Domain-general brain regions do not track linguistic input as closely as language-selective regions

Idan Blank[1] and Evelina Fedorenko[2,3]

[1]Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139, USA
[2]Department of Psychiatry, Massachusetts General Hospital, Charlestown, MA 02129, USA
[3]Department of Psychiatry, Harvard Medical School, Boston, MA 02115, USA

Abstract

Language comprehension engages a cortical network of left frontal and temporal regions. Activity in this network is language-selective, showing virtually no modulation by non-linguistic tasks. In addition, language comprehension engages a second network consisting of bilateral frontal, parietal, cingulate, and insular regions. Activity in this "Multiple Demand (MD)" network scales with comprehension difficulty, but also with cognitive effort across a wide range of non-linguistic tasks in a domain-general fashion. Given the functional dissociation between the language and MD networks, their respective contributions to comprehension are likely distinct, yet such differences remain elusive. Critically, given that each network is sensitive to some linguistic features, prior research has assumed – implicitly or explicitly – that both networks track linguistic input closely, and in a manner consistent across individuals. Here, we used fMRI to directly test this assumption by comparing the BOLD signal time-courses in each network across different people listening to the same story. Language network activity showed fewer individual differences, indicative of closer input tracking, whereas MD network activity was more idiosyncratic and, moreover, showed lower reliability within an individual across repetitions of a story. These findings constrain cognitive models of language comprehension by suggesting a novel distinction between the processes implemented in the language and MD networks.

Significance Statement: Language comprehension recruits both language-specific mechanisms and domain-general mechanisms that are engaged in many cognitive processes. In the human cortex, language-selective mechanisms are implemented in the left-lateralized "core language network", whereas domain-general mechanisms are implemented in the bilateral "Multiple Demand (MD)" network. Here, we report the first direct comparison of the respective contributions of these networks to naturalistic story comprehension. Using a novel combination of neuroimaging approaches we find that MD regions track stories less closely than language regions. This finding constrains the possible contributions of the MD network to comprehension, contrasts with accounts positing that this network has continuous access to linguistic input, and suggests a new typology of comprehension processes based on their extent of input tracking.

Introduction

A key desideratum for a theory of language comprehension is to specify the division of linguistic labor across distinct cognitive mechanisms. Insofar as distinct mechanisms are implemented in separable neural populations, one promising approach to advance such theories is to functionally characterize brain regions or networks that are engaged in comprehension and establish their respective roles. Indeed, high-level language processing recruits several large-scale networks, each exhibiting a unique functional profile. Among these, the network most critical to comprehension is the "core language network", a set of left frontal and temporal regions. This network is robustly engaged in language processing (e.g., Binder et al., 1997; Jung-Beeman, 2005; Fedorenko et al., 2010; Menenti et al., 2011) across languages (e.g., Sebastian et al., 2011), presentation modalities (Chee et al., 1999; Buchweitz et al., 2009; Fedorenko et al., 2010; Braze et al., 2011; Vagharchakian et al., 2012), and developmental experiences (Neville et al., 1998; Bedny et al., 2011). It exhibits sensitivity to linguistic features such as lexico-semantic information and syntactic structure (Keller et al., 2001; Fedorenko et al., 2012b; Bautista and Wilson, 2016; Blank et al., 2016) but, critically, shows virtually no engagement in non-linguistic tasks and is therefore language-selective (Fedorenko et al., 2011; for a review, see: Fedorenko and Varley, 2016).

In addition, language comprehension engages the "multiple demand (MD)" network (Duncan, 2010), consisting of frontal, parietal, cingulate, and insular regions bilaterally. Activity in this network is sensitive to comprehension difficulty, increasing in response to e.g., temporary ambiguity and non-local syntactic dependencies (for a review, see: Fedorenko, 2014). However, this network similarly scales its response with cognitive effort across a wide range of non-linguistic tasks (Duncan and Owen, 2000; Miller and Cohen, 2001; Braver et al., 2003; Dosenbach et al., 2006; Cole and Schneider, 2007; Fedorenko et al., 2013; Hugdahl et al., 2015) and is therefore domain-general.

