Flexing the principal gradient of the cerebral cortex to suit changing semantic task demands

Zhiyao Gao[1], Li Zheng[2], Katya Krieger-Redwood[1], Ajay Halai[3], Daniel S. Margulies[4], Jonathan Smallwood[5], Elizabeth Jefferies[1]

[1]Department of Psychology, University of York, Heslington, York YO10 5DD, United Kingdom

[2] Department of Psychology, University of Arizona, Tucson, AZ 85719, USA

[3]MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge CB2 1TN, United Kingdom

[4]Integrative Neuroscience and Cognition Center (UMR 8002), Centre National de la Recherche Scientifique (CNRS), France

[5]Department of Psychology, Queen's University, Kingston, ON K7L 3N6, Canada

Abbreviated title:

Flexing the gradient of semantic representation

Acknowledgements

This work was sponsored by the European Research Council (Project ID: 771863 - FLEXSEM).

Author contributions

Z.G. K.K.R and E.J. designed the experiment. K.K.R. performed the study. A.J. provided the multiple echo fMRI pre-processing analysis code. L.Z. provided assistance with data analysis and explanation. Z.G. analysed the data. Z.G. and E.J. wrote the original manuscript. All authors edited the manuscript.

Corresponding Authors

Zhiyao Gao: zhiyao.gao@york.ac.uk; Elizabeth Jefferies: beth.jefferies@york.ac.uk

Competing financial interests

The authors declare no competing financial interests.

Abstract

Understanding how thought emerges from the topographical structure of the cerebral cortex is a primary goal of cognitive neuroscience. Recent work has revealed a principal gradient of intrinsic connectivity capturing the separation of sensory-motor cortex from transmodal regions of the default mode network (DMN); this is thought to facilitate memory-guided cognition. However, studies have not explored how this dimension of connectivity changes when conceptual retrieval is controlled to suit the context. We used gradient decomposition of informational connectivity in a semantic association task to establish how the similarity in connectivity across brain regions changes during familiar and more original patterns of retrieval. Multivoxel activation patterns at opposite ends of the principal gradient were more divergent when participants retrieved stronger associations; therefore, when long-term semantic information is sufficient for ongoing cognition, regions supporting heteromodal memory are functionally separated from sensory-motor experience. In contrast, when less related concepts were linked, this dimension of connectivity was reduced in strength as semantic control regions separated from the DMN to generate more flexible and original responses. We also observed fewer dimensions within the neural response towards the apex of the principal gradient when strong associations were retrieved, reflecting less complex or varied neural coding across trials and participants. In this way, the principal gradient explains how semantic cognition is organised in the human cerebral cortex: the separation of DMN from sensory-motor systems is a hallmark of the retrieval of strong conceptual links that are culturally shared.

Significance statement

A central task in neuroscience is to understand how cognition emerges from the topographical structure of cerebral cortex. We used fMRI during a semantic task to assess informational connectivity as participants retrieved stereotypical or more unusual associations. The principal gradient of variation in informational connectivity captured the separation of heteromodal memory regions from unimodal cortex. This separation was reduced when weaker associations were retrieved; these trials also produced higher-dimensional neural responses in heteromodal regions. We conclude that the separation of DMN from sensory-motor systems is a hallmark of the retrieval of strong conceptual links that are culturally shared, while more complex and diverse cognition can be generated as this separation in connectivity is reduced.

Recent work has shown that cortical function and intrinsic connectivity change systematically along a ‘principal gradient’, which has primary sensory and motor cortex at one end, and transmodal regions implicated in abstract and memory-based functions at the other (Margulies et al., 2016). This principal gradient is the dimension of intrinsic connectivity that typically explains the most variance in whole-brain decompositions of human resting-state fMRI data (Dong et al., 2021b, Hong et al., 2020, Huntenburg et al., 2018). It is correlated with physical distance along the cortical mantle from primary sensory-motor landmarks. It captures the order of large-scale networks in frontal, temporal, and parietal regions – all of which show transitions from sensory-motor regions, through attention networks to frontoparietal control and then default mode network (DMN) regions. It also captures functional transitions: for instance, opposing ends of the principal gradient are implicated in sensory and memory-based decisions (Murphy et al., 2019, Murphy et al., 2018, Lanzoni et al., 2020). Theories suggest the principal gradient supports increasingly abstract levels of representation, allowing heteromodal regions in the default mode network (DMN) to take on roles that are less directly influenced by externally-driven neural activity (Smallwood et al., 2021, Margulies et al., 2016, Huntenburg et al., 2018, Wang et al., 2020). In line with this account, the representations of events extracted from movie-watching data are thought to vary in length along this hierarchy, with short events represented in sensory areas and longer events in transmodal regions including within the default mode network (Baldassano et al., 2017).

Semantic representation is thought to draw on both ends of this principal gradient (Huth et al., 2016, Barsalou, 2008, Kiefer and Pulvermüller, 2012, Meteyard et al., 2012, Patterson et al., 2007, Wang et al., 2010, Visser and Lambon Ralph, 2011,

Murphy et al., 2019). The graded hub and spoke model proposes that heteromodal concepts are formed at a distance from visual and auditory input regions: diverse sensory processing pathways gradually converge in transmodal areas to form abstract generalisable semantic representations (Patterson et al., 2007, Lambon Ralph et al., 2017). Moreover, we flexibly deploy conceptual information to suit our goals or context: this involves the recruitment of several heteromodal networks, including the DMN, semantic control network (SCN) and multiple-demand network (MDN), which lie at different points along the principal gradient (Wang et al., 2020, Gao et al., 2021a). These networks show differential recruitment across trials depending on the semantic similarity of the items being linked, with higher conceptual overlap associated with stronger activation at the top end of the principal gradient (Wang et al., 2020). Anterior temporal lobe (ATL) and angular gyrus (AG), which are allied with DMN, contribute to the relatively automatic retrieval of semantic information strongly supported by long-term memory (Davey et al., 2015, Jefferies et al., 2020b, Teige et al., 2019, Teige et al., 2018). In contrast, when unusual aspects of knowledge are required – for example, for weak associations – there is greater recruitment of the semantic control network (SCN) to shape the meaning that is retrieved (Davey et al., 2015, Jackson, 2021, Jefferies, 2013, Jefferies et al., 2020b), along with other control regions within the MDN, which responds to higher task demands across domains (Fedorenko et al., 2013, Erb et al., 2013, Assem et al., 2020).

While recruitment is known to vary along the principal gradient according to the demands of the task (Murphy et al., 2019, Wang et al., 2020), there have been few if any investigations of the way in which neural representation changes over the length of this whole-brain gradient depending on the type of information that is retrieved.

Local transitions in the information that is represented neurally have been observed for specific domains. Visual object recognition is supported by a processing hierarchy extending from primary visual cortex, through higher visual regions in inferior temporal cortex, to ventral anterior temporal areas that capture the meaning of concepts across modalities (Murphy et al., 2017, Ju and Bassett, 2020). A similar functional transition is proposed in medial temporal cortex, which supports detailed, fine-grained spatial representations in posterior hippocampus and more gist or schema like, coarse-grained memory representations in anterior hippocampus (Poppenk et al., 2013). Lateral prefrontal cortex has also been proposed to support a rostral–caudal gradient, representing specific motor actions in posterior regions through to more abstract goals in anterior regions (Badre and D'esposito, 2009, Badre and Nee, 2018). However, these studies have not examined changes in connectivity gradients as a function of task demands. Dimensionality reduction techniques like principal component analysis (PCA) can also quantify the number of components needed to explain a specific amount of variance in the neural response of a given brain region, providing an index of the complexity or dimensionality of the representational space over time (Ahlheim and Love, 2018). Neural representation in unimodal regions is thought to have many dimensions, allowing the cortex to discriminate between similar inputs, while the representational space may have fewer dimensions in transmodal DMN regions, allowing a variety of inputs to be encoded into common activity patterns that generalise knowledge across circumstances (Ju and Bassett, 2020). For example, we can easily tell the difference between ‘Dalmatian’ and ‘Bulldog’ yet understand their abstract similarity as ‘dogs’. Category learning based on an abstract rule is also associated with a prefrontal response characterised by fewer dimensions (Mack et al., 2020).

In the current study, we manipulated the degree to which ongoing semantic cognition was aligned with long-term semantic knowledge and quantified the similarity of the multivariate response to each trial along the principal gradient. Our task asked participants to generate links between pairs of words that varied in their associative strength, including strongly associated items which often co-occur, weakly linked items and unrelated pairs. We asked whether functional connectivity patterns during this task would mirror the principal gradient and vary flexibly according to the associative strength of the items in each trial. We also asked whether neural representations along the principal gradient differed in terms of their complexity. With less support from long-term memory, unusual patterns of conceptual retrieval might be supported by control regions that promote specific sensory-motor features, reducing the functional separation of the two ends of the principal gradient in terms of their patterns of connectivity. In contrast, frequently occurring patterns of conceptual retrieval may emerge from long-term memory in the absence of additional constraints from brain regions; this retrieval state might be associated with increased functional separation between DMN and sensory-motor regions, allowing memory to underpin cognition irrespective of ongoing sensory-motor experience. We might also expect the dimensionality of neural representations in a semantic task to decrease from unimodal to transmodal areas along the principal gradient, reflecting increasingly abstract and culturally shared representations towards the apex of the gradient, particularly when participants retrieve strong associations that link words together in specific well-established ways.

Results

Behavioural results

Participants' in-scanner ratings of the strength of the link they made between the words in each pair were strongly correlated with word2vec values as expected (Pearson $r = .84$, $p < 0.001$). We also predicted that the links formed by participants would be more consistent across individuals for word pairs that were more strongly associated, as these links could be generated from dominant aspects of long-term semantic memory, which should be shared across individuals. To test this hypothesis, we collected the subjective reports of the links that participants formed on each trial and counted how many unique meanings/links were produced. There was a significant negative correlation between word2vec and the number of meanings generated ($r = - 0.53$, $p < 0.001$), confirming that participants generated more variable patterns of semantic retrieval for weakly-associated word pairs.

