Similar neural networks respond to coherence during comprehension and production of discourse

MATÍAS MORALES

Tanvi PATEL

Andres TAMM

Martin J. PICKERING

Paul HOFFMAN*

School of Philosophy, Psychology & Language Sciences, University of Edinburgh, UK

* Correspondence to:

Dr. Paul Hoffman

School of Philosophy, Psychology & Language Sciences,

University of Edinburgh,

7 George Square, Edinburgh, EH8 9JZ, UK

Tel: +44 (0) 131 650 4654

Email: p.hoffman@ed.ac.uk

Acknowledgements

PH was supported by a BBSRC grant (BB/T004444/1). We are grateful to Beth Jefferies and Xiuyi Wang for helpful discussions regarding these data. Imaging was carried out at the Edinburgh Imaging Facility (www.ed.ac.uk/edinburgh-imaging), University of Edinburgh, which is part of the SINAPSE collaboration (www.sinapse.ac.uk). We are grateful to the University of Minnesota Center for Magnetic Resonance Research for sharing their neuroimaging sequences.

Abstract

When comprehending discourse, listeners engage default mode regions associated with integrative semantic processing to construct a situation model of its content. We investigated how similar networks are engaged when we produce, as well as comprehend, discourse. During fMRI, participants spoke about a series of specific topics and listened to discourse on other topics. We tested how activation was predicted by natural fluctuations in the global coherence of the discourse, i.e., the degree to which utterances conformed to the expected topic. The neural correlates of coherence were similar across speaking and listening, particularly in default mode regions. This network showed greater activation when less coherent speech was heard or produced, reflecting updating of mental representations when discourse did not conform to the expected topic. In contrast, regions that exert control over semantic activation showed task-specific effects, correlating negatively with coherence during listening but not during production. Participants who showed greater activation in left inferior prefrontal cortex also produced more coherent discourse, suggesting a specific role for this region in goal-directed regulation of speech content. Results suggest strong correspondence of discourse representations during speaking and listening. However, they indicate that the semantic control network plays different roles in comprehension and production.

Introduction

When people comprehend discourse, they construct a mental model of its content (often termed a "situation model") by integrating the incoming information with their prior semantic knowledge (Kintsch & van Dijk, 1978; Zwaan & Radvansky, 1998). These discourse comprehension processes engage regions in the default mode network (DMN), such as medial prefrontal cortex, precuneus, posterior cingulate, anterior temporal and lateral parietal cortex (Ferstl, Neumann, Bogler, & von Cramon, 2008; Yeshurun, Nguyen, & Hasson, 2021). This role in the construction of discourse models is supported by evidence implicating the DMN in semantic processing in general (Binder & Desai, 2011; Binder, Desai, Graves, & Conant, 2009) and in integration of multiple semantic cues in particular (Lanzoni et al., 2020). More broadly, the role of the DMN in the highest levels of language comprehension fits with data indicating that this network occupies the endpoint of a functional gradient in the brain that varies from low-level sensorimotor processes to high-level multimodal abstract thought (Margulies et al., 2016).

The DMN's response to variation in the coherence of language input is a major source of evidence for its role in discourse-level processes. Coherence refers to the degree to which successive utterances relate meaningfully to each other and to the overall topic of the discourse (Gloser & Deser, 1992). Neuroimaging studies have typically manipulated discourse coherence by taking narratives and scrambling the order of the sentences so that they no longer tell a coherent story, or by constructing sets of sentences that are each individually meaningful but have no meaningful relationships between them. When presented with incoherent materials like this, it is impossible to construct a coherent situation model of the text. Accordingly, DMN regions show reduced activation to incoherent relative to coherent texts (Ferstl et al., 2008; Ferstl & von Cramon, 2002; Yarkoni, Speer, & Zacks, 2008) and DMN activation patterns become less strongly correlated across participants when narratives are incoherent (Lerner, Honey, Katkov, & Hasson, 2014; Lerner, Honey, Silbert, & Hasson, 2011; Simony et al., 2016).

Experimental manipulations of coherence have been important in establishing a role for the DMN in models of discourse content. These manipulations are, however, not representative

of natural language. Healthy individuals never produce strings of entirely unrelated sentences when speaking. Instead, low coherence in natural language manifests as a weakening of the connections between utterances and an increased frequency of off-topic or tangential statements (Gloser & Deser, 1992; McNamara, Kintsch, Songer, & Kintsch, 1996). When listeners comprehend low coherence discourse of this type, they attempt to construct a situation model of its content but the process is made difficult by the fact that the model must be frequently updated and adjusted to accommodate shifts in topic and violations of the comprehender's expectations (Kurby & Zacks, 2012; Zacks, Speer, Swallow, Braver, & Reynolds, 2007). Little is known about how the brain responds to natural discourse that is low in coherence but is not entirely incoherent. Some evidence from experimental manipulations of coherence suggests that activation in some parts of the DMN is *higher* for weakly coherent passages, relative to either highly coherent samples or completely incoherent material (Kuperberg, Lakshmanan, Caplan, & Holcomb, 2006; Maguire, Frith, & Morris, 1999). This is consistent with the idea that weakly coherent discourse places high demands on DMN regions that support the construction of mental models. One of the aims of the present study was to test whether a similar effect occurs in response to natural fluctuations in coherence in real samples of spoken discourse.

Situation models are thought to govern production as well as comprehension of discourse (Garrod & Pickering, 2004; Kintsch & van Dijk, 1978), but little is known about how these processes overlap in the brain. Few neuroimaging studies have investigated discourse production, but those that have reveal that generating discourse engages an extensive left-lateralized network including anterior temporal and inferior parietal DMN regions, as well as prefrontal regions that have been associated with planning and cognitive control (AbdulSabur et al., 2014; Awad, Warren, Scott, Turkheimer, & Wise, 2007; Blank, Scott, Murphy, Warburton, & Wise, 2002; Stephens, Silbert, & Hasson, 2010). Importantly, people activate a similar cortical network during the comprehension of discourse, suggesting a common neural basis (AbdulSabur et al., 2014; Awad et al., 2007; Silbert, Honey, Simony, Poeppel, & Hasson, 2014; Stephens et al., 2010; Wise & Geranmayeh, 2016). In particular, Pickering and Garrod (2004) have argued that, during conversation, interlocutors' situation models (together with other linguistic representations) become aligned with each other (i.e., more similar), thus facilitating communication (Garrod & Pickering, 2004; Pickering & Garrod, 2021). Recent

neuroimaging studies have provided evidence for this assertion. Studies measuring inter-subject correlations in neural activity indicate that speakers and listeners show correlated activation patterns in lateral parietal, and medial and lateral prefrontal cortex when speech is coherent (Heidlmayr, Weber, Takashima, & Hagoort, 2020; Silbert et al., 2014). This suggests a coupling of neural activity between individuals that facilitates communication (Silbert et al., 2014). Within individuals, however, it is not known whether producing speech places similar demands on model construction as comprehending it. A second aim of the present study was therefore to assess the degree to which neural correlations with coherence during comprehension are similar to those that occur when participants produce their own discourse. If the DMN plays a similar role in generating models for production as well as comprehension, we would expect to see similar effects.

In addition to representational processes which may be shared with comprehension, production of discourse relies on executive processes that regulate its content (Barker, Young, & Robinson, 2017; Kintz, Fergadiotis, & Wright, 2016). Executive control is necessary to ensure that coherence is maintained during speech production, such that discourse remains focused on the topic at hand and avoids irrelevant statements (Arbuckle & Gold, 1993; Marini & Andreetta, 2016). Supporting this view, performance on cognitive control tasks predicts the coherence of speech in older adults (Gold, Andres, Arbuckle, & Schwartzman, 1988; Kintz et al., 2016; Wright, Koutsoftas, Capilouto, & Fergadiotis, 2014) and deficits in coherence and narrative organization are commonly found in the speech of patients with impaired executive functions (Ash et al., 2006; Coelho, 2002; Hoffman, Cogdell-Brooke, & Thompson, 2020; Marini, Zettin, & Galetto, 2014). Recent evidence has identified a role for a specific aspect of executive function, namely semantic control, in the maintenance of coherence (Hoffman, Loginova, & Russell, 2018a). Semantic control processes regulate access to knowledge so that currently relevant concepts are retrieved and selected for use (Hoffman, McClelland, & Lambon Ralph, 2018b; Jefferies, 2013; Lambon Ralph, Jefferies, Patterson, & Rogers, 2017). Hoffman et al. (2018b) found that more coherent speech in young and older adults was predicted by better performance in a semantic control task that required participants to inhibit irrelevant semantic knowledge. In line with this finding, recent neuroimaging evidence links the activation of semantic control regions with the production of highly coherent discourse. Hoffman (2019) asked older adults to produce discourse on different topics. He

found that activation increases in inferior frontal regions, implicated in semantic control (Badre & Wagner, 2007; Jackson, 2021), predicted the production of more coherent speech. Thus, while many discourse processes appear to be shared across production and comprehension, there is evidence that executive control networks may play specific roles in regulating content during production.