The strikingly different functional profiles of the language and MD networks are evident not only in task-based neuroimaging studies but also in relatively unconstrained, task-free paradigms where these networks show unsynchronized activity fluctuations during naturalistic cognition (Blank et al., 2014). A similar double dissociation is reported by neuropsychological studies: damage to the language network leads to language impairments (Broca, 1861/2006; Dax, 1863; Wernicke, 1874/1969; Geschwind, 1970; Bates et al., 2003) but leaves other high-level cognitive functions largely intact (Fedorenko and Varley, 2016); whereas damage to the MD network impairs executive functions (Luria, 1966/2012; Fuster, 1989; Woolgar et al., 2010), but can leave comprehension mostly unimpaired (for a review, see: Fedorenko, 2014). Finally, language and executive functions follow different developmental trajectories (Fedorenko, 2014). These converging findings, of course, do not imply that domain-general executive

functions do not play a role in language processing (for discussions, see Fedorenko, 2014; Geranmayeh et al., 2014). Indeed, a number of studies have provided causal evidence linking executive control to language (Wiener et al., 2004; Fridriksson et al., 2006; Amici et al., 2007; Murray, 2012). However, they provide evidence that the language and MD networks support distinct computations and their contributions to comprehension fundamentally differ.

Nonetheless, the precise nature of the respective contributions remains elusive, as most prior neuroimaging studies of comprehension have not couched their findings in terms of the functional distinction between the language and MD networks. To the extent that available accounts do draw this distinction, however implicitly, they suggest that the two networks differ in either the input features that they each track or the operations they each engage in while processing such input (e.g., Novick et al., 2005; Thompson-Schill et al., 2005; Hickok and Poeppel, 2007; Friederici, 2012; Hagoort, 2013; Fedorenko, 2014). Critically, the various postulated roles of these two networks appear to depend on continuous access to the unfolding input: such access underlies any suggestion that neural activity co-varies with a certain linguistic feature, e.g., lexical information, compositional structure, local ambiguity or parsing difficulty. However, this assumption – which is crucial for understanding the contributions of the two networks to language processing – has not been empirically evaluated.

Here, we use fMRI to directly test this assumption. Namely, we measure activity fluctuations in the language and MD networks during story comprehension and estimate how tightly coupled those fluctuations are to the story. Consistent with current views, one might hypothesize that both networks exhibit equally close tracking of linguistic input. Alternatively, one network might track the input less closely than the other, in which case the space of operations such a network could support would be importantly constrained. Such a finding would indicate that the contributions of the two networks to comprehension differ more fundamentally than is presently assumed.

Materials and Methods

Below, we outline and motivate our methodology. Specifically, we describe a novel combination of existing approaches that is designed to meet four criteria: (i) high functional resolution for identifying brain networks; (ii) a naturalistic paradigm suitable for studying comprehension in all its richness (cf. traditional task-based paradigms); (iii) direct comparisons of brain networks for valid statistical inferences; and (iv) reproducibility of results.

In order to evaluate the extent of input tracking in the language and MD networks, we first must define the cortical regions of interest that constitute these networks. In this process, we must account for the fact that individual brains are highly variable in the

mapping of high-level cognitive functions onto macro-anatomical landmarks. This variability, evident in the temporal cortex (Jones and Powell, 1970; Gloor, 1997; Wise et al., 2001) and especially in the frontal cortex (Amunts et al., 1999; Tomaiuolo et al., 1999; Chein et al., 2002) where language and MD regions lie side-by-side (Fedorenko et al., 2012a), renders anatomical localization precarious (Juch et al., 2005; Poldrack, 2006; Fischl et al., 2008; Tahmasebi et al., 2011; Frost and Goebel, 2012). For these reasons, we similarly cannot rely on functional localization at the level of an entire sample using group-based analyses (Saxe et al., 2006; Fedorenko and Kanwisher, 2009). Therefore, we functionally localize language and MD regions individually in each participant. This approach allows us to pool data from the same functional regions across participants even when those regions do not align well spatially.