On the chevron task, the average reaction time was 0.336s and the accuracy was 72.6%; these data suggest the task was challenging enough to engage the participant but not so difficult that they could not do it. The fast pace of this task reduced participants' ability to think about semantic links across trials.

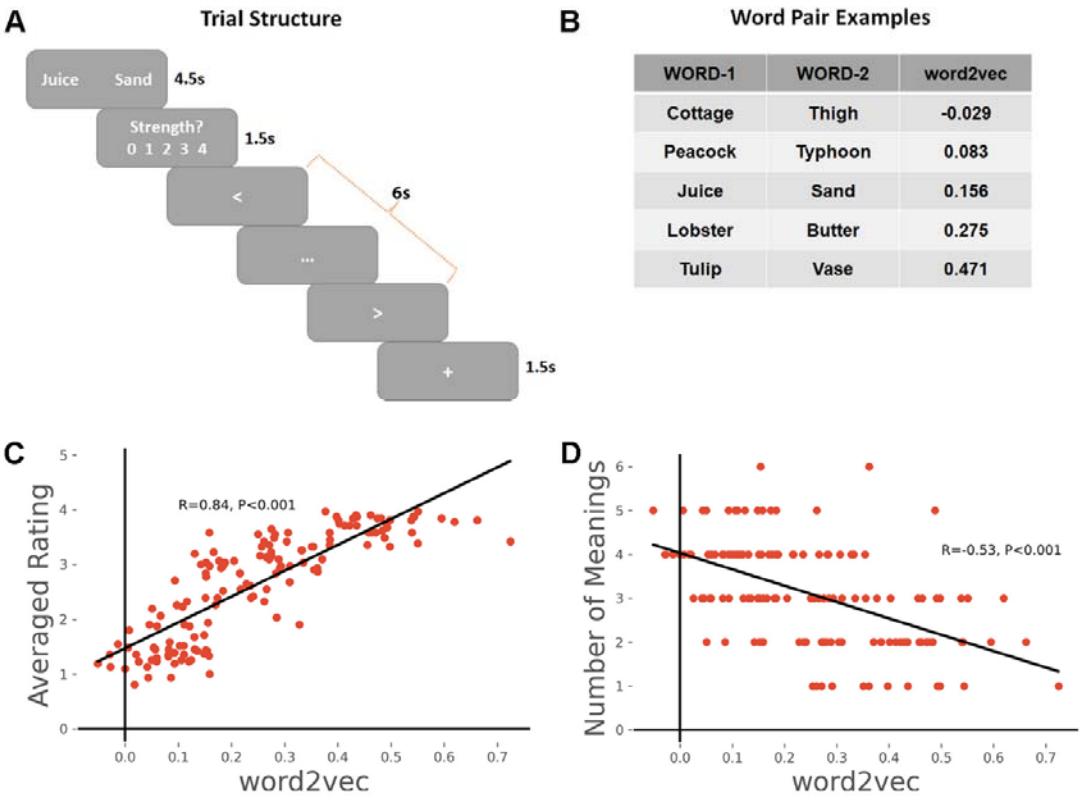

Figure 1**: A.** Task schematic; participants were asked to produce a link between the words and rate the strength of the link. B. Word pair examples with association strength ranging from weak to strong, based on word2vec scores. C. The average rating of link strength for each word-pair was positively associated with its word2vec score. D. The number of meanings produced by the group of participants for each word-pair was negatively associated with the word2vec score.

fMRI results

The principal gradient of cortical organization was modulated by strength of association

fMRI was used to assess how the topographical organisation of the neural response to the task changed as a function of strength of association. Recent research decomposing whole-brain intrinsic connectivity patterns in resting-state fMRI into their spatial components has revealed a ‘principal gradient’ of intrinsic functional connectivity, which captures the separation between transmodal memory systems

and unimodal sensory-motor cortex at rest (Margulies et al., 2016). Here, we decomposed informational connectivity matrices in the same way.

We first verified that the principal gradient of cortical organization could be identified from task-based informational connectivity matrices as well as from intrinsic patterns of connectivity at rest, and that this dimension of cortical organisation was robust across different cognitive tasks. PCA was used to decompose the group-averaged connectivity matrices, with the resulting components (or gradients) reflecting spatial patterns of connectivity across the cortex. These components were ordered by the variance they explained in the initial connectivity matrix. Within each gradient, similar values reflected similarity in patterns of connectivity. The decomposition was performed separately for the semantic task (TR 4 to 6) and the chevron task (TR 9 and 10). This similarity between cortical gradients was established by using spin permutation tests, which preserve the spatial autocorrelation present within gradients. Statistical significance was determined by comparing with a null distribution using the permutation approach (5000 permutations), in which we randomly rotated the spatial patterns and measured the correlations between maps that preserve spatial autocorrelation (Alexander-Bloch et al., 2018).The first gradient, accounting for the greatest variance in informational connectivity, was highly overlapping with the principal gradient of intrinsic connectivity identified by prior work (Margulies et al., 2016), for both the semantic and chevron tasks (Spin permutation p value: $P_{spin} < 0.001$), see Supplementary Figure S1. The gradient which explained the second most variance for the semantic task was also positively correlated with the corresponding gradient derived from resting data ($P_{spin} < 0.001$), demonstrating a pattern of distinguishing visual areas from auditory/motor areas. The gradient which explained third most variance for the semantic task was only positively correlated

with the corresponding gradient derived from resting data in the left hemisphere ($P_{spin} < 0.001$), while a negative relationship was found for the right hemisphere, see Supplementary Figure S2.

Next, we examined our core hypothesis that the principal gradient of cortical organization is flexed according to the demands of cognitive tasks. Since our behavioural data highlights how more strongly associated word pairs elicit more similar patterns of semantic retrieval across participants, the neurocognitive response in these trials is expected to reflect the retrieval of readily accessible semantic information from long-term memory, which is shared across people. This retrieval state is assumed to be driven by relatively abstract conceptual representations in transmodal regions as opposed to sensory-motor features that vary across experiences; consequently, the retrieval of strong associations might strengthen the principal gradient (i.e., it might drive functional separation between transmodal and unimodal cortex), relative to the more idiosyncratic and controlled patterns of retrieval required for weaker associations and unrelated words.

We derived estimates of the principal gradient of informational connectivity for strong and weak associations separately, dividing trials according to their word2vec values and focussing on the semantic response in TR 4 to 6. This analysis revealed a significant increase in principal gradient values for strong versus weak associations in transmodal areas implicated in semantic cognition, including in bilateral prefrontal cortex, temporal lobe and presupplementary motor cortex. These regions, shown in warm colours on Figure 2B, are found towards the DMN apex of the principal gradient. The reverse of this pattern (i.e., lower principal gradient values for strong versus weak associations) was found in sensory-motor regions towards the bottom of the principal gradient. These regions are shown in cool colours on Figure 2A. The

principal gradient difference between middle and weak association trials of the semantic task showed a similar pattern, see Supplementary Figure S10.

In a control analysis, we also examined the chevron task period (TR 9 and 10) for the same subsets of trials (i.e. divided on the basis of word2vec values for the immediately preceding semantic decision). A different pattern was found: although there were still differences between these sets of trials that might have reflected the consequences of earlier variation in cognitive effort or retrieval, these differences were substantially weaker for the chevron than the semantic period in transmodal regions (see Figure 2C). In addition, these differences were not systematically aligned with the principal gradient: some DMN regions showed higher gradient values for chevron decisions following strong versus weak associations (for example, in right lateral temporal and right medial prefrontal cortex), while other DMN regions showed the reverse pattern (for example, in left angular gyrus). In addition, visual and motor cortex did not generally show lower principal gradient values for chevron trials following strong versus weak associations; these regions often showed opposing effects for the chevron and semantic TRs, see Supplementary Figure S3.

To examine how the principal gradient was modulated by the tasks within pre-defined functional networks that are relevant to semantic representation and control demands, we conducted a two-way repeated-measures ANOVA examining the within-participant factor of task (semantic versus chevron) in different brain networks. These networks corresponded to (i) regions implicated in semantic cognition that fell outside a meta-analytic map of semantic control (Jackson, 2021) – focused on anterior temporal cortex and angular gyrus; (ii) semantic control network regions

(SCN) showing activation across a range of manipulations of semantic task difficulty (Jackson, 2021) yet which are outside the multiple-demand network (MDN) (Fedorenko et al., 2013) – including left anterior inferior frontal gyrus and posterior middle temporal gyrus; (iii) regions falling in both SCN and MDN (SCN+MDN) – including inferior frontal sulcus and pre-supplementary motor area, and (iv) MDN regions not implicated in semantic cognition; see methods). We found a significant interaction between brain network and task ($F(2.53, 75.887) = 8.345$, $p < 0.001$, $\eta_p^2 = 0.218$) and significant main effects of task ($F(1, 30) = 21.947$, $p < 0.001$, $\eta_p^2 = 0.414$) and network ($F(2.01, 60.288) = 3.338$, $p = 0.042$, $\eta_p^2 = 0.100$), see Figure 2D. Simple t-tests revealed significant differences in principal gradient values for the semantic task across SCN and SCN+MDN compared with MDN regions ($t(30) = 3.259$, $p = 0.0167$, 95% CI = [0.279 ~ 1.215], Cohen's d = 0.827; and $t(30) = 3.904$, $p = 0.003$, 95% CI = [0.334 ~ 1.066], Cohen's d = 0.965, respectively), while no significant differences between networks were found for the chevron task ($p > 0.682$). Bonferroni correction was applied. These results show that networks implicated in semantic representation and control, but not executive control across domains, have lower gradient values when weaker associations are retrieved, suggesting less separation in connectivity from sensory-motor cortex.

To quantify the semantic task principal gradient difference between strong and weak associations at a finer-grained scale, we used Schaefer's 400-region whole-brain parcellation. We assessed the correlation between the gradient difference value for each parcel (between strong and weak associations) and resting state gradient reported by Margulies et al. (2016). There was a positive correlation for the semantic task: parcels higher on the resting state gradient tended to show a larger effect of associative strength (strong > weak) on task-derived gradient values ($r = 0.369$, p <

0.001), while no correlation was found in the chevron task (r = -0.038, p = 0.45), see Figure 2E~F. These results confirmed that variation in the strength of association between words influenced task-induced gradients of informational connectivity, with brain regions at the heteromodal end of the principal gradient showing more similar connectivity patterns to each other, and more separation from unimodal regions, during semantic decisions involving strongly linked items. These results were confirmed using spin permutation to control for the effects of spatial autocorrelation: again, there was a significant correlation between the resting state gradient and the difference of gradient values between strong and weak associations for the semantic response (Pspin < 0.001), while no such effect was found for the chevron task (Pspin = 0.195).

