, *9*, e56603. https://doi.org/10.7554/eLife.56603

Hegdé, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Prog Neurobiol*, *84*(4), 405–439. https://doi.org/10.1016/j.pneurobio.2007.09.001

Hillis, A. E., & Caramazza, A. (1995). Cognitive and neural mechanisms underlying visual and semantic processing: Implications from 'optic aphasia'. *Cognitive Neuropsychology*, *7*, 457–478.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804. https://doi.org/S0896627302010917 [pii]

Isik, L., Meyers, E. M., Leibo, J., & Poggio, T. (2014). The dynamics of invariant object recognition in the human visual system. *Journal of Neurophysiology*, *111*, 91–102.

Jozwik, K. M., Kriegeskorte, N., & Mur, M. (2016). Visual features as stepping stones toward semantics: Explaining object similarity in IT and perception with non-negative least squares. *Neuropsychologia*, *83*, 201–226. https://doi.org/10.1016/j.neuropsychologia.2015.10.023

Kar, K., & DiCarlo, J. J. (2021). Fast Recurrent Processing via Ventrolateral Prefrontal Cortex Is Needed by the Primate Ventral Stream for Robust Core Visual Object Recognition. *Neuron*, *109*(1), 164-176.e5. https://doi.org/10.1016/j.neuron.2020.09.035

Kar, K., Kubilius, J., Schmidt, K., Issa, E. B., & DiCarlo, J. J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscience*, *22*(6), Article 6. https://doi.org/10.1038/s41593-019-0392-5

Karimi-Rouzbahani, H., Ramezani, F., Woolgar, A., Rich, A., & Ghodrati, M. (2021). Perceptual difficulty modulates the direction of information flow in familiar face recognition. *NeuroImage*, *233*, 117896. https://doi.org/10.1016/j.neuroimage.2021.117896

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences*, *116*(43), 21854–21863. https://doi.org/10.1073/pnas.1905544116

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends Cogn Sci*, *17*(1), 26–49. https://doi.org/10.1016/j.tics.2012.10.011

Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*. https://doi.org/10.3389/neuro.06.004.2008

Kubilius, J., Schrimpf, M., Kar, K., Rajalingham, R., Hong, H., Majaj, N., Issa, E., Bashivan, P., Prescott-Roy, J., Schmidt, K., Nayebi, A., Bear, D., Yamins, D. L., & DiCarlo, J. J. (2019). Brain-Like Object Recognition with High-Performing Shallow Recurrent ANNs. *Advances in Neural Information Processing Systems*, *32*. https://papers.nips.cc/paper/2019/hash/7813d1590d28a7dd372ad54b5d29d033-Abstract.html

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends Cogn Sci*, *7*(1), 12–18. https://doi.org/S136466130200013X [pii]

Leonardelli, E., & Fairhall, S. L. (2022). Similarity-based fMRI-MEG fusion reveals hierarchical organisation within the brain's semantic system. *NeuroImage*, *259*, 119405. https://doi.org/10.1016/j.neuroimage.2022.119405

Mack, M. L., & Palmeri, T. J. (2011). The timing of visual object categorization. *Frontiers in Psychology*, *2*, 165. https://doi.org/10.3389/fpsyg.2011.00165

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Martin, C. B., Douglas, D., Newsome, R. N., Man, L. L., & Barense, M. D. (2018). Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. *ELife*, *7*, e31873. https://doi.org/10.7554/eLife.31873

Mohsenzadeh, Y., Mullin, C., Lahner, B., Cichy, R. M., & Oliva, A. (2019). Reliability and Generalizability of Similarity-Based Fusion of MEG and fMRI Data in Human Ventral and Dorsal Visual Streams. *Vision*, *3*(1), Article 1. https://doi.org/10.3390/vision3010008

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

Mollo, G., Cornelissen, P. L., Millman, R. E., Ellis, A. W., & Jefferies, E. (2017). Oscillatory Dynamics Supporting Semantic Cognition: MEG Evidence for the Contribution of the Anterior Temporal Lobe Hub and Modality-Specific Spokes. *PLOS ONE*, *12*(1), e0169269. https://doi.org/10.1371/journal.pone.0169269

Mummery, C. J., Patterson, K., Wise, R. J. S., Vandenbergh, R., Price, C. J., & Hodges, J. R. (1999). Disrupted temporal lobe connections in semantic dementia. *Brain*, *122*, 61–73.