The aim of the present study was to investigate the neural correlates of coherence in naturalistic discourse passages, directly contrasting effects in speech comprehension with production. Previous studies have manipulated coherence using experimenter-designed discourse stimuli (Ferstl & von Cramon, 2002; Kuperberg et al., 2006). Here, we instead used a parametric approach to capitalise on the fluctuations in coherence that occur in natural speech. Participants underwent fMRI while producing and listening to discourse about a range of specified topics. We used previously validated methods from computational linguistics to quantify the global coherence of each passage of speech, i.e. the degree to which it related to the overall discourse topic (Gloser & Deser, 1992; Kintsch & van Dijk, 1978). We then investigated how neural activation correlated with variations in coherence.

As speech at the discourse level relies on widely distributed neural networks, we took a network-level approach to analysis. Traditionally, cortical networks have been identified in a discrete fashion, e.g., by using resting-state connectivity patterns to segregate the cortex into a set of distinct networks (e.g., Yeo et al., 2011). This approach assumes hard boundaries between networks. However, other theories propose that cortical function varies in a more continuous fashion, postulating graded variation in connectivity and function as one moves across the cortical surface (Lambon Ralph et al., 2017; Margulies et al., 2016). On this view, network membership is a graded property. In the present study, we used two approaches, one continuous and one discrete, to independently identify the DMN and other functional networks and to seek converging evidence for their roles.

In the first approach, we investigated effects of coherence along the continuous gradient of cortical organisation described by Margulies et al. (2016). Margulies et al. mapped the organisation of the cortex along a single dimension, such that regions of the brain that shared similar patterns of functional connectivity were located at similar points on the spectrum. At

one end of this spectrum lie sensorimotor cortices, which are strongly functionally connected with one another. At the other end lie DMN-associated regions, which also show correlated patterns of activity but are anti-correlated with sensorimotor systems (see Figure 3B). Margulies et al. argued that the principal gradient reflects the organisation of cognitive processes in the brain, varying from stimulus-driven sensorimotor processes at one extreme to multimodal, internally generated thought at the other. We investigated the degree to which the brain's response to coherence in discourse aligned with this gradient of neural organisation. In the second approach, we analysed the effect of coherence in three discrete networks of interest: the DMN, the semantic control network (SCN) and a network implicated in domain-general executive control (the multiple demand network or MDN; Fedorenko, Duncan, & Kanwisher, 2013). We tested how each of these networks responded to variations in the coherence of discourse and, in particular, the degree to which these effects were consistent for speaking and listening.

Methods

Participants

25 native English speakers from the University of Edinburgh participated in the study in exchange for payment. Their mean age was 24 (SD = 4.4, range = 18-35 years old) and they were all right-handed based on the Edinburgh Handedness Inventory (Oldfield, 1971). The study was approved by the Psychology Research Ethics Committee of the University of Edinburgh and all participants gave informed consent.

Materials

In the production task, discourse was elicited by 12 prompts that asked about common semantic knowledge on particular topics (e.g., *Describe how you would make a cup of tea or coffee*; see Supplementary Materials for a complete list of prompts used in the production and comprehension tasks). Discourse production was contrasted with a baseline of automatic speech that involved the reciting of the well-known English nursery rhyme, Humpty Dumpty. For the comprehension task, we selected 24 samples of speech discussing 12 different topics. These were selected from a corpus of responses provided by participants in a previous behavioural study (Hoffman et al., 2018b). Speech comprehension used

different topics to production, to avoid priming participants' production responses with information presented in the comprehension trials. For each topic, we selected one highly coherent and one less coherent response for use in the study, to ensure sufficient variance in coherence across samples (highly coherent: M = 0.61, SD = 0.08; less coherent: M = 0.31, SD = 0.06; global coherence values, described in detail below). In addition, highly coherent and less coherent passages were of similar word length (highly coherent: M = 146, SD = 23.62; less coherent: M = 155, SD = 23.03; t = 1.33, p = 0.21). All comprehension speech passages were recorded by the same male native English speaker and their duration was 50 s each. For an example passage, see Supplementary Materials. The Humpty Dumpty rhyme was also pre-recorded as a comprehension baseline lasting 10 s.

Design and procedure

Each participant performed two production and two comprehension runs, presented in an alternating sequence. Order of presentation was counterbalanced over participants, i.e., half of them began the experiment with a production run, whereas the other half began with a comprehension run. Each run was approximately 8 min and included six topics and five baseline trials, the order of which was fully randomised within each run. For both tasks, we created two sets of stimuli whose topics were matched for difficulty based on ratings obtained with a normative pre-test. In the pre-test, participants (N = 15), who were not part of the fMRI study, rated how difficult it would be to talk about each topic from 1 (easy) to 7 (difficult). The final sets of topics did not vary in difficulty rating (Production: M = 2.8, SD = 0.5; Comprehension: M = 2.4, SD = 0.76; t= 1.65, p = 0.11). In the comprehension task, half of the discourse passages were high coherent, while the other half were low coherent, counterbalanced over participants. This manipulation ensured that each participant was exposed to passages with a similar range of coherence scores, although we modelled coherence as a continuous parameter in our analyses.

The procedure for a single trial is shown in Figure 1A. Production trials started with the presentation of a written prompt on screen for 6 s. Participants were asked to prepare to speak during this period and to start speaking when a green circle replaced the prompt in the centre of screen. They were instructed to speak about the topic for 50 s, after which a red cross would replace the green circle. At this point participants were instructed to wait for the

next prompt to appear on screen. The procedure for speech comprehension was the same except that participants were asked to listen to the speech attentively for 50 s while the green circle was on screen. For the baseline conditions, participants were instructed to recite (production task) or to listen (comprehension task) to the Humpty Dumpty rhyme for a 10 s period (in production they were asked to start again from the beginning if they reached the end of the nursery rhyme before the 10s had elapsed). The baseline conditions therefore involved grammatically well-formed continuous speech, but without the requirement to generate or understand novel utterances. The baseline trials were made shorter than the discourse trials to reduce boredom and to maximise the number of discourse trials in the study. All trials were presented with an interstimulus interval jittered between 3 s and 7 s (M = 5 s).

Before scanning, participants were presented with two training trials to familiarise them with both tasks. They were also informed that they would receive a memory test after scanning to increase their motivation and attention during comprehension runs. In this test, participants answered 12 multiple choice questions, one for each topic presented during speech comprehension (each question had three choices). On average participants responded to 10/12 questions correctly (SD = 1.6; one-sample t-test comparing with chance performance: $t = 31.21$, $p < 0.001$). All participants responded above the chance level of 4/12. They were also asked to rate how well they could hear the speech samples in the scanner from 1 (inaudible) to 7 (perfectly audible), and rated auditability with a mean of 5.5/7 (SD = 1.0).