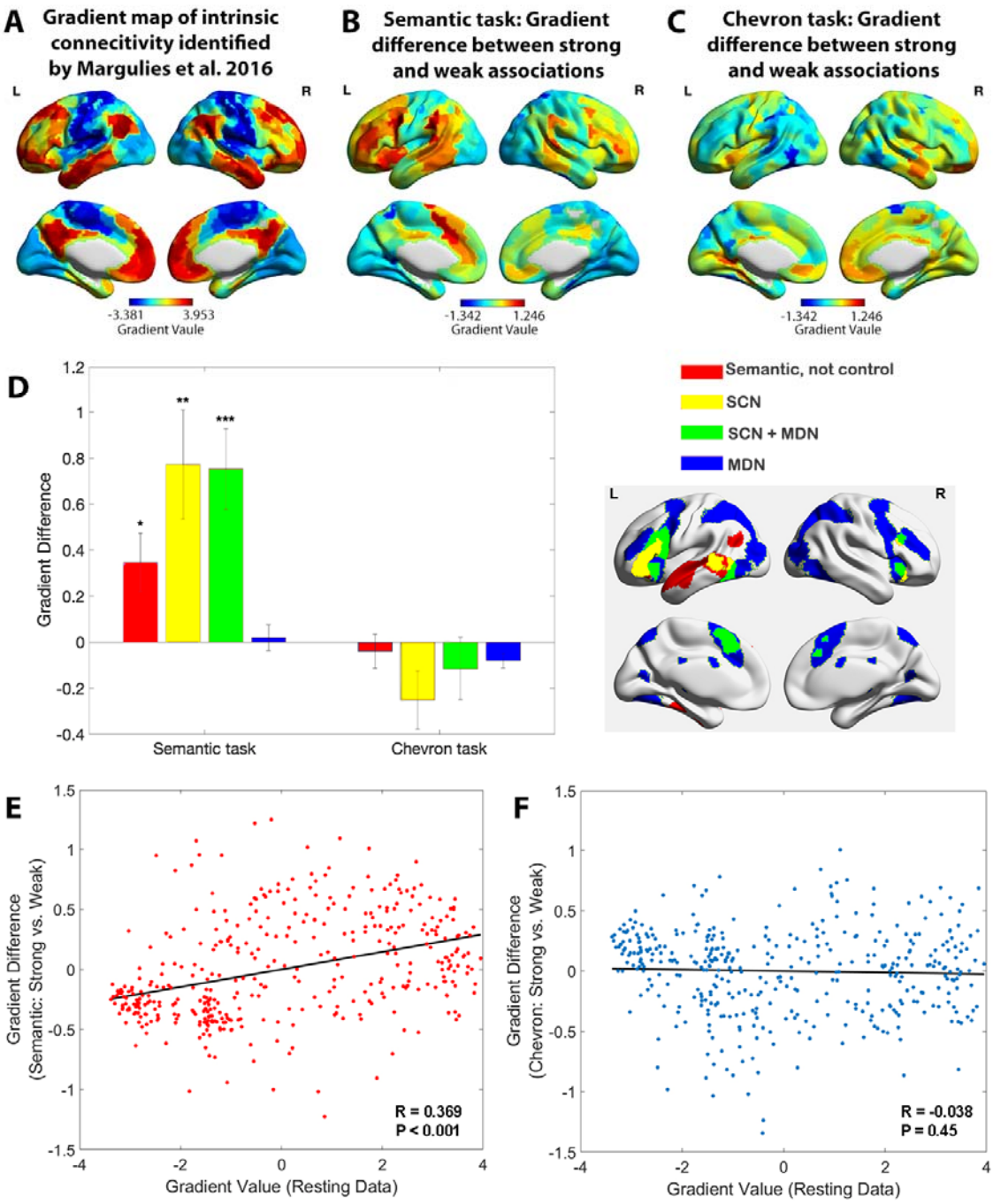

Figure 2. A. The principal gradient map identified from intrinsic connectivity reported by Margulies et al. (2016). B. The difference map of principal gradient values during the semantic task comparing strong and weak associations estimated using the informational connectivity approach and brain responses in TR 4-6. C. The difference map of principal gradient values during the chevron task, comparing time points that followed strong and weak associations, estimated using the informational connectivity approach and brain responses in TR 9-10. D. Left panel. The difference in principal

gradient values between strong and weak association trials showed a significant interaction effect between functional networks and tasks ($F(2.53, 75.887) = 8.345$, $p < 0.001$, $\eta_p^2 = 0.218$). Error bars represent the standard error of gradient difference across parcels within each network. Right panel. Functional networks: (i) semantic not control, (ii) within the semantic control network (SCN) but outside multiple-demand cortex (DMN), (iii) within both SCN and MDN, and (iv) falling in MDN regions not implicated in semantic cognition. E. Using Schaefer's 400 parcellation, gradient difference values in each parcel for strong and weak associations in the semantic task were positively associated with the principal gradient values of these parcels, estimated from independent resting data. F. No correlation between informational connectivity differences per parcel and principal gradient values estimated from resting-state fMRI data was observed for the chevron task when comparing timepoints that followed strong and weak semantic associations.

Representational space changed along the principal gradient

The analysis above suggests that the functional role of cortical parcels in semantic cognition depends on their location on the principal gradient. This gradient captures a functional hierarchy of processing that extends from sensory-motor features to more abstract long-term memory representations (Wang et al., 2020, Smallwood et al., 2021, Margulies et al., 2016, Huntenburg et al., 2018), echoing the way in which semantic cognition draws on representations from sensory-motor features to abstract and heteromodal conceptual information in long-term memory (Lambon Ralph et al., 2017, Patterson et al., 2007). Given that strong associations are associated with greater separation of sensory-motor features from memory-based cognition along the gradient, we might also expect regions at the apex of the gradient to reflect simpler representational states when strong associations are being retrieved. This is because the long-term semantic knowledge underpinning strong associations is highly predictable, abstract and shared across participants.

We used the dimensionality or heterogeneity of the voxels' responses in each parcel to index the complexity of neural representation. We applied a searchlight-based PCA approach to the neuroimaging data for each trial, concatenated across participants. By calculating the number of components which together explained more than 90% of the variance in each local cube, we generated a neural dimensionality map for each trial. Figure 3A shows this dimensionality map averaged across all trials. Some heteromodal regions at the top of the principal gradient overlapping with DMN showed activity characterised by many dimensions (e.g., lateral and medial temporal cortex), while other regions with similar gradient values had fewer dimensions (inferior parietal and precuneus cortex).

To quantify the degree to which the dimensionality followed the resting state principal gradient, we correlated dimensionality estimates with principal gradient values derived from intrinsic connectivity data by Margulies et al. (2016). We also assessed whether this relationship was influenced by associative strength, averaging estimates for sets of four trials with similar word2vec values using a 'sliding window' approach. The results showed a weak negative correlation overall (when all trials were included) between dimensionality and principal gradient values ($r = -0.155$, $p = 0.0018$), suggesting that the dimensionality of neural coding decreased, at least to some extent, along the principal gradient from unimodal to heteromodal regions. Importantly, the magnitude of this negative correlation increased as a function of association strength ($r = -0.425$, $p < 0.001$, see Figure 3D): more strongly associated word pairs generated neural responses at the heteromodal end of the gradient which had even lower dimensionality. This pattern remained significant even when regions in limbic network were removed due to their low tSNR ($r = -0.346$, $p = 0.038$). Additional control analyses assessed the dimensionality in the chevron task and its

relation to the strength of association in the preceding semantic trial. There was a weak negative corelation between overall dimensionality (averaged across all trials) and the principal gradient estimated by resting state data ($r = -0.08$, $p < 0.001$), in line with the expectation of more abstract representations at the top of the gradient. The relationship between the resting state principal gradient and dimensionality (estimated for sets of four trials with similar strength of association) did not vary with the association strength of the preceding semantic trial ($r = -0.18$, $p = 0.28$, see Supplementary Figure S5).

A direct comparison of weak and strong associations (contrasting the bottom and top third of the trials arranged according to associative strength) revealed that the neural response had significantly higher dimensionality for weakly-related trials in regions implicated in semantic control, including inferior frontal gyrus, presupplementary motor cortex and posterior middle temporal gyrus in the left hemisphere, as well as in regions of the default mode network including the bilateral angular gyrus, precuneus, medial prefrontal cortex and left anterior temporal pole (Figure 3B). No clusters showed the reverse pattern (i.e., higher dimensionality for activation patterns during the retrieval of strong associations). Confirmatory analyses examined the effect of association strength on dimensionality in networks defined using resting state data (Yeo et al., 2011) and functional recruitment during semantic and control tasks. This confirmed the strongest dimensionality difference was within semantic control areas, with no effect of association strength in multiple demand areas, see Figure 3C and Supplementary Figure S6. To check the robustness of our results, we performed a series of control analyses in which the dimensions of the activation pattern were determined using different criteria (explaining 60% or 75% of the variance) and obtained highly similar results, see Supplementary Figure S7. Using

Schaefer's 400-region parcellation, we established a significant positive correlation between the dimensionality change across weak and strong associations and the principal gradient value for each parcel obtained from both intrinsic connectivity (r = 0.352, $p < 0.001$, Figure 3E) and the informational connectivity of our semantic task (r = 0.549, $p < 0.001$, Figure 3F). These results show that the semantic representational space is more strongly modulated by association strength towards the heteromodal end of the principal gradient, implicating these regions in individual differences in semantic cognition.