Muttenthaler, L., & Hebart, M. N. (2021). *THINGSvision: A Python toolbox for streamlining the extraction of activations from deep neural networks* [Preprint]. Neuroscience. https://doi.org/10.1101/2021.03.11.434979

Naspi, L., Hoffman, P., Devereux, B., & Morcom, A. M. (2021). Perceptual and Semantic Representations at Encoding Contribute to True and False Recognition of Objects. *Journal of Neuroscience*, *41*(40), 8375–8389. https://doi.org/10.1523/JNEUROSCI.0677-21.2021

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), Article 12. https://doi.org/10.1038/nn.3248

O'Reilly, R., Wyatte, D., Herd, S., Mingus, B., & Jilk, D. (2013). Recurrent Processing during Object Recognition. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00124

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), 8768–8776. https://doi.org/10.1523/JNEUROSCI.4888-14.2015

Rahimi, S., Farahibozorg, S.-R., Jackson, R., & Hauk, O. (2022). Task modulation of spatiotemporal dynamics in semantic brain networks: An EEG/MEG study. *NeuroImage*, *246*, 118768. https://doi.org/10.1016/j.neuroimage.2021.118768

Rahimi, S., Jackson, R., Farahibozorg, S.-R., & Hauk, O. (2022). *Time Lagged Multidimensional Pattern Connectivity (TL MDPC): An EEG/MEG Pattern Transformation Based Functional Connectivity Metric* (p. 2022.05.21.492913). bioRxiv. https://doi.org/10.1101/2022.05.21.492913

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), e1007001. https://doi.org/10.1371/journal.pcbi.1007001

Riès, S. K., Dhillon, R. K., Clarke, A., King-Stephens, D., Laxer, K. D., Weber, P. B., Kuperman, R. A., Auguste, K. I., Brunner, P., Schalk, G., Lin, J. J., Parvizi, J., Crone, N. E., Dronkers, N. F., & Knight, R. T. (2017). Spatiotemporal dynamics of word retrieval in speech production revealed by cortical high-frequency band activity. *Proceedings of the National Academy of Sciences*, *114*(23), E4530–E4538. https://doi.org/10.1073/pnas.1620669114

Ries, S., Tavildar, S., Rohilla, R., Sperling, C., & Ashrafi, A. (2020). Patterns of cortical interactivity supporting speech production and lexical retrieval: A graph signal processing approach at the individual level. *Journal of Neurolinguistics*, *56*, 100936. https://doi.org/10.1016/j.jneuroling.2020.100936

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition. *Nature*, *2*(11), 1019–1025.

Rupp, K., Roos, M., Milsap, G., Caceres, C., Ratto, C., Chevillet, M., Crone, N. E., & Wolmetz, M. (2017). Semantic attributes are encoded in human electrocorticographic signals during visual object recognition. *NeuroImage*, *148*, 318–329. https://doi.org/10.1016/j.neuroimage.2016.12.074

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), doi: 10.3389/fpsyg.2012.00329.

Schendan, H. E., & Maher, S. M. (2009). Object knowledge during entry-level categorization is activated and modified by implicit memory after 200 ms. *NeuroImage*, *44*, 1423–1438.

Seeliger, K., Fritsche, M., Güçlü, U., Schoenmakers, S., Schoffelen, J.-M., Bosch, S. E., & van Gerven, M. A. J. (2017). Convolutional neural network-based encoding and decoding of visual object recognition in space and time. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2017.07.018

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *PNAS*, *104*(15), 6424–6429.

Shafto, M. A., Tyler, L. K., Dixon, M., Taylor, J. R., Rowe, J. B., Cusack, R., Calder, A. J., Marslen-Wilson, W. D., Duncan, J., Dalgleish, T., Henson, R. N., Brayne, C., Cam-CAN, & Matthews, F. E. (2014). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: A cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC Neurology*, *14*(204). https://doi.org/10.1186/s12883-014-0204-1

Spoerer, C. J., McClure, P., & Kriegeskorte, N. (2017). Recurrent Convolutional Neural Networks: A Better Model of Biological Object Recognition. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01551

Su, L., Fonteneau, E., Marslen-Wilson, W., & Kriegeskorte, N. (2012). *Spatiotemporal searchlight representational similarity analysis in EMEG source space*. 97–100. https://doi.org/10.1109/PRNI.2012.26

Taylor, J. R., Williams, N., Cusack, R., Auer, T., Shafto, M. A., Dixon, M., Tyler, L. K., Cam-CAN, & Henson, R. N. (2017). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage*, *144*, 262–269. https://doi.org/10.1016/j.neuroimage.2015.09.018

Tyler, L. K., Stamatakis, E. A., Bright, P., Acres, K., Abdallah, S., Rodd, J. M., & Moss, H. E. (2004). Processing objects at different levels of specificity. *Journal of Cognitive Neuroscience*, *16*(3), 351–362.

Ungerleider, L., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfiled (Eds.), *Analysis of visual behavior* (pp. 549–586). MIT Press.

Warrington, E. K. (1975). The selective impairments of semantic memory. *Quarterly Journal of Experimental Psychology*, *27*, 635–657.

Wurm, M. F., & Caramazza, A. (2022). Two 'what' pathways for action and object recognition. *Trends in Cognitive Sciences*, *26*(2), 103–116. https://doi.org/10.1016/j.tics.2021.10.003

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