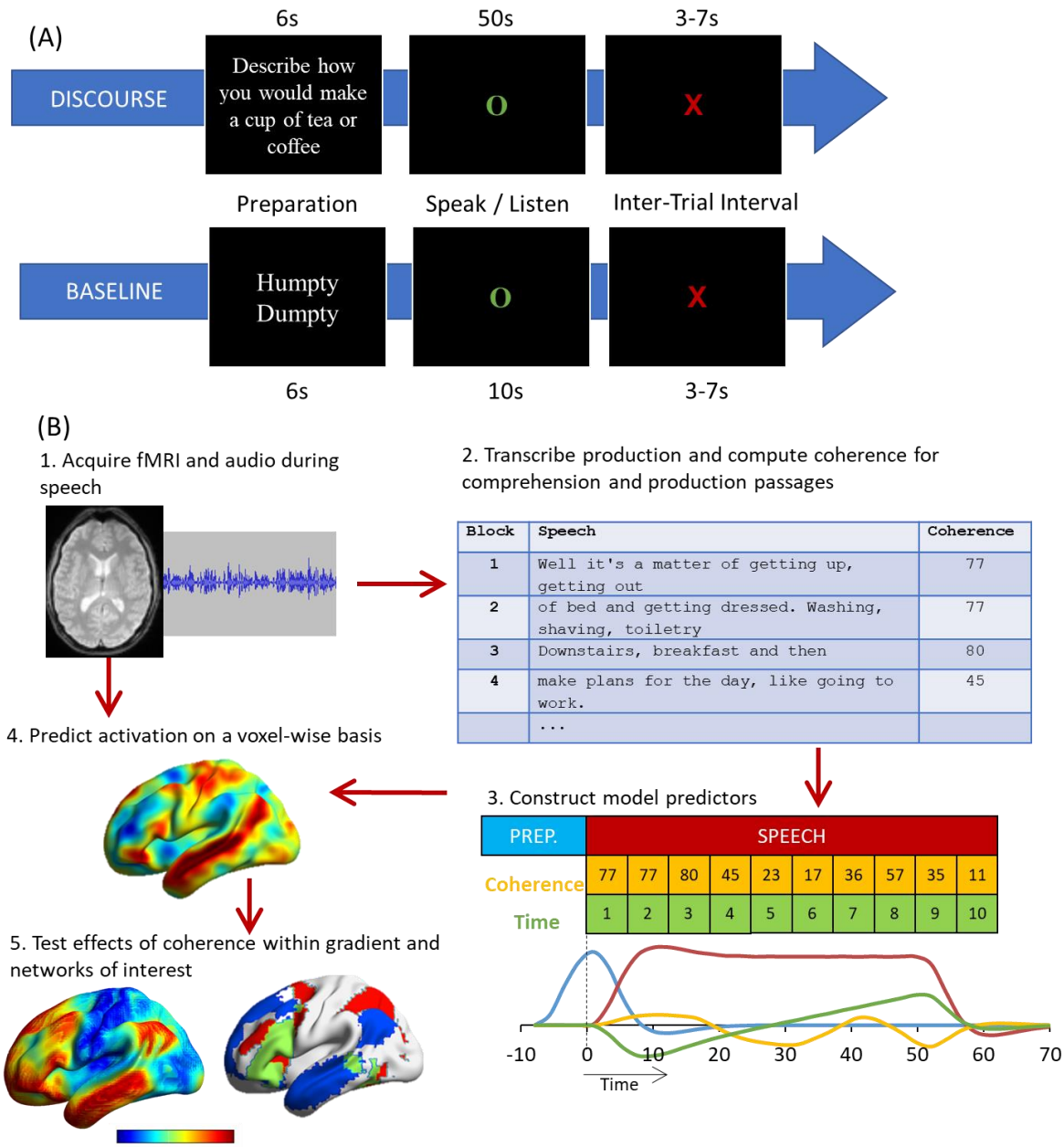

Figure 1. A – Illustration of a single trial of the discourse and baseline tasks. **B** – Stages in analysis.

Processing of speech samples

The stages in analysis are illustrated in Figure 1B. Responses to each prompt were digitally recorded with an MRI-compatible microphone and processed with noise cancellation software (Cusack, Cumming, Bor, Norris, & Lyzenga, 2005) to reduce noise from the scanner. They were then transcribed and the global coherence for each response was computed using computational analyses based on Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997).

In LSA, each word is represented as a high-dimensional vector in which words with similar meanings have similar vectors. Critically, the vectors for individual words can be combined linearly to represent the meaning of speech or text (Foltz, Kintsch, & Landauer, 1998). We used these representations of participants' speech to quantify global coherence, the degree to which the utterances produced by a participant are related to topic they were asked about (Gloser & Deser, 1992). We measured this by calculating the similarity between utterances produced in response to a prompt and the prototypical semantic content expected for that prompt, in the following manner.

Global coherence was measured using the method first described in (Hoffman et al., 2018b). For each production response (target response), we began by computing an LSA representation for each participant's response to the target prompt. Then these representations were averaged (excluding the target response) to obtain a composite vector for the prompt. This composite vector is therefore a semantic representation of a prototypical response to the target prompt, obtained by combining the responses of the whole group. Next, we divided the target response into a series of moving windows of 20 words in length, which each contained the current word and the 19 words preceding it. We computed an LSA vector for each of these windows. Then, these vectors were each compared to the prototypical vector using a cosine similarity metric that outputs a value ranging from 0 to 1. This value indicates how similar the speech produced in each window to the typical semantic content of responses to the same prompt. A high coherence value suggests that speech was strongly semantically related to the prompt being tested, whereas a low coherence value indicates speech that was semantically less related to the topic being probed. This method provided us with a moving estimate of coherence at the point each word was produced. To aggregate these values and relate them to the neural signal, we divided each 50 s speech period into blocks of 5 s and computed the mean global coherence for the words produced in each block (following the procedure in (Hoffman, 2019).

A similar approach was used to quantify global coherence for the speech samples presented in the comprehension task. Here, each speech window was compared with a composite prototype vector generated from the responses given to the prompt by 30 young adult participants in Hoffman et al. (2018).

Image acquisition and processing

Participants were scanned on a 3T Siemens Prisma scanner using a 32-channel head coil. fMRI data were acquired at three echo times (13ms, 31ms, and 48ms) using a whole-brain multiband multi-echo acquisition protocol (Feinberg et al., 2010; Moeller et al., 2010; Xu et al., 2013). Data from these three echo series were weighted and combined, and the resulting time-series denoised using independent components analysis (ICA). This approach improves the signal quality in regions that typically suffer from susceptibility artefacts (e.g., the ventral anterior temporal lobes, (Kundu et al., 2017). The TR was 1.7 s and images consisted of 46 slices with an 80 x 80 matrix and isotropic voxel size of 3mm. Multiband acceleration with a factor of 2 was used and the flip angle was 73°. Four runs of 281 volumes (477.7s) were acquired. A high-resolution T1-weighted structural image was also acquired for each participant using an MP-RAGE sequence with 1mm isotropic voxels, TR = 2.5 s, TE = 4.6 ms.

Images were pre-processed and analysed using SPM12 and the TE-Dependent Analysis Toolbox 0.0.7 (Tedana) (DuPre et al., 2019). Estimates of head motion were obtained using the first BOLD echo series. Slice-timing correction was carried out and images were then realigned using the previously obtained motion estimates. Tedana was used to combine the three echo series into a single-time series and to divide the data into components classified as either BOLD-signal or noise-related based on their patterns of signal decay over increasing TEs (Kundu et al., 2017). Components classified as noise were discarded. After that, images were unwarped with a B0 fieldmap to correct for irregularities in the scanner's magnetic field. Finally, functional images were spatially normalised to MNI space using SPM's DARTEL tool (Ashburner, 2007).

Images were smoothed with a kernel of 8mm FWHM. Data were treated with a high-pass filter with a cut-off of 128 s and the four experimental runs were analysed using a single general linear model. Four speech periods were modelled as different event types: discourse production, baseline production, discourse comprehension, and baseline comprehension. Discourse periods were modelled as a series of concatenated 5 s blocks. This allowed us to include parametric modulators that coded the characteristics of speech: coherence of the

speech, as computed in the earlier description, and time within the 50 s speech period. Time was included as a modulator because it is correlated with coherence: the later stages of a speech sample tend to be less coherent than the earlier ones (Hoffman, 2019). Modulators were mean-centred for each run. Additional regressors modelled the preparation periods for discourse and baseline in each task (6s events). Covariates consisted of six motion parameters and their first-order derivatives.

Analyses

We first analysed the whole-brain activation for speech production and speech comprehension using the contrast discourse minus baseline in each task. For these analyses we employed a voxel-height threshold of $p < 0.005$ for one-sample t-tests with correction for multiple comparisons at the cluster level ($p < 0.05$ corrected), using SPM's random field theory.

To investigate the effect of global coherence on activation during production and comprehension, we correlated effects with the principal connectivity gradient described in Margulies et al. (2016). This gradient provides a macroscale framework of functional and spatial cortical organisation, characterising a spectrum of activation that ranges from unimodal/sensory to multimodal/association areas associated with the DMN. Based on this connectivity gradient, we conducted two analyses using similar methods to Wang et al. (2020). First, we examined the effect of coherence on speech processing at different points on the gradient at the voxel level. We extracted the coherence effect (beta values) for production and comprehension in each voxel and computed voxelwise correlations between these values and the position of each voxel on the Margulies gradient. These correlations were performed in the whole brain and also within a mask of regions associated with semantic processing (defined as regions associated with the term "semantic" in the Neurosynth database; (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011). Second, to determine the reliability of these relationships across participants, we used linear mixed models to test how coherence varied along the cortical gradient. To do this, we divided the voxels along the gradient into 10 equally-sized bins (from bin 1 at the unimodal end to bin 10 at the multimodal end) and extracted a mean coherence effect for each participant for the voxels contained in each bin. We then conducted linear mixed models for production and

comprehension to examine whether there was a linear (or higher order) relationship between the coherence effect and the connectivity gradient.