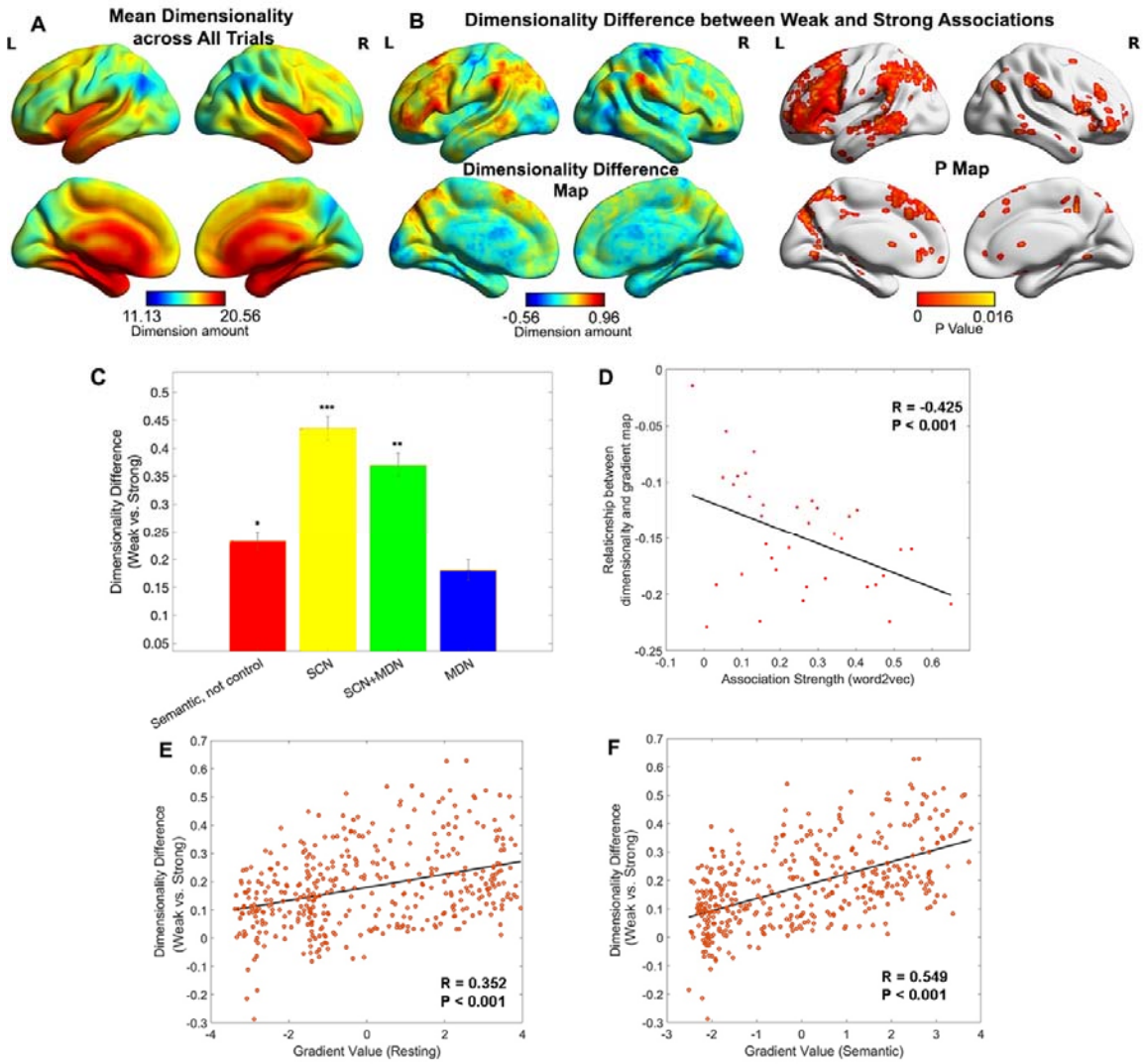

Figure 3. A. Dimensionality map averaged across all trials. B. Dimensionality difference between weak and strong associations (left panel) and the corresponding p map (right panel), with FDR correction applied, q = 0.05. C. Dimensionality difference between weak and strong in functional networks implicated in semantic representation and control demands (Semantic, not control: $p = 0.017$, SCN: $p < 0.0001$, SCN+MDN: $p = 0.002$, MDN: $p = 0.245$, Wilcoxon rank sum test was conducted and p values were adjusted using Bonferroni correction). Error bars represent the standard error of dimensionality differences across parcels within each network. D. The relationship between dimensionality and the principal gradient was generally negative (all data points are below 0) and negatively correlated with the associative strength between word-pairs. We averaged the dimensionality map for every four trials arranged in order from weak to strong according to their word2vec values for this analysis. E and F. The dimensionality difference between weak and strong trials was positively correlated with the location of each parcel on the principal gradient estimated from resting-state fMRI (in E) and informational connectivity during the semantic task (in F). Each datapoint represents a parcel from the Schaffer 400 parcellation. Note: the dimensionality difference was examined by subtracting strong from weak associations; this contrast is in the opposite to the gradient difference analysis, since we anticipate higher gradient values and fewer neural dimensions for strong associations in similar brain regions.

Relationships between the principal gradient, representational dimensionality, and semantic cognition

To identify regions in which neural activity patterns could predict differences in semantic cognition between participants, we performed a second-order RSA to examine the semantic-brain alignment using a whole-brain searchlight approach. A semantic similarity matrix, based on the correlation of participants' ratings of associative strength across trials (reflecting the global similarity of neural states of retrieval between participants; left-hand panel of Figure 4A), was positively associated with neural pattern similarity in inferior frontal gyrus, posterior middle temporal gyrus, right anterior temporal lobe, bilateral lateral and medial parietal cortex, pre-supplementary motor area, and middle and superior frontal cortex (right-

hand panel of Figure 4A). Functional masks defined by semantic and control tasks and classical resting state networks (Yeo et al., 2011) confirmed that semantic control regions and DMN showed the strongest association with individual differences in associative strength ratings across participants, see Supplementary Figure S8. Therefore, heteromodal regions implicated in semantic representation (within the default mode network) and semantic control were associated with individual differences in semantic cognition across participants.

Next, we examined whether the strength of the semantic-brain alignment could be predicted by either dimensionality differences between strongly associated and weakly associated trials or the semantic task principal gradient. Using Schaefer's 400-region parcellation, we assessed the correlation between these three variables across parcels. Semantic-brain alignment was positively associated with dimensionality differences between strongly associated and weakly associated trials ($r = 0.506$, $p < 0.001$, Figure 4C), and with the principal gradient derived from the semantic task ($r = 0.42$, $p < 0.001$, Figure 4B) and from intrinsic connectivity ($r = 0.255$, $p < 0.001$; see Supplementary Figure S9). These analyses show that individual differences in semantic cognition were associated with variation in the neural response more strongly in parcels showing less variable or complex neural responses to strongly associated items, and towards the DMN apex of the principal gradient.

Given that the principal gradient correlated with both differences in the dimensionality of neural responses between strong and weaker associations and semantic-brain alignment, we reasoned that the hierarchy of brain organization captured by the principal gradient might influence semantic cognition via effects on neural dimensionality. To test this hypothesis, we performed a mediation analysis in

which the principal gradient was the predictor, semantic-brain alignment was the dependent variable, and the effect of strength of association on the dimensionality of the neural response was the mediator. The significance of the mediation effect was tested by using the bootstrapping approach with 5000 samples. Consistent with our prediction, we found the correlation between the principal gradient and semantic-brain alignment could be partially explained by this dimensionality difference (indirect effect: a x b = 0.0053, 95% CI = [0.0039 ~ 0.01], $p < 0.001$, explaining 51.7% variance; direct effect: c' = 0.0049, 95% CI = [0.0026 ~ 0.01], $p < 0.001$). This result suggests the apex of the principal gradient might relate to individual differences in semantic cognition because its behaviour reflects the extent to which information is common across individuals: when there is greater semantic-brain alignment at the top of the gradient, the neural patterns for strong associations also show less variation and complexity than for weak associations (Figure 4D).