Following functional localization, we evaluate how closely the language and MD networks track linguistic input during naturalistic comprehension. Our interest in naturalistic input is threefold: first, some brain regions respond more reliably to richly structured natural input compared to experimentally controlled input (Hasson et al., 2010). Second, unlike traditional experimental paradigms which often require participants to perform artificial tasks on linguistic materials, naturalistic comprehension more closely approximates language processing "in the wild", where the primary goal is the extraction of meaning. Therefore, this "task free" paradigm provides an important complementary approach for evaluating the contributions of the MD regions to comprehension, especially given that these regions operate in a task-dependent manner (Miller and Cohen, 2001; Sreenivasan et al., 2014; D'Esposito and Postle, 2015). And third, naturalistic comprehension requires all aspects of the input to be combined into a single rich representation, unlike experimental stimuli and tasks that focus on particular linguistic features and have lower ecological validity. Therefore, we record the BOLD signal fluctuations of language and MD regions while participants passively listen to stories, where the only explicit task is to comprehend the story's content.

Following Lerner et al. (2011), we reasoned that if a given network closely tracked the story such that fluctuations in its BOLD signal were stimulus-locked, then its signal time-course would be similar across participants and would thus show a high Inter-Subject Correlation (ISC) (Hasson et al., 2004). Hence, we use ISC as an index of input tracking. Critically, ISC is a "model-free" measure: instead of testing how well signal time-courses can be explained by certain pre-specified, hypothesis-driven predictors, each participant's empirical data serve as the model compared against the data from the other participants.

This data-driven method has been successfully used to demonstrate that broad cortical swathes do track stories to significant extents (Wilson et al., 2008; Lerner et al., 2011; Honey et al., 2012; Regev et al., 2013; Silbert et al., 2014; Schmälzle et al., 2015), proposing a neural correlate of "shared understanding" across individuals (Hasson et al.,

2012). Nevertheless, prior studies have measured ISCs in a voxel-wise fashion, whereby brains were first anatomically aligned and, then, each stereotaxic location served in turn as a basis for comparing signal time-courses across participants. Relating the resulting cortical topology of ISCs to the topology of known functional brain networks could then proceed only through "reverse inference" (Poldrack, 2006). Moreover, voxel-wise comparisons across participants rely on the invalid assumption that a given anatomical location has a common function across individuals. To relax this assumption, here we augment the ISC framework by comparing signal time-courses across regions that are functionally defined. This allows us to focus on, and compare between, language and MD regions, such that we can tie our findings to the wealth of prior literature characterizing the response profiles of those networks.

In addition, we augment the statistical approach adopted in early studies of ISCs by directly testing the correlations in the language network against those in the MD network. Such an explicit comparison between networks allows for more nuanced inferences compared to those licensed when each network is separately tested against a null baseline and differences across networks are indirectly inferred (cf. Lerner et al., 2011, for the latter) (Nieuwenhuis et al., 2011).

Finally, we demonstrate that our results are reproducible, by reporting two replications of our main, story comprehension experiment: the first is a direct replication with a subset of the original stories; the second is a conceptual replication with a new, even more naturalistic story.

Participants

Fifty participants between the ages of 18 and 47, recruited from the MIT student body and the surrounding community, were paid for participation. Two participants were removed from the analysis due to poor quality of the functional localizer data and three more were removed due to poor segmentation of their anatomical scan. Of the remaining 45 participants (30 females; mean age 23.5, SD 4.8), 19 were tested in the main experiment, 13 in the first replication and 19 in the second replication (the first and third groups were partially overlapping). In addition, 15 of these participants were tested in a control experiment (described below): these included 8 participants from the main experiment, 2 from the first replication, and one who participated in both the main experiment and the second replication. Forty-one participants were right-handed (based on the Edinburgh Handedness Inventory; Oldfield, 1971), and the remaining four left-handed participants had a left-lateralized language network (for motivation to include left-handers in cognitive neuroscience research, see Willems et al., 2014). All participants were native English speakers and gave informed consent in accordance with the requirements of MIT's Committee on the Use of Humans as Experimental Subjects (COUHES).