The gradient analysis tested how neural responses to discourse varied along a single, continuous dimension of neural organisation with the DMN as one of its endpoints. However, we were also interested in the effects on coherence on regions associated with semantic control and general executive function that fall at midpoints along the cortical gradient. To isolate effects in these regions more precisely, we extracted the mean effects of coherence from the following networks of interest (see Figure 5A):

1. The DMN, identified from Yeo et al's (2011) 7-network parcellation of resting-state fMRI in 1000 participants.
2. The semantic control network (SCN), defined as significant voxels in Jackson's (2021) meta-analysis. This meta-analysis identified regions that consistently show effects of cognitive control demands in semantic processing tasks.
3. The multiple-demand network (MDN), defined as regions responding to cognitive demands in multiple domains in Fedorenko et al. (2013). The MDN encompasses prefrontal and parietal and other regions (Duncan, 2010) and frequently acts in functional opposition to the DMN: while the DMN relates to more internally-oriented processes like mind-wandering and remembering the past, the MDN is present in more goal-oriented tasks (Mineroff, Blank, Mahowald, & Fedorenko, 2018).

As expected, many of the voxels in the SCN were also part of either the MDN or DMN (principally in lateral prefrontal cortex). To ensure independence between the three networks, voxels that fell within the SCN were excluded from the masks defining the DMN and MDN. Thus, MDN and DMN results reflect the functions of the portions of these networks that are *not* specifically implicated in controlled semantic cognition. A small number of voxels were also shared between DMN and MDN; these were excluded from the DMN mask.

To test effects of coherence, the mean coherence beta values for comprehension and production were extracted for each network for each participant. For each participant, we extracted data only from the top 20% of voxels that showed the largest effect in the relevant discourse-minus-baseline contrast (Tong et al., 2016). In other words, we restricted our

analysis to those voxels within the network that were most strongly engaged by the discourse processing tasks. Coherence effect sizes were entered into a 2 x 3 (task x network) ANOVA.

Finally, we conducted an exploratory analysis of individual differences to determine whether brain activity during speech production differed between highly coherent and less coherent speakers. Following Hoffman (2019), we calculated the mean coherence for each participant over all of their responses and included this value as a covariate in the group-level analysis of the BOLD response. This allowed us to explore whether participants who tended to speak coherently about the topics provided showed different patterns of activation to those who were less coherent.

Results

Characteristics of speech

In speech production, participants produced 128 words to each prompt on average (range 120-145) and the mean global coherence was 0.49 across all responses (range 0.43-0.66).

We divided each 50 s speech period into ten segments of 5 s each to examine the variation of global coherence over time in speech production and in speech presented for comprehension. For production, we observed a high negative correlation between global coherence and the position of segments in the 50s discourse period (r = -0.9) (see Supplementary Materials). In other words, the longer participants spoke for, the less related their speech became to the prescribed topic. Coherence and time segments were also strongly correlated in speech presented in comprehension runs (r = -0.93). To avoid a confounding effect of this factor, we included segment position as a covariate in the brain analysis for speech production and comprehension, as in Hoffman (2019) (see Figure 1B). We found that the number of words produced in each segment did not vary over time (r = 0), indicating that participants did not slow their speech rate or run out of things to say during these later periods of speech. These patterns were similar in speech samples used for comprehension. Finally, there was no correlation between number of words in a segment and its coherence, for either production (r = - 0.2, p = 0.58) or comprehension (r = 0.3, p = 0.34).

Activation during speech comprehension and speech production

Figure 2 shows activation for discourse production and comprehension against automatic speech baselines. For production, whole-brain analysis revealed a left-biased activation pattern that included inferior frontal gyrus (pars triangularis and pars opercularis), angular gyrus, frontal orbital cortex, the middle frontal gyrus, the anterior temporal lobe and the length of the superior temporal sulcus. A strikingly similar pattern was found in the comprehension task, but with more activation in left posterior temporal regions and additional activation in the right anterior temporal lobe. This is consistent with the semantic-related network found in other studies relating to discourse processing (e.g., (Hoffman, 2019; Silbert et al., 2014) and indicates a high degree of overlap in the networks recruited for high-level discourse comprehension and production.

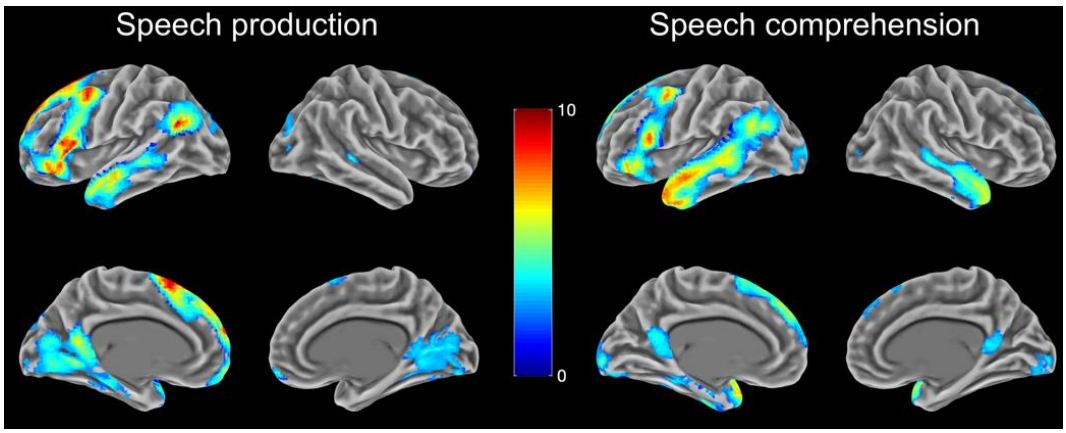

Figure 2. Neural activation during speech comprehension and speech production over automatic speech baseline. Images thresholded at cluster-corrected $p < 0.05$.

Activation during speech comprehension and speech production as a function of coherence

Coherence effects across the whole brain are shown in Figure 3A. Activation in blue areas was positively correlated with coherence, increasing when participants produced or comprehended more coherent speech, whereas red areas were negatively correlated with coherence, showing greater activation when speech was less coherent. For speech production, positive effects of coherence were found principally within the supramarginal gyrus, posterior inferior and middle temporal gyri and the precentral gyrus. In contrast,

negative effects of coherence were found predominantly in DMN regions such as the angular gyrus, anterior middle and superior temporal gyri, the posterior cingulate and ventromedial prefrontal cortex. A similar set of regions showed negative effects of coherence in speech comprehension, with a particularly strong effect in the lateral temporal cortices.

Figure 3B shows the principal functional gradient reported by Marguiles et al. (2016), which arranges regions along a continuum from primary sensory and motor regions to the multimodal cortex of the DMN. Visual inspection suggests alignment between this gradient and the effects of coherence: areas towards the DMN end of the cortical gradient appear to show the strongest negative response to coherence. We assessed this pattern formally in the next section.

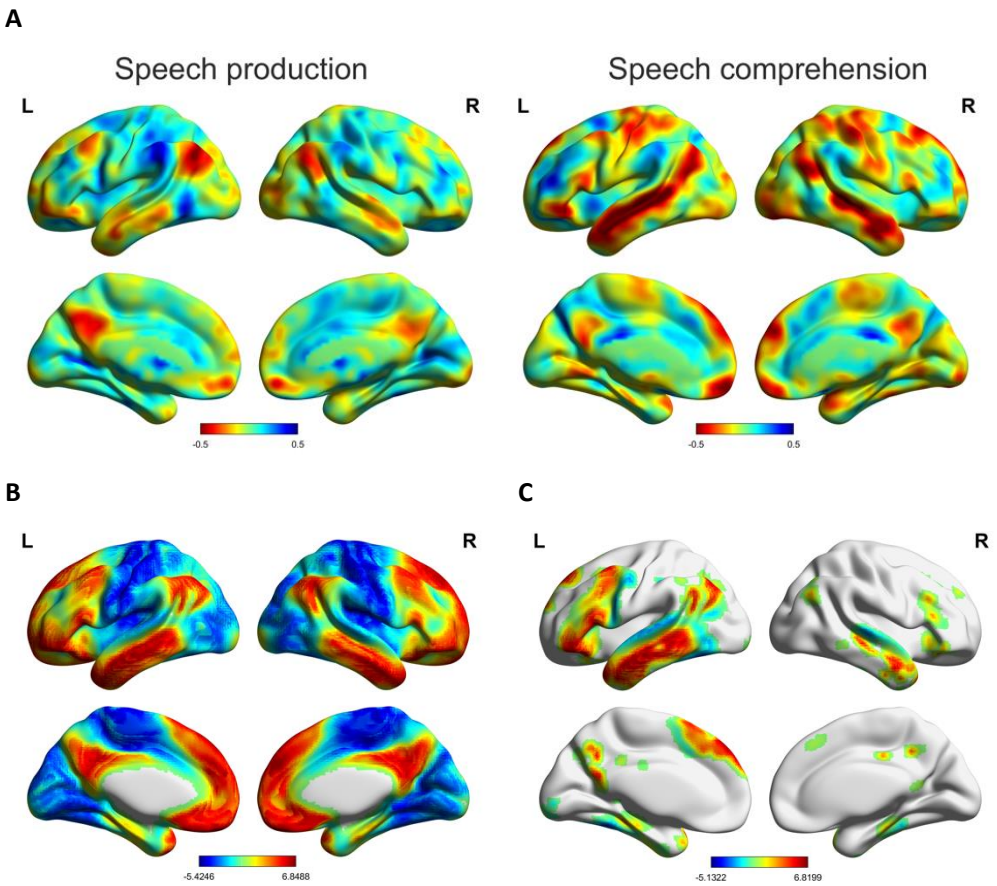

Figure 3. A - Unthresholded maps of the coherence effect (beta values) on speech production and speech comprehension. **B** - Principal connectivity gradient map from Margulies et al. (2016), with the unimodal extreme of the gradient shown in blue and the DMN extreme in red. **C** – Areas of the Marguiles et al. gradient falling within the mask of semantic processing regions.