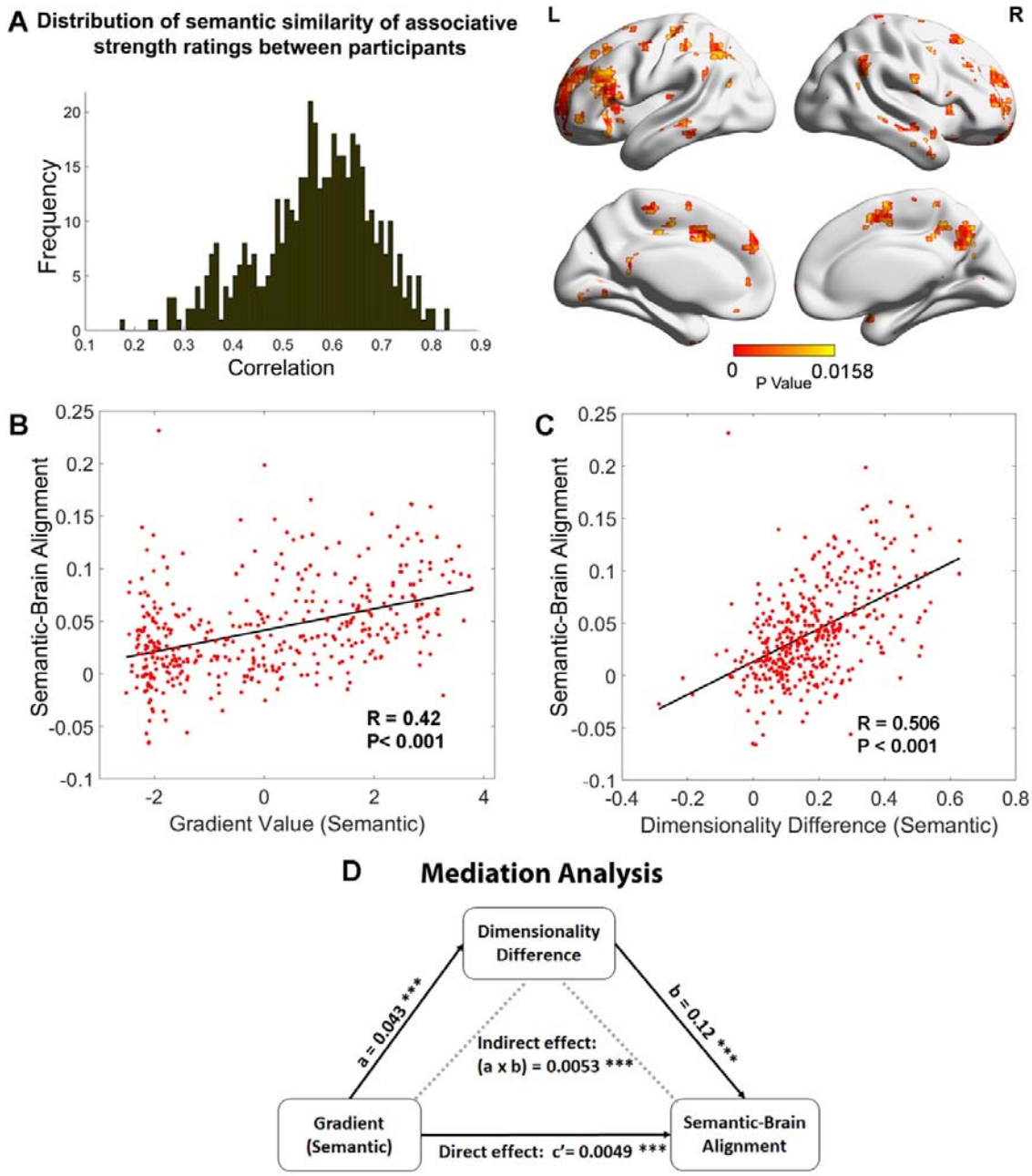

Figure 4. A. Left: The distribution of similarity in associative strength ratings between participants. We measured the semantic similarity by correlating the rating patterns of associative strength between any pair of participants (x-axis: correlation) and plotted the distribution of semantic similarity (y-axis: frequency). Right: More similar neural patterns between participants and trials were positively associated with similarity in semantic ratings; FDR correction was applied, q = 0.05. B. The semantic task principal gradient was positively associated with semantic-brain alignment shown in the right-side of Figure 4A; each point represents one parcel. C. The difference in the dimensionality of the neural

response between weak and strong associations (weak - strong) was also positively associated with the semantic-brain alignment estimated from Figure 4A; each point represents one parcel. D. Mediation analysis. The semantic task principal gradient was the independent variable, the dimensionality difference between weak and strong associations was the moderator and semantic-brain alignment was the dependent variable. This analysis revealed that the relationship between the semantic task principal gradient and semantic-brain alignment could be partially explained by the difference in dimensionality of the neural representational space between strong and weak trials (explained 51.7% variance), Ps < 0.001.

Discussion

This study modulated the associative strength between word-pairs to delineate the neural patterns underpinning semantic retrieval along the principal gradient of cortical organization. We combined dimensionality reduction methods (gradient analysis) with informational connectivity capturing the multivariate similarity between brain regions during the retrieval of semantic associations. This analysis showed that the presentation of strong associations increased the functional separation between unimodal and transmodal regions at opposite ends of the principal gradient. Stronger associations elicited higher gradient values in transmodal regions and lower gradient values in unimodal regions, suggesting that commonly occurring patterns of cognition mediated by long-term memory involved greater informational separation from sensory-motor features. In contrast, more idiosyncratic patterns of retrieval involved less separation between transmodal and unimodal representations, perhaps because participants fell back on specific features to generate a link when dominant aspects of long-term memory were insufficient for the task. This change was greatest in semantic control regions, which were close to the apex of the principal gradient along with DMN at rest and during the retrieval of strong associations, yet showed substantially lower gradient values during the retrieval of weak associations. We also

found a larger decrease in the dimensionality of neural representations from unimodal to transmodal regions along the principal gradient when strong associations were retrieved; the response towards the DMN apex of the gradient was less complex or variable across participants for strong associations, reflecting the way that these trials relied on aspects of knowledge or retrieval states that could generalise across the specific experiences of individuals. The neural representation of weakly associated word-pairs was associated with higher dimensionality or heterogeneity in distributed transmodal areas, including in lateral DMN and semantic control regions. Finally, mediation analysis showed that the relevance of cortical gradients to individual differences in rated semantic similarity was partially explained by neural dimensionality differences between weak and strong associations. Therefore, the neural organisation of semantic cognition not only reflects global patterns of brain connectivity but changes in global connectivity patterns associated with variation in transmodal neural representation.

Notably, while most gradient studies generate connectivity matrices by correlating mean univariate signals between ROIs over time, normally using resting-state fMRI data (Huntenburg et al., 2018, Margulies et al., 2016), we employed a novel informational connectivity approach to measure the similarity of multivariate response patterns between ROIs during a task (Anzellotti and Coutanche, 2018, Coutanche and Thompson-Schill, 2013). The cortical gradient explaining the most variance in informational connectivity was highly correlated with the principal gradient estimated from resting data (Margulies et al., 2016) using the traditional univariate approach[1]. Moreover, the variance explained by the principal gradient was

[1] This relationship between task-based informational connectivity and resting-state fMRI analysis was also found for the gradient which explained the second most variance for the semantic task. The gradient which explained the third most variance showed a moderate correlation with the third

higher for the decomposition of informational connectivity than for time-series correlation (40~50% versus 25%). Therefore, the current study helps to validate a new approach to deriving cortical gradients from task data which was sensitive to the effects of associative strength in semantic cognition.

The human capacity to retrieve abstract meaning from sounds, visual objects, words and phrases exceeds the capacities of other animal species; however, the development of this ability is poorly understood (Enge et al., 2021). One recent study (Dong et al., 2021a) found that, in children, the overarching organizational gradient is anchored within the unimodal cortex, between somatosensory/motor and visual territories, while in adolescence, the principal gradient of connectivity transitions into an adult-like spatial framework, with the default network at the opposite end of a spectrum from primary sensory and motor regions. Whether the development of abstract semantic representations is related to this gradual emergence of the principal gradient in the brain remains an open question which can be addressed in future studies.

Our findings are broadly consistent with the Controlled Semantic Cognition framework (Jefferies et al., 2020b, Lambon Ralph et al., 2017), which proposes that heteromodal semantic cognition emerges from the distillation of sensory-motor features into more abstract concepts; semantic control processes then shape the responses across this network to suit the demands of the task. There is strong overlap between the semantic network recruited during tasks and DMN regions identified from the parcellation of resting-state data, for example in inferior and lateral temporal and lateral parietal cortices (Davey et al., 2016, Jackson et al., 2016,

---

gradient derived from resting state data in the left hemisphere. However, the correlation was negative in the right hemisphere, suggesting an influence of the task: semantic cognition might be expected to separate the connectivity patterns of left and right hemispheres, since it is left-lateralised.

Seghier et al., 2010, Humphreys and Lambon Ralph, 2015b). When semantic retrieval is focused on dominant features of conceptual knowledge, automatic retrieval might emerge from the unconstrained behaviour of these semantic 'hub' regions, including ATL and potentially other sites with similar patterns of connectivity, such as left angular gyrus (Humphreys and Lambon Ralph, 2015a). Gradient analysis of both task and rest data shows that these transmodal regions have a high degree of pattern similarity to each other, and weaker similarity to sensory-motor cortex. Long-term knowledge amplifies the informational distance between the opposite ends of the principal gradient, suggesting that when strong associations are retrieved, areas associated with abstract and heteromodal conceptual representation separate from specific sensory-motor features and attentional networks at the opposite end of the gradient, as these elements are not needed for task-relevant patterns of retrieval. In contrast, when retrieval is focused on unusual aspects of semantic knowledge, semantic control processes shape retrieval to suit the demands of the task – for example, inhibiting dominant associative links and retrieving specific features that support specific and unusual connections between words (Jackson, 2021, Jefferies et al., 2020a, Gao et al., 2021b, Noonan et al., 2013a). In these circumstances, stronger activation is seen within the 'semantic control network' comprising inferior frontal gyrus, posterior middle temporal gyrus and pre-supplementary cortex (Jackson, 2021, Noonan et al., 2013b, Gao et al., 2021a).

We found a positive linear relationship between the difference of gradient values across weak and strong associations and principal gradient location, providing evidence that regions towards the DMN apex can flex their patterns of connectivity depending on the demands of the task. Semantic control regions showed the largest changes in gradient values and dimensionality between strong and weak

associations, suggesting that these regions separate from other heteromodal areas towards the DMN apex of the gradient and interact more strongly with attentional networks and unimodal regions to generate unusual connections between words. While our results are in line with the controlled semantic cognition framework in general, multiple cognitive processes might contribute to the differences between stronger and weaker associations in our task, including the requirement to generate a novel link, as opposed to recognising an existing link. There are also likely to be differing demands on more automatic versus more controlled retrieval, imagery and processes associated with creativity.

Our gradient analyses are highly consistent with inter-voxel similarity analysis across participants and trials. Using this approach to index the complexity of neural representations, researchers have revealed a graded transition from coarse to fine-grained representations along the long axis of the hippocampus from posterior to anterior regions (Brunec et al., 2018). Lower inter-voxel similarity over time in the hippocampus has been linked to more differentiated memory representations (Chanales et al., 2017). Motivated by these studies, we measured the heterogeneity of the voxels' responses in local areas to assess the dimensionality of representational space during semantic retrieval. We found a dimensionality decrease from unimodal to transmodal regions along the principal gradient which was larger for strong associations. This finding is consistent with recent studies showing that multivariate patterns within transmodal areas of ventral ATL represent meaning irrespective of presentation modality, feature input or language (Murphy et al., 2017, Coutanche and Thompson-Schill, 2015, Correia et al., 2014) – this abstraction of conceptual representations away from surface features of the input is likely to be critical for the generalisation of knowledge across modalities and diverse

contexts and to support shared understanding across individuals. In line with this view, higher dimensional spaces are thought to enable the discrimination of similar inputs but consume more resources to embed the information, while lower dimensional spaces facilitate generalization and novel learning across stimuli or situations (Badre et al., 2021, Ju and Bassett, 2020). DMN regions near the apex of the principal gradient may support more abstract representations of meaning when the retrieval demands of a task are well-aligned with the information stored in long-term semantic memory.