Design, stimuli and procedure

Language localizer task. The task used to localize the language network is described in detail in Fedorenko et al. (2010). Briefly, we used a reading task contrasting sentences and lists of unconnected, pronounceable nonwords (Figure 1a) in a standard, deterministic blocked design with a counterbalanced order across runs (for timing parameters, see Table 1). Stimuli were presented one word / nonword at a time. For the first ten participants only, each trial ended with a memory probe and they had to indicate, via a button press, whether or not that probe had appeared in the preceding sequence of words / nonwords. The remaining participants instead read the materials passively (we included a button-pressing task at the end of each trial, to help participants remain alert). Importantly, this localizer has been shown to generalize across task manipulations: the sentences > nonwords contrast robustly activates the fronto-temporal language network regardless of the task (Fedorenko et al., 2010). The regions identified by this contrast engage in a broad range of linguistic processes including (but not limited to) lexico-semantic processes and combinatorial syntactic and semantic processes (Fedorenko et al., 2012b; Blank et al., 2016; Fedorenko et al., 2016; Fedorenko et al., 2017). Moreover, this localizer identifies the same regions that are localized with a broader contrast, between recorded natural speech and its acoustically-degraded version (Scott et al., 2016).

Table 1. Timing parameters for the different versions of the language localizer task.

	Version		
	A	**B**	**C**
Number of participants	35	5	5
Task: Passive Reading or Memory?	PR	M	M
Words / nonwords per trial	12	12	12
Trial duration (ms)	6,000	6,000	6,000
Fixation	100	---	---
Presentation of each word / nonword	450	350	350
Fixation	500	300	300
Memory probe	---	1,000	1,000
Fixation	---	500	500
Trials per block	3	3	3
Block duration (s)	18	18	18
Blocks per condition (per run)	8	8	6
Conditions	Sentences Nonwords	Sentences Nonwords	Sentences Nonwords Word-lists[1]
Fixation block duration (s)	14	18	18
Number of fixation blocks	5	5	4
Total run time (s)	358	378	396
Number of runs	2	2	2-3

[1] Used for the purposes of another experiment; see (Fedorenko et al., 2010)

MD localizer task. Regions of the MD network were localized using a spatial working-memory task contrasting a hard version with an easy version (Figure 1b). On each trial (8s), participants saw a 3×4 grid and kept track of eight (hard version) or four (easy version) randomly generated locations that were sequentially flashed two at a time or one at a time, respectively (1s per flash). Then, participants indicated their memory for these locations in a two-alternative, forced-choice paradigm via a button press (3s total). Feedback was immediately provided upon choice (or lack thereof) (250ms). Hard and easy conditions were presented in a standard blocked design (4 trials in a 32s block, 6 blocks per condition per run) with a counterbalanced order across runs. Each run included 4 blocks of fixation (16s each) and lasted a total of 448s. Thirty-nine participants completed 1-2 runs of the localizer. The remaining six participants either provided poor-quality data (5 participants) or were not run on this task (1 participant). For this latter group, MD regions were localized with data from the language localizer task, using the (reverse) nonwords > sentences contrast. Both the hard > easy contrast and the nonwords > sentences contrast have been previously demonstrated to robustly and reliably identify the MD network (Fedorenko et al., 2013) (these participants did not differ from the rest of the sample in the dependent variables; see Table 2a).

Story comprehension task. In the main experiment, each subject listened to 1-4 stories (one story: n=7; two: n=3; three: n=2; four: n=7; duration: 270s-364s) over scanner-safe headphones (Sensimetrics, Malden, MA). Stories were constructed based on publicly available fairy tales and short stories:

(i) "The Legend of the Bradford Boar" (by E. H. Hopkinson; unedited version: www.make4fun.com/stories/British-short-story/3917-The-Legend-of-the-Bradford-Boar-by-E-H-Hopkinson)

(ii) "Aqua; or the Water Baby" (by Kate Douglas Wiggin; unedited version: fullreads.com/literature/aqua-or-the-water-baby/)

(iii) "The King of the Birds" (by The Brothers Grimm; unedited version: www.apples4theteacher.com/holidays/bird-day/short-stories/the-king-of-the-birds.html)

(iv) "Elvis Died at the Florida Barber College" (by Roger Dean Kiser; unedited version: www.eastoftheweb.com/short-stories/UBooks/ElvDie.shtml).