We analysed our data in two different ways. First, we correlated coherence effects with the connectivity gradient at the voxel level (Wang, Margulies, Smallwood, & Jefferies, 2020). Correlation tests were conducted between the cortical gradient values and group-mean beta values of the coherence effect in production and comprehension (Figure 4, panels A-B, D-E). For production, we found a negative correlation between the coherence effect and the principal gradient ($r = -0.4$, $p < 0.001$). The correlation was stronger when the analysis was restricted to semantic regions only ($r = -0.5$, $p < 0.001$; voxels shown in Figure 3C). The closer voxels were to the multimodal, DMN end of the gradient, the more likely they were to show a strong negative correlation with coherence. For comprehension, activation as a function of coherence was weakly correlated with the gradient at the whole-brain level ($r = -0.09$, $p < 0.001$), but more strongly correlated when analysing only voxels within semantic areas ($r = -0.35$, $p < 0.001$). However, Figure 4D also suggests the possibility of a non-linear relationship between coherence and the gradient at the whole-brain level, in which voxels at both extremes of the gradient show greater activation for less coherent speech.

Next, to account for the possibility of a non-linear relationship between coherence and the gradient, we segmented it into ten bins from unimodal (bin 1) to multimodal DMN regions (bin 10) (Wang et al., 2020). Each voxel was assigned to one of the bins based on its position on the gradient and the coherence effects in the voxels for each bin were averaged for each participant. We then fitted linear mixed models predicting these coherence values for comprehension and production separately (Figure 4, panels C and F). Our fixed effect was the number of the gradient bin (from 1 to 10), treated as a numeric variable and centred. To test for higher order relationships, we included a quadratic term for gradient bin in all models as this predictor improved model fits ($p < 0.05$) over and above the linear term alone (see Supplementary Materials for model comparisons). All models included random intercepts and slopes by participant and p-values were obtained with the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017) via Satterthwaite's method.

We found a significant linear effect for speech production ($\beta = -0.022$, $SE = 0.007$, $t = -2.946$, $p = 0.007$). As shown in Figure 4C, voxels towards the sensorimotor end of the gradient showed greater activation when speech was more coherent, while those towards the DMN

end showed more activation when participants produced less coherent speech. The quadratic term did not reach significance for the production task (β = -0.005, SE = 0.003, t = -1.789, p = 0.086). For speech comprehension, we found no linear relationship between the coherence effect and the gradient (β = -0.002, SE = 0.011, t = -0.220, p = 0.828) but the quadratic term showed a significant effect (β = -0.006, SE = 0.003, t = -2.116, p = 0.045). As shown in Figure 4F, there was a strong negative effect of coherence at the sensorimotor end of the gradient, which gradually decreased at middle parts of the spectrum before increasing at the DMN end. Thus, comprehension and production effects converged in DMN regions, where there was greater activation for less coherent discourse, but were less consistent in sensorimotor regions.

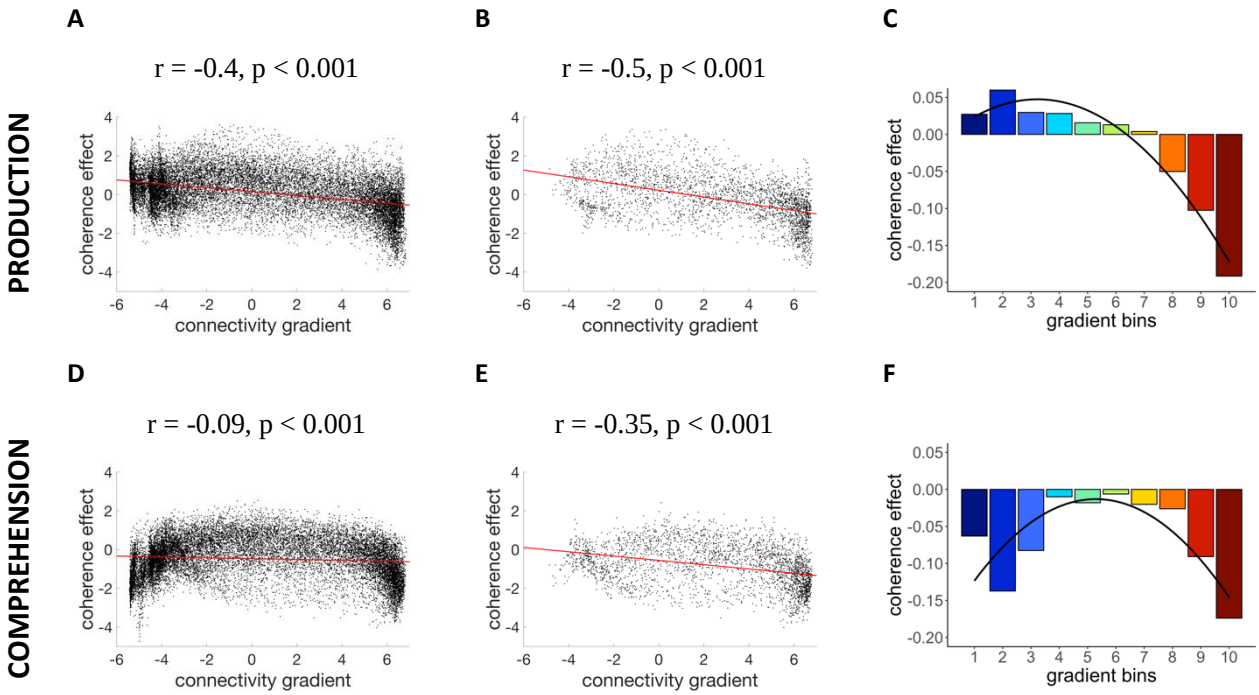

Figure 4. **A** and **B** – Correlations between coherence beta values and the connectivity gradient in production at the whole-brain level and within semantic regions, respectively. **C** – Model fit for the mixed effects model that examined the effect of the gradient on coherence activation during speech production. **D** and **E** – Correlations between the coherence beta values and the connectivity gradient in comprehension at the whole-brain level and within semantic regions. **F** – Model fit for the mixed effects model that tested the effect of the gradient on coherence activation during speech comprehension.

Coherence effects in networks of interest

To confirm and extend the results found for the cortical gradient, we compared the effects of coherence across language tasks in regions belonging to the DMN, SCN and MDN (see Figure 5). We computed a 2 x 3 (task x network) ANOVA, revealing a main effect of network (F (2, 48) = 10.2, $p < 0.001$) and a task x network interaction (F (2, 48) = 3.62, p = 0.034). Post-hoc t-tests with Holm-Bonferroni correction indicated that, during comprehension, DMN and SCN exhibited similar effects of coherence (p = 0.38) which were more negative than those seen in the MDN ($p < 0.05$). In production, however, effects in DMN and SCN differed from one another (p = 0.04). MDN differed from DMN (p = 0.005) but not SCN (p = 0.64). Thus, the DMN showed activation increases during low coherence speech that were consistent across tasks, while less coherent discourse produced greater activation in SCN regions only in comprehension. There were no effects of coherence in MDN regions that were not specifically associated with control over semantic processing.