Consistent with this hypothesis, neural representations showed higher dimensionality for weak associations in transmodal regions, including the semantic ‘hub’ regions in ATL and angular gyrus, and semantic control regions in lateral and medial prefrontal, parietal and temporal cortex (Humphreys and Lambon Ralph, 2015b, Jackson et al., 2016, Jefferies, 2013, Jefferies et al., 2020b, Ralph et al., 2017, Seghier et al., 2010). Notably, the dimensionality difference peaked in semantic control regions, suggesting the semantic control network is important for promoting flexible semantic cognition (Jefferies et al., 2020b). In this way, our results extended prior observations from univariate analyses by revealing that more complex representations and/or processes were implemented in transmodal areas during more controlled patterns of retrieval. Although regions towards the DMN apex of the principal gradient were associated with lower dimensionality of neural representation, we also observed higher sensitivity to the manipulation of associative strength in these areas. This suggests that transmodal regions adopt low-dimensional states when retrieval is adequately constrained by semantic long-term memory and higher-dimensional states when more idiosyncratic patterns of retrieval are generated.

It is worth noting that not all brain regions showed the expected pattern in the dimensionality analysis – especially when considering the global dimensionality of all semantic trials, as opposed to the influence of strength of association in the semantic task. In particular, the limbic network, including regions of ventral ATL thought to support a heteromodal semantic hub, showed significantly higher dimensionality than sensory-motor areas – these higher-order regions are expected to show lower dimensionality corresponding to more abstract representations. However, this analysis does not assess the psychological significance of data dimensionality differences (unlike our contrast of strong and weak associations, which are more interpretable in terms of semantic cognition). Limbic regions are subject to severe distortion and signal loss in functional MRI, which might strongly influence this metric. Future studies using data acquisition and analysis techniques that are less susceptible to this problem are required to fully characterize global dimensionality and its relation to the principal gradient.

In conclusion, we found that when strong associations were retrieved, and therefore cognition was well-aligned with semantic information in long-term memory, the principal gradient was strengthened, suggesting greater separation of heteromodal and unimodal connectivity. When less related concepts were linked, this dimension of connectivity was reduced in strength as semantic control regions near the apex of the gradient separated from DMN to generate more flexible and original responses. In addition, semantic regions near the apex of the principal gradient showed higher dimensionality for weak associations, with activation patterns in these regions predicting individual differences in the similarity of semantic ratings. These results reveal that the principal gradient provides an organizing principle for semantic cognition in the cerebral cortex.

Method:

Participants:

A group of 36 healthy participants aged 19 to 35 years (mean age = 21.97 ±□3.47 years; 19 females) was recruited from the University of York. They were all right-handed, native English speakers, with normal or corrected-to-normal vision and no history of psychiatric or neurological illness. The study was approved by the Research Ethics Committee of the York Neuroimaging Centre. All volunteers provided informed written consent and received monetary compensation or course credit for their participation. Five participants were not included in the data analysis: one participant had poor behavioural performance (no link made on 32% of trials), another withdrew during scanning, one scan showed a structural anomaly, one had missing volumes and another showed excessive head movement (movement > 1.55mm). Of the 31 participants included in the analysis, none showed (absolute) movement greater 1.2 mm. This study provides a novel analysis of a dataset first reported by Krieger-Redwood et al. (Krieger-Redwood et al., 2022).

Task materials and procedure:

The stimuli were 144 English concrete noun pairs. Because abstract and concrete meanings, and taxonomic and thematic semantic relations, at least partially recruit different brain mechanisms (Mkrtychian et al., 2019, Mirman et al., 2017, Montefinese, 2019, Wang et al., 2010, Schwartz et al., 2011), we excluded abstract nouns and pairs of items drawn from the same taxonomic category, so that only thematic links were evaluated. We manipulated semantic association strength between the pairs of words using word2vec, with word-pairs ranging from completely unrelated (minimum word2vec = -.05) to highly related (maximum word2vec = .72).

The association strength did not show significant correlation with word frequency ($r = -0.010$, $p = 0.392$), concreteness ($r = -0.092$, $p = 0.285$) or imageability ($r = 0.074$, $p = 0.377$). Participants were viewed these word-pairs on screen, for 4.5 seconds, and were asked to identify a link between the items during this time. Participants rated the strength of the link they formed after each trial on a 5-point scale (0, no link; 1~4, weak to strong).

We used a slow event-related design, with trials lasting 13.5s. Each trial began with a visually presented word-pair for 4.5s; during this period, participants were tasked with identifying a link between the words. Following the link generation for each trial, there was 1.5s to rate the strength of the link they retrieved. Semantic trials were separated by an easy chevron task for 6s: participants pressed buttons to indicate whether each chevron faced left or right, with 10 chevrons presented. Finally, there was 1.5 seconds of fixation to alert the participant to the upcoming trial. Participants finished 3 functional runs, each lasting 11 minutes and 45 seconds. Immediately following the scanning session, participants were asked to recall and describe the link that they formed in the scanner. Using participants' post-scan recall of the links that they formed, we analysed the uniqueness of each response. These values were expressed as a proportion of the total sample who gave that particular response, ranging from .03 (a minimum of 1/31 participants) to 1 (a maximum of 31/31 participants).

FMRI acquisition:

Whole brain structural and functional MRI data were acquired using a 3T Siemens Prisma MRI scanner utilising a 64-channel head coil, tuned to 123 MHz at the York Neuroimaging Centre, University of York. Three whole brain functional runs were

acquired using a multi-band multi-echo (MBME) EPI sequence, each lasting for 11.45 minutes (TR = 1.5; TEs = 12ms, 24.83ms, 37.66ms; flip angle θ = 75°; FoV = 240 x 240 mm, resolution matrix = 80 x 80, and slice thickness = 3mm; 455 volumes per run; GRAPPA acceleration factor = 3; multi-band acceleration factor = 2). Structural T1-weighted images were acquired using an MPRAGE sequence (TR = 2.3 s, TE = 2.26 s; flip angle = 8; FOV = 256 × 256 mm, matrix = 256 × 256, and slice thickness = 1 mm). A total of 176 sagittal slices were acquired to provide high-resolution structural images of the whole brain.

fMRI data analysis:

Multi-echo data pre-processing:

This study used a multiband multi-echo (MBME) scanning sequence. We used TEDANA to combine the images (DuPre et al., 2021, Kundu et al., 2013, Kundu et al., 2012). Before images were combined, some pre-processing was performed. FSL_anat (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/fsl_anat) was used to process the anatomical images, including re-orientation to standard (MNI) space (fslreorient2std), automatic cropping (robustfov), bias-field correction (RF/B1 – inhomogeneity-correction, using FAST), linear and non-linear registration to standard-space (using FLIRT and FNIRT), brain extraction (using FNIRT, BET), tissue-type segmentation (using FAST) and subcortical structure segmentation (FAST). The multi-echo data were pre-processed using AFNI (https://afni.nimh.nih.gov/), including de-spiking (3dDespike), slice timing correction (3dTshift), and motion correction (3daxialise, deoblique) of all echoes aligned to the first echo (with a cubic interpolation). The script used to implement the pre-processing TEDANA pipeline is available at OSF (https://osf.io/mkgcy/). The TEDANA outputs (dn_ts_oc) in native space were filtered

in the temporal domain using a nonlinear high-pass filter with a 100 s cut-off. A two-step registration procedure was used whereby EPI images were first registered to the MPRAGE structural image (Jenkinson and Smith, 2001). Registration from MPRAGE structural image to standard space was further refined using FNIRT nonlinear registration (Andersson et al., 2007). Functional images were resampled to 3 mm isotropic voxels for all analyses. Because the BOLD signal reaches its peak around 4 ~ 8s after the stimuli's onset, the brain data were averaged within-trial across three time points for the semantic task (4.5~9s after trial onset; TRs 4-6). To avoid potential contamination of neural responses from task-switching processes and the forthcoming semantic trial, we selected two time points for the chevron task (12~15s after trial onset; TRs 9-10), resulting in one pattern of brain activity for the semantic and chevron tasks per trial. All analyses were performed in standard volume space.

Network definition

The networks were taken from previous meta-analytic studies of the semantic control network (SCN) and multiple-demand network (MDN) (Jackson, 2021, Fedorenko et al., 2013). Within these networks, we selected 1) semantic control specific areas, which did not overlap with MDN; 2) multiple-demand specific regions, which did not overlap with SCN; 3) shared control regions, identified from the overlap between MDN and SCN; and 4) semantic regions not implicated in control; these were identified using Neurosynth (search term 'semantic'; 1031 contributing studies; http://www.neurosynth.org/analyses/terms/), removing regions that overlapped with the two control networks to identify regions associated with semantic representation or more automatic aspects of semantic retrieval, mostly within DMN (e.g. in lateral temporal cortex and angular gyrus).

Cortical gradient analysis

Cortical gradient analysis was performed using The BrainSpace Toolbox (https://brainspace.readthedocs.io/). Due to limitation of computational resources, we chose to use the default Schafer's 400 region parcellation in Brainspace. We performed gradient decomposition on a group-averaged parcel-by-parcel matrix of informational connectivity, rather than extracting the simple Pearson time series correlation between pairs of regions, since informational connectivity corresponds to the similarity in the multivariate response of pairs of brain regions at different points in time (reflecting the multivariate response to trials in this experiment). In this way, informational connectivity is thought to reflect similarity in the representational states of parcels during a task.

We grouped the semantic judgements into strongly associated (top 1/3 trials, with a mean rating strength of 3.51) and weakly associated (bottom 1/3 trials, with a mean rating strength of 1.59) sets, according to the strength of association between the words within each trial (as measured by word2vec, widely used in the field (Kivisaari et al., 2019, Pereira et al., 2016)). Between-trial pattern similarity matrices were produced for each condition. All within-run pairs of trials were excluded from the analysis to reduce the influence of temporal autocorrelation on pattern similarity. Next, the informational connectivity matrix was constructed by correlating the between-trial pattern similarity matrix between pairs of parcels for each condition. Lastly, the BrainSpace Toolbox was applied to the informational connectivity matrix to estimate the principal gradient, with the following parameters: kernel function: cosine similarity; dimensionality reduction: PCA.