These stories were edited to include a variety of linguistic phenomena that have been shown to increase processing difficulty in numerous prior behavioral sentence processing studies and which recruit the MD network (Figure 1c; see also Supplementary Materials). As a result of these edits, comprehension difficulty was robustly modulated across each story. Namely, self-paced reading times in a separate sample (n=181 participants) were reliably predicted by measures of linguistic complexity (Shain et al., 2016). Moreover, in

these stories, some measures of complexity influenced online behavior more robustly than in studies that have used unedited texts, plausibly because the relevant linguistic phenomena do not naturally occur with sufficiently high frequency (Collins, 1996; Roland et al., 2007; Futrell et al., 2015). Further, even though the stories in the current experiments were presented via the auditory rather than visual modality, we still expect them to successfully modulate processing difficulty because reading-time effects generalize to on-line listening (Ferreira et al., 1996; Waters and Caplan, 2001) (see also Table 2b for evidence that our neuroimaging results generalize to visual story presentation).

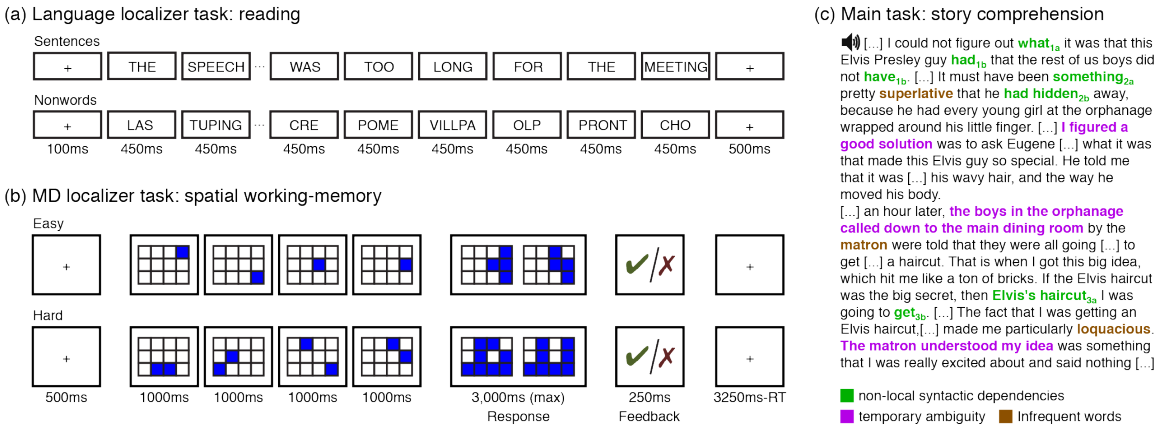

Figure 1. Experimental tasks. (a) The reading task used to localize language regions, based on the critical contrast sentences > nonwords. (b) The spatial working-memory task used to localize MD regions, based on the critical contrast hard > easy. (c) An excerpt from a story used in the main comprehension experiment. Linguistic phenomena that increase processing difficulty and have been shown to recruit the MD network, but are naturally infrequent, were edited into the text. These include non-local syntactic dependencies (green; words in this relation have subscripts with the same number but different letters); temporary ambiguity (purple), where a likely initial parse is later revealed to be wrong; and low-frequency words (brown).

In the first replication, participants listened to stories (i) and (iii) used in the main experiment (these data were originally collected for the purpose of a separate experiment; participants also listened to the other two stories, but performed a simultaneous, unrelated task during those trials). In the second replication, participants listened to an autobiographical story ("Pie-man," told by Jim O'Grady) recorded at a live storytelling event ("The Moth" storytelling event, NYC). This story (duration: 420s) did not undergo linguistic editing and was thus even more naturalistic than the previous stories. Each story started and ended with 16s seconds of fixation (and music, for the Pie-man story) that were not analyzed.

To test the reliability of signal time-courses in the language and MD networks, participants in the control experiment listened to the same stories twice, either within the

same scanning session (approximately one hour apart, *n*=7) or in separate sessions (6.5-21.5 months apart, *n*=8) (4 participants listened to the same story twice within the same session and then, once more, in a separate session).