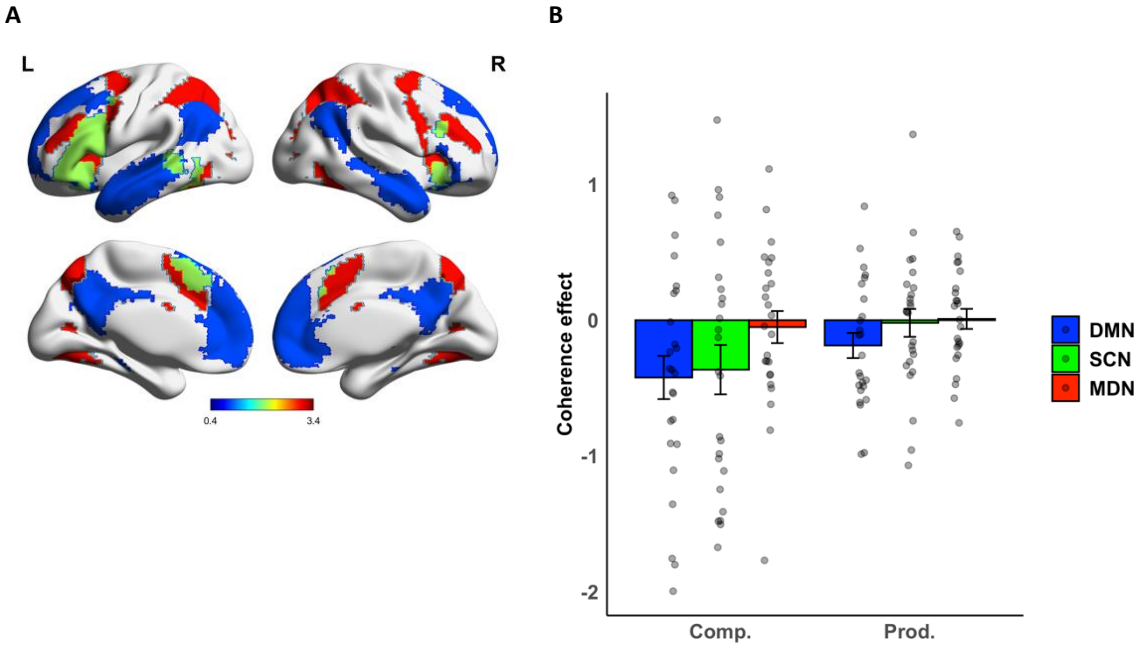

Figure 5. A – Masks used for the default mode network (DMN - blue), semantic control network (SCN - green), and multiple demand network (MDN - red). **B** – Effects of coherence in each network.

Effects of individual differences in coherence on activation during speech production

The prior within-subject analyses all investigated how activation covaried with fluctuations in the coherence of speech over time. To explore how activation differed between more and less coherent speakers, we included the mean coherence of the speech produced by each participant as a predictor of activation. Activation during speech production was positively correlated with the coherence of the speaker in a cluster of voxels in the pars triangularis portion of the left inferior frontal gyrus (17 voxels at uncorrected $p < 0.001$ with a peak at [-54, 36, 12], $t = 5.08$). Participants who produced more coherent speech showed greater activation in this area while speaking. This effect did not survive correction for multiple comparisons across the whole brain; however, it is consistent with previously-found associations of this area with coherence (Alyahya, Lambon Ralph, Halai, & Hoffman, submitted for publication; Hoffman, 2019) and with the theory that the semantic selection processes served by the inferior frontal gyrus are important for maintaining coherence (Hoffman et al., 2018). No other regions showed a positive association with coherence at $p < 0.001$. Less coherent speakers showed more activation in the cerebellum (50 voxels at uncorrected $p < 0.001$ with a peak at [-3, -39,-15], $t = 4.37$).

Discussion

In this fMRI study, we investigated how neural activity covaried with the coherence of discourse during speech production and comprehension. We used a global measure of coherence which tracked the degree to which speech conformed to the expected discourse topic (Gloser & Deser, 1992; Hoffman et al., 2018b). DMN regions showed activation increases when discourse was lower in coherence, whether participants were listening to another speaker or producing their own narratives. The brain's response to coherence was closely aligned with a cortical connectivity gradient that arranges brain regions on a continuum from sensorimotor processing to multimodal abstract thought (Margulies et al., 2016). The multimodal end of this gradient, most associated with the DMN and with integrative higher-order cognition, showed the strongest increases for lower coherence in speech. Coherence effects in comprehension and production diverged in the semantic control network. Regions involved in semantic control showed negative coherence correlations in comprehension that did not occur during production. Our results have implications for understanding the shared

processes that support the representation of discourse during speaking and listening, and the points at which these processes diverge.

When listening, participants showed greater DMN activation when discourse was less coherent. This result is consistent with the hypothesized role of this network in constructing situation models that represent the elements of discourse, their meanings and their relations with one another (Baldassano et al., 2017; Heidlmayr et al., 2020; Lerner et al., 2011). We begin by noting that, unlike many previous studies, we did not use speech passages which were entirely incoherent (e.g., sentences in a scrambled order). Under these conditions, DMN activation declines (Ferstl et al., 2008; Ferstl & von Cramon, 2002; Yarkoni et al., 2008), presumably because attempts to build a model of the discourse are abandoned. In contrast, we assume that when people listened to the speech in our study, they always attempted to form a situation model by integrating the incoming speech input with their prior knowledge and expectations of the topic under discussion. When the speech input was less coherent, we assume that this mental representation required frequent updating and reconfiguration, resulting in increased DMN activation. This interpretation is supported by evidence that DMN activity increases when participants encounter event boundaries in narratives and have to update their mental representations (Speer, Zacks, & Reynolds, 2007; Whitney et al., 2009). Our results are consistent with previous evidence showing that DMN areas are involved in constructing coherent mental models of discourse (Ferstl et al., 2008; Ferstl & von Cramon, 2002; Heidlmayr et al., 2020; Jacoby & Fedorenko, 2020; Kuperberg et al., 2006; Yeshurun et al., 2021). Uniquely, however, our study reveals that similar effects are present when people produce their own discourse. This suggests that the people experience similar difficulty in developing less coherent situation models to use as the basis for their own utterances as when they are comprehending other people's utterances.

Overall, we found that listening and speaking both activated a left-dominant network of frontal, temporal, and inferior parietal regions, consistent with shared activation patterns across comprehension and production in previous studies (e.g., e.g., Silbert et al., 2014). Beyond mere coactivation, however, the present study reveals that much of this network responds to the content of speech in similar ways across comprehension and production. Specifically, participants activated DMN regions more when their own discourse was less

related to the topic of discussion. This negative effect could partly be a consequence of participants' self-monitoring of their own utterances. If participants monitored and updated their situation models on the basis of their speech output, this would result in greater DMN activation during passages of less coherent speech production, just as it does during comprehension. It is likely, however, that situation models also play a much earlier role in planning and generating speech content. Studies of discourse comprehension suggest that increases in DMN activation are indicative of major shifts in the content of situation models (Speer et al., 2007; Whitney et al., 2009). When such shifts occur during production, the likely consequence is speech that lacks coherence and fails to conform to the prescribed topic.

Overall, then, our results suggest that the mechanisms involved in constructing situation models are largely shared between production and comprehension. Previous studies have provided support for this idea by comparing the brain activation of the person producing an utterance with that of the comprehender (Heidlmayr et al., 2020; Silbert et al., 2014; Stephens et al., 2010). In contrast, here we have shown that convergence occurs within individuals when comparing their activation profile as a speaker and as a listener. Our results can be explained by models of language processing which assume that semantic representations are shared between comprehension and production (Gambi & Pickering, 2017; Hagoort, 2013; Kintsch & van Dijk, 1978; Levelt, Roelofs, & Meyer, 1999). This shared representation may allow the production system to make forward predictions that aid comprehension (Dell & Chang, 2014; Pickering & Garrod, 2007) and help interlocutors to align their situation models during conversation (Garrod & Pickering, 2004).

Our interpretation of our results has focused on the idea that the DMN supports situation models of discourse content. This notion is highly compatible with other theories implicating this network in representing event characteristics during episodic recall (Ranganath & Ritchey, 2012) and with the notion that inferior parietal cortex in particular functions as a multimodal buffer of recent experience (Humphreys & Lambon Ralph, 2015). These theories share the view that the DMN generates mental models of events and situations, whether these draw from current experience, recall from memory or from processing discourse. Thus the role of the DMN is not limited to verbal processing; indeed, DMN responses are influenced

by manipulations of coherence when people watch movies as well as when they listen to spoken stories (Lerner et al., 2011).

Another set of researchers have investigated the involvement of DMN in mind-wandering and internally generated thoughts (Andrews-Hanna, Smallwood, & Spreng, 2014; Christoff, Gordon, Smallwood, Smith, & Schooler, 2009). These studies are consistent with the view that the DMN represents mental models of events and situations, as these are a common constituent of self-generated thought (Andrews-Hanna et al., 2014; Binder et al., 1999). Increases in DMN activity have also been associated with task-unrelated thoughts and attention lapses, potentially reflecting unwanted engagement of self-generated thoughts (Christoff et al., 2009; Mason et al., 2007). In our study, the DMN appeared to be making an active contribution to task performance as large parts of the network were *positively* engaged by the discourse tasks (see Figure 2). Is it possible, nevertheless, that the effects of coherence we have observed reflect the intrusion of self-generated thoughts that are not relevant to the task at hand? During comprehension, this account would imply that participants experienced more task-unrelated thoughts during less coherent passages of speech. Although we cannot rule out this possibility, we consider it unlikely since making sense of less coherent speech requires greater attentional focus, not less. During the production task, participants generated their own speech and here it is likely that irrelevant thoughts would lead to the production of a weakly coherent response.