For each trial, we also extracted the TRs associated with the chevron task. These judgements were again divided into two sets, based on the strength of the word association that preceded them. This allowed us to conduct a control analysis, examining the semantic manipulation at a time point when it should not affect the gradient decomposition results in a systematic fashion.

To confirm the advantage of the informational connectivity method, a supplementary analysis was conducted using a more standard measure of functional connectivity (i.e., Pearson correlation of averaged time series between any pair of ROIs). Gradient components extracted in this fashion proved to be insensitive to the effects of associative strength. These results are provided in the Supplementary Materials.

Second-order representational similarity analysis

A semantic model capturing the expected similarity in semantic processing across individuals was constructed by measuring the correlation of their ratings of associative strength across trials. For the brain data, we then calculated the between-trial pattern similarity for each participant using a searchlight approach: for each voxel, signals were extracted from a cube region containing 125 surrounding voxels (15 x 15 x 15 mm). A brain model capturing the neural similarity between each pair of participants was constructed for each voxel in MNI space by calculating the correlation of representational similarity patterns at that location. A second-order representational similarity analysis was then conducted by correlating the semantic and brain models to assess brain-behaviour alignment at the group level. All *r* values were Fisher z transformed before carrying them forwards into next-step analyses. To reduce the effect of temporal autocorrelation on pattern similarity, within-run pairs of trials were excluded from the analyses.

Estimating the dimensionality of representational space

Brain regions that show a more diverse pattern of response across their voxels across trials and/or participants are expected to contribute in a more complex or varied fashion to a task (Diedrichsen et al., 2013, Rigotti et al., 2013). To estimate the dimensionality of the representational space supporting the semantic processing of each word-pair in each parcel, we concatenated brain activation maps across participants into a four-dimensional data matrix for each trial, then applied PCA using a searchlight approach as above to measure the voxels' heterogeneity locally. The input matrix contained 31 x 125 features (31 participants by 125 voxels in each 5 x 5 x 5 searchlight cube). PCA is a linear approach that transforms the data to a low-dimensional space represented by a set of orthogonal components that explain maximal variance. We measured the dimensionality of each cube by calculating how many components were needed to explain more than 90 percent of the variance, reflecting the voxels' functional heterogeneity in each local area. To check the robustness of our findings, we also assessed dimensionality by calculating how many components were required to explain more than 60 and 75 percent of the variance. Next, we examined the effect of strength of association on the dimensionality of the neural response (i.e., between strongly and weakly associated trials) in the semantic task, across the whole group of participants. As a control analysis, we also examined dimensionality differences between strongly and weakly associated trials in the chevron task as a function of the associative strength of the semantic trial that preceded it. These results are provided in Supplementary Materials.

Data availability

Neuroimaging data at the group-level are openly available in Neurovault at https://neurovault.org/collections/12539/. Experiment materials and scripts for the task, source data, and analysis codes are accessible in the Open Science Framework at https://osf.io/mkgcy/. The conditions of our ethical approval do not permit public archiving of the data because participants did not provide sufficient consent. Researchers who wish to access the data should contact the Research Ethics and Governance Committee of the York Neuroimaging Centre, University of York, or the corresponding authors. Data will be released to researchers when this is possible under the terms of the GDPR (General Data Protection Regulation).

References

AHLHEIM, C. & LOVE, B. C. 2018. Estimating the functional dimensionality of neural representations. *NeuroImage,* 179**,** 51-62.

ALEXANDER-BLOCH, A. F., SHOU, H., LIU, S., SATTERTHWAITE, T. D., GLAHN, D. C., SHINOHARA, R. T., VANDEKAR, S. N. & RAZNAHAN, A. 2018. On testing for spatial correspondence between maps of human brain structure and function. *NeuroImage,* 178**,** 540-551.

ANDERSSON, J. L., JENKINSON, M. & SMITH, S. 2007. Non-linear registration aka Spatial normalisation FMRIB Technial Report TR07JA2. *FMRIB Analysis Group of the University of Oxford*.

ANZELLOTTI, S. & COUTANCHE, M. N. 2018. Beyond functional connectivity: investigating networks of multivariate representations. *Trends in cognitive sciences,* 22**,** 258-269.

ASSEM, M., GLASSER, M. F., VAN ESSEN, D. C. & DUNCAN, J. 2020. A Domain-General Cognitive Core Defined in Multimodally Parcellated Human Cortex. *Cerebral Cortex*.

BADRE, D., BHANDARI, A., KEGLOVITS, H. & KIKUMOTO, A. 2021. The dimensionality of neural representations for control. *Current Opinion in Behavioral Sciences,* 38**,** 20-28.

BADRE, D. & D'ESPOSITO, M. 2009. Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience,* 10**,** 659-669.

BADRE, D. & NEE, D. E. 2018. Frontal cortex and the hierarchical control of behavior. *Trends in cognitive sciences,* 22**,** 170-188.

BALDASSANO, C., CHEN, J., ZADBOOD, A., PILLOW, J. W., HASSON, U. & NORMAN, K. A. 2017. Discovering event structure in continuous narrative perception and memory. *Neuron,* 95**,** 709-721. e5.

BARSALOU, L. W. 2008. Grounded cognition. *Annu. Rev. Psychol.,* 59**,** 617-645.

BRUNEC, I. K., BELLANA, B., OZUBKO, J. D., MAN, V., ROBIN, J., LIU, Z.-X., GRADY, C., ROSENBAUM, R. S., WINOCUR, G., BARENSE, M. D. & MOSCOVITCH, M. 2018. Multiple Scales of Representation along the Hippocampal Anteroposterior Axis in Humans. *Current Biology,* 28**,** 2129-2135.e6.

CHANALES, A. J. H., OZA, A., FAVILA, S. E. & KUHL, B. A. 2017. Overlap among Spatial Memories Triggers Repulsion of Hippocampal Representations. *Current Biology,* 27**,** 2307-2317.e5.

CORREIA, J., FORMISANO, E., VALENTE, G., HAUSFELD, L., JANSMA, B. & BONTE, M. 2014. Brain-Based Translation: fMRI Decoding of Spoken Words in Bilinguals Reveals Language-Independent Semantic Representations in Anterior Temporal Lobe. *The Journal of Neuroscience,* 34**,** 332-338.

COUTANCHE, M. N. & THOMPSON-SCHILL, S. L. 2013. Informational connectivity: identifying synchronized discriminability of multi-voxel patterns across the brain. *Frontiers in human neuroscience,* 7**,** 15.

COUTANCHE, M. N. & THOMPSON-SCHILL, S. L. 2015. Creating Concepts from Converging Features in Human Cortex. *Cereb Cortex,* 25**,** 2584-93.

DAVEY, J., CORNELISSEN, P. L., THOMPSON, H. E., SONKUSARE, S., HALLAM, G., SMALLWOOD, J. & JEFFERIES, E. 2015. Automatic and controlled semantic retrieval: TMS reveals distinct contributions of posterior middle temporal gyrus and angular gyrus. *Journal of Neuroscience,* 35**,** 15230-15239.

DAVEY, J., THOMPSON, H. E., HALLAM, G., KARAPANAGIOTIDIS, T., MURPHY, C., DE CASO, I., KRIEGER-REDWOOD, K., BERNHARDT, B. C., SMALLWOOD, J. & JEFFERIES, E. 2016. Exploring the role of the posterior middle temporal