Emerging evidence suggests that cognitive control networks are critical for task-based regulation of DMN activity (Andrews-Hanna et al., 2014). With respect to discourse production in particular, previous studies have found that participants with better executive and semantic control abilities produce more coherent speech (Hoffman et al., 2018b; Wright et al., 2014). In the present study, evidence for this executive regulation is mixed. In an analysis of individual differences, we found that the more coherent speakers in our study activated a left inferior prefrontal region (BA45) more than those who were less coherent. This region is the dominant node in a network of semantic control regions and is specifically implicated in inhibition of irrelevant semantic knowledge (Badre & Wagner, 2007; Jefferies, 2013). This finding supports the idea that semantic control regions regulate the selection of ideas during speech production. When we looked at neural correlations within participants,

we found the SCN's association with coherence was task-dependent. During comprehension, less coherent speech activated the SCN more, in line with the idea that less predictable speech input requires greater executive control. In production, this negative effect was not present. However, we did not find the positive association with coherence that the executive regulation account would predict, and which we found in our previous study in older adults (Hoffman, 2019). Thus, there is mixed evidence as to whether greater SCN engagement leads to the production of more coherent discourse and more research is needed to investigate how and when such regulation occurs.

Beyond the SCN, we found no effects of discourse coherence in the MDN that supports domain-general cognitive control. Some researchers have proposed a role for domain-general MDN regions in regulating speech production (Wise & Geranmayeh, 2016). Here, however, we found these regions were not modulated by the coherence of speech when either speaking or listening. These results suggest that involvement in discourse processing is restricted to those parts of the MDN that are specifically implicated in control over semantic processing (which, in our analysis, were assigned to the SCN instead). This aligns with the general view that cognitive control at the higher levels of language processing draws on specific neural resources distinct from those engaged in other cognitive domains (Fedorenko, Behr, & Kanwisher, 2011; Mineroff et al., 2018).

Our results add to an emerging body of literature examining how effects of semantic processing vary along a cortical gradient from sensorimotor regions to the DMN. Other studies have found that when processing word pairs, pairs with a stronger semantic relationship cause greater activation in areas towards the DMN (Gao et al., 2021; Wang et al., 2020). There are many differences between word-pair semantic judgements and the more natural form of comprehension studied here. However, one potential way of integrating these findings is to assume that semantically related word pairs bring to mind a mental model of particular situation or context, which is supported by the DMN.

One major contribution of the present study is to extend findings relating to comprehension into the domain of discourse production. Although we found convergence in cortical regions towards the DMN end of the cortical gradient, coherence effects in speaking and listening

appeared to dissociate at the sensorimotor extreme of the principal cortical gradient. These regions showed more activation to less coherent discourse in comprehension but not in production. We suggest that these effects may relate to lower levels of speech perception. It is well-known that when perceiving speech in noisy conditions, such as those experienced in a MRI scanner, processing is easier when speech content is highly semantically coherent (Boothroyd & Nittrouer, 1988). Accordingly, activation in auditory cortices is increased when people hear less semantically predictable words during narrative comprehension (see also see also Friederici, Rüschemeyer, Hahne, & Fiebach, 2003; Willems, Frank, Nijhof, Hagoort, & van den Bosch, 2016). Thus, the enhanced response of sensorimotor regions to less coherent passages may reflect the greater demands that these place on speech perception systems.

In conclusion, we found increased DMN activation when participants produced and listened to less coherent discourse. This result suggests that people update and reconfigure representations of discourse content in similar ways during production and comprehension. However, our results also revealed activation increases in semantic control regions during the comprehension of less coherent discourse, which were not present during production. This indicates that the relationship between coherence and executive control systems is task-dependent. Thus, our findings contribute to understanding of the ways neural processes are shared between the production and comprehension of discourse, but also identify the points at which they diverge.

Data and code availability

Analysis code and study data are available at: https://osf.io/nwbk3/

References.

AbdulSabur, N. Y., Xu, Y., Liu, S., Chow, H. M., Baxter, M., Carson, J., & Braun, A. R. (2014). Neural correlates and network connectivity underlying narrative production and comprehension: a combined fMRI and PET study. *Cortex, 57*, 107-127. doi:10.1016/j.cortex.2014.01.017

Alyahya, R. S. W., Lambon Ralph, M. A., Halai, A. D., & Hoffman, P. (submitted for publication). The cognitive and neural underpinnings of discourse coherence in post-stroke aphasia.

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Ann N Y Acad Sci, 1316*, 29-52. doi:10.1111/nyas.12360

Arbuckle, T. Y., & Gold, D. P. (1993). Aging, Inhibition, and Verbosity. *Journal of Gerontology: PSYCHOLOGICAL SCIENCES, 48*(5), 225-232.

Ash, S., Moore, P., Antani, S., McCawly, G., Work, M., & Grossman, M. (2006). Trying to tell a tale: discourse impairments in progressive aphasia and frontotemporal dementia. *Neurology, 66*, 1405-1413.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage, 38*(1), 95-113. doi:DOI 10.1016/j.neuroimage.2007.07.007

Awad, M., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. (2007). A common system for the comprehension and production of narrative speech. *Journal of Neuroscience, 27*(43), 11455-11464. doi:10.1523/JNEUROSCI.5257-06.2007

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia, 45*(13), 2883-2901.

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*(3), 709-721 e705. doi:10.1016/j.neuron.2017.06.041

Barker, M. S., Young, B., & Robinson, G. A. (2017). Cohesive and coherent connected speech deficits in mild stroke. *Brain Lang, 168*, 23-36. doi:10.1016/j.bandl.2017.01.004

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends Cogn Sci, 15*(11), 527-536. doi:10.1016/j.tics.2011.10.001

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex, 19*(12), 2767-2796. doi:10.1093/cercor/bhp055

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Rao, S. M., & Cox, R. W. (1999). Conceptual processing during the conscious resting state: A functional MRI study. *Journal of Cognitive Neuroscience, 11*(1), 80-93.

Blank, S. C., Scott, S. K., Murphy, K., Warburton, E., & Wise, R. J. (2002). Speech production: Wernicke, Broca and beyond. *Brain, 125*, 1829-1838.

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *The Journal of the Acoustical Society of America, 84*, 101-114.

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proc Natl Acad Sci U S A, 106*(21), 8719-8724. doi:10.1073/pnas.0900234106

Coelho, C. A. (2002). Story Narratives of Adults With Closed Head Injury and Non- Brain-Injured Adults: Influence of Socioeconomic Status, Elicitation Task, and Executive Functioning. *Journal of Speech, Language, and Hearing Research, 45*, 1232-1248.

Cusack, R., Cumming, N., Bor, D., Norris, D., & Lyzenga, J. (2005). Automated post-hoc noise cancellation tool for audio recordings acquired in an MRI scanner. *Hum Brain Mapp, 24*(4), 299-304. doi:10.1002/hbm.20085

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634), 20120394.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn Sci, 14*(4), 172-179. doi:10.1016/j.tics.2010.01.004

DuPre, E., Salo, T., Markello, R., Kundu, P., Whitaker, K., & Handwerker, D. (2019). ME-ICA/tedana: 0.07 (Version 0.0.7): Zenodo. Retrieved from http://doi.org/10.5281/zenodo.2558498

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proc Natl Acad Sci U S A, 108*(39), 16428-16433. doi:10.1073/pnas.1112937108

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proc Natl Acad Sci U S A, 110*(41), 16616-16621. doi:10.1073/pnas.1315235110

Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Glasser, M. F., . . . Yacoub, E. (2010). Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. *PLoS One, 5*(12), e15710.

Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: a meta-analysis of neuroimaging studies on text comprehension. *Hum Brain Mapp, 29*(5), 581-593. doi:10.1002/hbm.20422

Ferstl, E. C., & von Cramon, D. Y. (2002). What does the frontomedian cortex contribute to language processing: coherence or theory of mind? *Neuroimage, 17*(3), 1599-1612. doi:10.1006/nimg.2002.1247

Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes, 25*(2-3), 285-307. doi:10.1080/01638539809545029

Friederici, A. D., Rüschemeyer, S.-A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cerebral Cortex, 13*(2), 170–177.

Gambi, C., & Pickering, M. J. (2017). Models linking production and comprehension. *The handbook of psycholinguistics*, 157-181.

Gao, Z., Zheng, L., Chiou, R., Gouws, A., Krieger-Redwood, K., Wang, X., . . . Jefferies, E. (2021). Distinct and Common Neural Coding of Semantic and Non-semantic Control Demands. *Neuroimage*, 118230.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends Cogn Sci, 8*(1), 8-11. doi:10.1016/j.tics.2003.10.016

Gloser, G., & Deser, T. (1992). A Comparison of Changes in Macrolinguistic and Microlinguistic Aspects of Discourse Production in Normal Aging. *Journal of Gerontology, 47*(4), 266-272.

Gold, D. P., Andres, D., Arbuckle, T. Y., & Schwartzman, A. (1988). Measurement and correlates of verbosity in elderly people. *Journal of Gerontology, 43*(2), 27-33.
Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology, 4*, 416. doi:10.3389/fpsyg.2013.00416
Heidlmayr, K., Weber, K., Takashima, A., & Hagoort, P. (2020). No title, no theme: The joined neural space between speakers and listeners during production and comprehension of multi-sentence discourse. *Cortex, 130*, 111-126. doi:10.1016/j.cortex.2020.04.035
Hoffman, P. (2019). Reductions in prefrontal activation predict off-topic utterances during speech production. *Nat Commun, 10*(1), 515. doi:10.1038/s41467-019-08519-0
Hoffman, P., Cogdell-Brooke, L., & Thompson, H. E. (2020). Going off the rails: Impaired coherence in the speech of patients with semantic control deficits. *Neuropsychologia, 146*, 107516. doi:10.1016/j.neuropsychologia.2020.107516
Hoffman, P., Loginova, E., & Russell, A. (2018a). Poor coherence in older people's speech is explained by impaired semantic and executive processes. *eLife, 7*, e38907. doi:10.7554/eLife.38907
Hoffman, P., McClelland, J. L., & Lambon Ralph, M. A. (2018b). Concepts, control, and context: A connectionist account of normal and disordered semantic cognition. *Psychol Rev, 125*(3), 293-328. doi:10.1037/rev0000094
Humphreys, G. F., & Lambon Ralph, M. A. (2015). Fusion and Fission of Cognitive Functions in the Human Parietal Cortex. *Cerebral Cortex, 25*, 3547-3560. doi:10.1093/cercor/bhu198
Jackson, R. L. (2021). The neural correlates of semantic control revisited. *Neuroimage, 224*, 117444. doi:10.1016/j.neuroimage.2020.117444
Jacoby, N., & Fedorenko, E. (2020). Discourse-level comprehension engages medial frontal Theory of Mind brain regions even for expository texts. *Lang Cogn Neurosci, 35*(6), 780-796. doi:10.1080/23273798.2018.1525494
Jefferies, E. (2013). The neural basis of semantic cognition: converging evidence from neuropsychology, neuroimaging and TMS. *Cortex, 49*(3), 611-625. doi:10.1016/j.cortex.2012.10.008
Kintsch, W., & van Dijk, T. A. (1978). Toward a Model of Text Comprehension and Production. *Psychological Review, 85*(5).
Kintz, S., Fergadiotis, G., & Wright, H. H. (2016). Aging effects on discourse production. In H. H. Wright (Ed.), *Cognition, Language and Aging* (pp. 81-106). Amsterdam/Philadelphia: John Benjamins Publishing Company.
Kundu, P., Voon, V., Balchandani, P., Lombardo, M. V., Poser, B. A., & Bandettini, P. A. (2017). Multi-echo fMRI: A review of applications in fMRI denoising and analysis of BOLD signals. *Neuroimage, 154*, 59-80. doi:10.1016/j.neuroimage.2017.03.033
Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., & Holcomb, P. J. (2006). Making sense of discourse: an fMRI study of causal inferencing across sentences. *Neuroimage, 33*(1), 343-361. doi:10.1016/j.neuroimage.2006.06.001
Kurby, C. A., & Zacks, J. M. (2012). Starting from scratch and building brick by brick in comprehension. *Memory & cognition, 40*(5), 812-826.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13). doi:10.18637/jss.v082.i13

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nat Rev Neurosci, 18*(1), 42-55. doi:10.1038/nrn.2016.150

Landauer, T., & Dumais, S. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review, 104*(2), 211-240. doi:10.1037/0033-295X.104.2.211

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. doi:10.1016/j.neuroimage.2020.117019

Lerner, Y., Honey, C. J., Katkov, M., & Hasson, U. (2014). Temporal scaling of neural responses to compressed and dilated natural speech. *J Neurophysiol, 111*(12), 2433-2444. doi:10.1152/jn.00497.2013

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J Neurosci, 31*(8), 2906-2915. doi:10.1523/JNEUROSCI.3684-10.2011

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1-38.

Maguire, E. A., Frith, C. D., & Morris, R. G. M. (1999). The functional neuroanatomy of comprehension and memory: the importance of prior knowledge. *Brain, 122*, 1839-1850.

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., . . . Smallwood, J. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc Natl Acad Sci U S A, 113*(44), 12574-12579. doi:10.1073/pnas.1608282113

Marini, A., & Andreetta, S. (2016). Age-related effects on language production *Cognition, Language and Aging* (pp. 55-79).

Marini, A., Zettin, M., & Galetto, V. (2014). Cognitive correlates of narrative impairment in moderate traumatic brain injury. *Neuropsychologia, 64*, 282-288. doi:10.1016/j.neuropsychologia.2014.09.042

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. *Science, 315*(5810), 393-395.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and instruction, 14*(1), 1-43.

Mineroff, Z., Blank, I. A., Mahowald, K., & Fedorenko, E. (2018). A robust dissociation among the language, multiple demand, and default mode networks: Evidence from inter-region correlations in effect size. *Neuropsychologia, 119*, 501-511. doi:10.1016/j.neuropsychologia.2018.09.011

Moeller, S., Yacoub, E., Olman, C. A., Auerbach, E., Strupp, J., Harel, N., & Uğurbil, K. (2010). Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic resonance in medicine, 63*(5), 1144-1153.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97-113.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169-190.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*(3), 105-110.
Pickering, M. J., & Garrod, S. (2021). *Understanding Dialogue: Language Use and Social Interaction*: Cambridge University Press.
Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience, 13*(10), 713-726.
Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proc Natl Acad Sci U S A, 111*(43), E4687-4696. doi:10.1073/pnas.1323812111
Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., & Hasson, U. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat Commun, 7*, 12141. doi:10.1038/ncomms12141
Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science, 18*(449-455).
Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proc Natl Acad Sci U S A, 107*(32), 14425-14430. doi:10.1073/pnas.1008662107
Tong, Y., Chen, Q., Nichols, T. E., Rasetti, R., Callicott, J. H., Berman, K. F., . . . Mattay, V. S. (2016). Seeking Optimal Region-Of-Interest (ROI) Single-Value Summary Measures for fMRI Studies in Imaging Genetics. *PLoS One, 11*(3), e0151391. doi:10.1371/journal.pone.0151391
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. doi:10.1016/j.neuroimage.2020.117074
Whitney, C., Huber, W., Klann, J., Weis, S., Krach, S., & Kircher, T. (2009). Neural correlates of narrative shifts during auditory story comprehension. *Neuroimage, 47*(1), 360-366. doi:10.1016/j.neuroimage.2009.04.037
Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & van den Bosch, A. (2016). Prediction During Natural Language Comprehension. *Cereb Cortex, 26*(6), 2506-2516. doi:10.1093/cercor/bhv075
Wise, R. J. S., & Geranmayeh, F. (2016). Sentence and Narrative Speech Production *Neurobiology of Language* (pp. 751-762).
Wright, H. H., Koutsoftas, A. D., Capilouto, G. J., & Fergadiotis, G. (2014). Global coherence in younger and older adults: Influence of cognitive processes and discourse type. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn, 21*(2), 174-196. doi:10.1080/13825585.2013.794894
Xu, J., Moeller, S., Auerbach, E. J., Strupp, J., Smith, S. M., Feinberg, D. A., . . . Uğurbil, K. (2013). Evaluation of slice accelerations using multiband echo planar imaging at 3 T. *Neuroimage, 83*, 991-1001.
Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods, 8*(8), 665-670. doi:10.1038/nmeth.1635
Yarkoni, T., Speer, N. K., & Zacks, J. M. (2008). Neural substrates of narrative comprehension and memory. *Neuroimage, 41*(4), 1408-1425. doi:10.1016/j.neuroimage.2008.03.062
Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., . . . Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by

intrinsic functional connectivity. *J Neurophysiol, 106*(3), 1125-1165. doi:10.1152/jn.00338.2011

Yeshurun, Y., Nguyen, M., & Hasson, U. (2021). The default mode network: where the idiosyncratic self meets the shared social world. *Nature Reviews Neuroscience, 22*(3), 181-192.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological Bulletin, 133*(2), 273.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation Models in Language Comprehension and Memory. *Psychological Bulletin, 123*(2), 162-185.