gyrus in semantic cognition: Integration of anterior temporal lobe with executive processes. *Neuroimage,* 137**,** 165-177.
DIEDRICHSEN, J., WIESTLER, T. & EJAZ, N. 2013. A multivariate method to determine the dimensionality of neural representation from population activity. *NeuroImage,* 76**,** 225-235.
DONG, H.-M., MARGULIES, D. S., ZUO, X.-N. & HOLMES, A. J. 2021a. Shifting gradients of macroscale cortical organization mark the transition from childhood to adolescence. *Proceedings of the National Academy of Sciences,* 118**,** e2024448118.
DONG, H.-M., MARGULIES, D. S., ZUO, X.-N. & HOLMES, A. J. 2021b. Shifting gradients of macroscale cortical organization mark the transition from childhood to adolescence. *Proceedings of the National Academy of Sciences,* 118.
DUPRE, E., SALO, T., AHMED, Z., BANDETTINI, P. A., BOTTENHORN, K. L., CABALLERO-GAUDES, C., DOWDLE, L. T., GONZALEZ-CASTILLO, J., HEUNIS, S. & KUNDU, P. 2021. TE-dependent analysis of multi-echo fMRI with* tedana. *Journal of Open Source Software,* 6**,** 3669.
ENGE, A., ABDEL RAHMAN, R. & SKEIDE, M. A. 2021. A meta-analysis of fMRI studies of semantic cognition in children. *NeuroImage,* 241**,** 118436.
ERB, J., HENRY, M. J., EISNER, F. & OBLESER, J. 2013. The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *Journal of Neuroscience,* 33**,** 10688-10697.
FEDORENKO, E., DUNCAN, J. & KANWISHER, N. 2013. Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences,* 110**,** 16616-16621.
GAO, Z., ZHENG, L., CHIOU, R., GOUWS, A., KRIEGER-REDWOOD, K., WANG, X., VARGA, D., RALPH, M. A. L., SMALLWOOD, J. & JEFFERIES, E. 2021a. Distinct and common neural coding of semantic and non-semantic control demands. *NeuroImage,* 236**,** 118230.
GAO, Z., ZHENG, L., CHIOU, R., GOUWS, A., KRIEGER-REDWOOD, K., WANG, X., VARGA, D., RALPH, M. A. L., SMALLWOOD, J. & JEFFERIES, E. 2021b. Distinct and common neural coding of semantic and non-semantic control demands. *NeuroImage,* 236**,** 118230-118230.
HONG, S.-J., XU, T., NIKOLAIDIS, A., SMALLWOOD, J., MARGULIES, D. S., BERNHARDT, B., VOGELSTEIN, J. & MILHAM, M. P. 2020. Toward a connectivity gradient-based framework for reproducible biomarker discovery. *NeuroImage,* 223**,** 117322.
HUMPHREYS, G. F. & LAMBON RALPH, M. A. 2015a. Fusion and Fission of Cognitive Functions in the Human Parietal Cortex. *Cereb Cortex,* 25**,** 3547-60.
HUMPHREYS, G. F. & LAMBON RALPH, M. A. 2015b. Fusion and fission of cognitive functions in the human parietal cortex. *Cerebral Cortex,* 25**,** 3547-3560.
HUNTENBURG, J. M., BAZIN, P.-L. & MARGULIES, D. S. 2018. Large-Scale Gradients in Human Cortical Organization. *Trends in Cognitive Sciences,* 22**,** 21-31.
HUTH, A. G., DE HEER, W. A., GRIFFITHS, T. L., THEUNISSEN, F. E. & GALLANT, J. L. 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature,* 532**,** 453-458.
JACKSON, R. L. 2021. The neural correlates of semantic control revisited. *NeuroImage,* 224**,** 117444.
JACKSON, R. L., HOFFMAN, P., POBRIC, G. & RALPH, M. A. L. 2016. The semantic network at work and rest: differential connectivity of anterior temporal lobe subregions. *Journal of Neuroscience,* 36**,** 1490-1501.
JEFFERIES, E. 2013. The neural basis of semantic cognition: converging evidence from neuropsychology, neuroimaging and TMS. *Cortex,* 49**,** 611-625.
JEFFERIES, E., THOMPSON, H., CORNELISSEN, P. & SMALLWOOD, J. 2020a. The neurocognitive basis of knowledge about object identity and events: dissociations reflect opposing effects of semantic coherence and control. *Philosophical Transactions of the Royal Society B,* 375**,** 20190300.
JEFFERIES, E., THOMPSON, H., CORNELISSEN, P. & SMALLWOOD, J. 2020b. The neurocognitive basis of knowledge about object identity and events: dissociations reflect opposing effects of semantic coherence and control. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 375**,** 20190300.
JENKINSON, M. & SMITH, S. 2001. A global optimisation method for robust affine registration of brain images. *Medical image analysis,* 5**,** 143-156.
JU, H. & BASSETT, D. S. 2020. Dynamic representations in networked neural systems. *Nature Neuroscience,* 23**,** 908-917.
KIEFER, M. & PULVERMÜLLER, F. 2012. Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *cortex,* 48**,** 805-825.
KIVISAARI, S. L., VAN VLIET, M., HULTÉN, A., LINDH-KNUUTILA, T., FAISAL, A. & SALMELIN, R. 2019. Reconstructing meaning from bits of information. *Nature Communications,* 10**,** 927.
KRIEGER-REDWOOD, K., STEWARD, A., GAO, Z., WANG, X., HALAI, A., SMALLWOOD, J. & JEFFERIES, E. 2022. Creativity in Verbal Associations is Linked to Semantic Control. *bioRxiv***,** 2022.02.08.479385.
KUNDU, P., BRENOWITZ, N. D., VOON, V., WORBE, Y., VÉRTES, P. E., INATI, S. J., SAAD, Z. S., BANDETTINI, P. A. & BULLMORE, E. T. 2013. Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proceedings of the National Academy of Sciences,* 110**,** 16187-16192.
KUNDU, P., INATI, S. J., EVANS, J. W., LUH, W.-M. & BANDETTINI, P. A. 2012. Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *NeuroImage,* 60**,** 1759-1770.
LAMBON RALPH, M. A., JEFFERIES, E., PATTERSON, K. & ROGERS, T. T. 2017. The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience,* 18**,** 42.

LANZONI, L., RAVASIO, D., THOMPSON, H., VATANSEVER, D., MARGULIES, D., SMALLWOOD, J. & JEFFERIES, E. 2020. The role of default mode network in semantic cue integration. *NeuroImage,* 219**,** 117019.

MACK, M. L., PRESTON, A. R. & LOVE, B. C. 2020. Ventromedial prefrontal cortex compression during concept learning. *Nature Communications,* 11**,** 46.

MARGULIES, D. S., GHOSH, S. S., GOULAS, A., FALKIEWICZ, M., HUNTENBURG, J. M., LANGS, G., BEZGIN, G., EICKHOFF, S. B., CASTELLANOS, F. X., PETRIDES, M., JEFFERIES, E. & SMALLWOOD, J. 2016. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences,* 113**,** 12574-12579.

METEYARD, L., CUADRADO, S. R., BAHRAMI, B. & VIGLIOCCO, G. 2012. Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex,* 48**,** 788-804.

MIRMAN, D., LANDRIGAN, J.-F. & BRITT, A. E. 2017. Taxonomic and thematic semantic systems. *Psychological bulletin,* 143**,** 499.

MKRTYCHIAN, N., BLAGOVECHTCHENSKI, E., KURMAKAEVA, D., GNEDYKH, D., KOSTROMINA, S. & SHTYROV, Y. 2019. Concrete vs. Abstract Semantics: From Mental Representations to Functional Brain Mapping. *Frontiers in Human Neuroscience,* 13.

MONTEFINESE, M. 2019. Semantic representation of abstract and concrete words: A minireview of neural evidence. *Journal of neurophysiology,* 121**,** 1585-1587.

MURPHY, C., JEFFERIES, E., RUESCHEMEYER, S.-A., SORMAZ, M., WANG, H.-T., MARGULIES, D. S. & SMALLWOOD, J. 2018. Distant from input: Evidence of regions within the default mode network supporting perceptually-decoupled and conceptually-guided cognition. *NeuroImage,* 171**,** 393-401.

MURPHY, C., RUESCHEMEYER, S.-A., SMALLWOOD, J. & JEFFERIES, E. 2019. Imagining sounds and images: Decoding the contribution of unimodal and transmodal brain regions to semantic retrieval in the absence of meaningful input. *Journal of cognitive neuroscience,* 31**,** 1599-1616.

MURPHY, C., RUESCHEMEYER, S.-A., WATSON, D., KARAPANAGIOTIDIS, T., SMALLWOOD, J. & JEFFERIES, E. 2017. Fractionating the anterior temporal lobe: MVPA reveals differential responses to input and conceptual modality. *NeuroImage,* 147**,** 19-31.

NOONAN, K. A., JEFFERIES, E., VISSER, M. & LAMBON RALPH, M. A. 2013a. Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *J Cogn Neurosci,* 25**,** 1824-50.

NOONAN, K. A., JEFFERIES, E., VISSER, M. & LAMBON RALPH, M. A. 2013b. Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *Journal of cognitive neuroscience,* 25**,** 1824-1850.

PATTERSON, K., NESTOR, P. J. & ROGERS, T. T. 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature reviews neuroscience,* 8**,** 976-987.

PEREIRA, F., GERSHMAN, S., RITTER, S. & BOTVINICK, M. 2016. A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. *Cognitive Neuropsychology,* 33**,** 175-190.

POPPENK, J., EVENSMOEN, H. R., MOSCOVITCH, M. & NADEL, L. 2013. Long-axis specialization of the human hippocampus. *Trends in cognitive sciences,* 17**,** 230-240.

RALPH, M. A. L., JEFFERIES, E., PATTERSON, K. & ROGERS, T. T. 2017. The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience,* 18**,** 42.

RIGOTTI, M., BARAK, O., WARDEN, M. R., WANG, X.-J., DAW, N. D., MILLER, E. K. & FUSI, S. 2013. The importance of mixed selectivity in complex cognitive tasks. *Nature,* 497**,** 585-590.

SCHWARTZ, M. F., KIMBERG, D. Y., WALKER, G. M., BRECHER, A., FASEYITAN, O. K., DELL, G. S., MIRMAN, D. & COSLETT, H. B. 2011. Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proceedings of the National Academy of Sciences,* 108**,** 8520-8524.

SEGHIER, M. L., FAGAN, E. & PRICE, C. J. 2010. Functional subdivisions in the left angular gyrus where the semantic system meets and diverges from the default network. *Journal of Neuroscience,* 30**,** 16809-16817.

SMALLWOOD, J., BERNHARDT, B. C., LEECH, R., BZDOK, D., JEFFERIES, E. & MARGULIES, D. S. 2021. The default mode network in cognition: a topographical perspective. *Nature Reviews Neuroscience,* 22**,** 503-513.

TEIGE, C., CORNELISSEN, P. L., MOLLO, G., ALAM, T. R. D. J. G., MCCARTY, K., SMALLWOOD, J. & JEFFERIES, E. 2019. Dissociations in semantic cognition: Oscillatory evidence for opposing effects of semantic control and type of semantic relation in anterior and posterior temporal cortex. *cortex,* 120**,** 308-325.

TEIGE, C., MOLLO, G., MILLMAN, R., SAVILL, N., SMALLWOOD, J., CORNELISSEN, P. L. & JEFFERIES, E. 2018. Dynamic semantic cognition: Characterising coherent and controlled conceptual retrieval through time using magnetoencephalography and chronometric transcranial magnetic stimulation. *Cortex,* 103**,** 329-349.

VISSER, M. & LAMBON RALPH, M. 2011. Differential contributions of bilateral ventral anterior temporal lobe and left anterior superior temporal gyrus to semantic processes. *Journal of cognitive neuroscience,* 23**,** 3121-3131.

WANG, J., CONDER, J. A., BLITZER, D. N. & SHINKAREVA, S. V. 2010. Neural representation of abstract and concrete concepts: A meta-analysis of neuroimaging studies. *Human brain mapping,* 31**,** 1459-1468.

WANG, X., MARGULIES, D. S., SMALLWOOD, J. & JEFFERIES, E. 2020. A gradient from long-term memory to novel cognition: Transitions through default mode and executive cortex. *NeuroImage,* 220**,** 117074.

YEO, B. T. T., KRIENEN, F. M., SEPULCRE, J., SABUNCU, M. R., LASHKARI, D., HOLLINSHEAD, M., ROFFMAN, J. L., SMOLLER, J. W., ZÖLLEI, L., POLIMENI, J. R., FISCHL, B., LIU, H. & BUCKNER, R. L. 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology,* 106**,** 1125-1165.