After each story, participants answered 6-12 comprehension questions that required attentive listening (i.e., could not have been answered correctly based on common knowledge). For the main experiment and the first replication, participants answered two-alternative forced-choice (2AFC) questions via a button press while in the scanner. For the second replication, participants filled in a 4AFC questionnaire after the scanning session. For eight participants, answers to these questions were not recorded due to equipment malfunction (these participants did not differ from the rest of the sample in the dependent variables; see table 2a). The remaining 37 participants demonstrated good comprehension, with a negatively skewed accuracy distribution (mode=100%, median=87.5%, semi-interquartile range=12.85%).

Data acquisition and preprocessing

Data acquisition. Whole-brain structural and functional data were collected on a whole-body 3 Tesla Siemens Trio scanner with a 32-channel head. T1-weighted structural images were collected in 176 axial slices with 1mm isotropic voxels (repetition time (TR) = 2,530ms; echo time (TE) = 3.48ms). Functional, blood oxygenation level-dependent (BOLD) data were acquired using an EPI sequence with a 90° flip angle and using GRAPPA with an acceleration factor of 2; the following parameters were used: thirty-one 4mm thick near-axial slices acquired in an interleaved order (with 10% distance factor), with an in-plane resolution of 2.1mm × 2.1mm, FoV in the phase encoding (A >> P) direction 200mm and matrix size 96mm × 96mm, TR = 2000ms and TE = 30ms. The first 10s of each run were excluded to allow for steady state magnetization.

Spatial preprocessing. Data preprocessing was carried out with SPM5 (using default parameters, unless specified otherwise) and supporting, custom MATLAB scripts. Preprocessing of anatomical data included normalization into a common space (Montreal Neurological Institute (MNI) template), resampling into 2mm isotropic voxels, and segmentation into probabilistic maps of the gray matter, white matter (WM) and cerebrospinal fluid (CSF). Preprocessing of functional data included motion correction (realignment to the mean image using 2nd-degree b-spline interpolation), normalization (estimated for the mean image using trilinear interpolation), resampling into 2mm isotropic voxels, smoothing with a 4mm FWHM Gaussian filter and high-pass filtering at 200s.

Temporal preprocessing. Additional preprocessing of data from the story comprehension runs was carried out using the CONN toolbox (Whitfield-Gabrieli and Nieto-Castanon, 2012) with default parameters, unless specified otherwise. Five temporal principal components of the BOLD signal time-courses extracted from the WM were

regressed out of each voxel's time-course; signal originating in the CSF was similarly regressed out. Six principal components of the six motion parameters estimated during offline motion correction were also regressed out, as well as their first time derivative. Next, the residual signal was bandpass filtered (0.008–0.09 Hz) to preserve only low-frequency signal fluctuations (Cordes et al., 2001). This filtering did not influence the results reported below.

Participant-specific functional localization of language and MD networks

Modeling localizer data. For each localizer task, a standard mass univariate analysis was performed in SPM5 whereby a general linear model estimated the effect size of each condition in each experimental run. These effects were each modeled with a boxcar function (representing entire blocks) convolved with the canonical Hemodynamic Response Function (HRF). The model also included first-order temporal derivatives of these effects, as well as nuisance regressors representing entire experimental runs and offline-estimated motion parameters. The obtained beta weights were then used to compute the functional contrast of interest: for the language localizer, sentences > nonwords, and for the MD localizer, hard > easy (or nonwords > sentence for 6 participants; see Design, stimuli and procedure).

Defining functional regions of interest (fROIs). Language and MD fROIs were defined based on functional contrast maps from the localizer experiments. These maps were first restricted to include only gray matter voxels by excluding voxels that were more likely to belong to either the white matter or the cerebrospinal fluid based on SPM's probabilistic segmentation of the participant's structural data.

Then, fROIs in the language network were defined using group-constrained, participant-specific localization (Fedorenko et al., 2010). For each participant, the map of the sentences > nonwords contrast was intersected with binary masks that constrained the participant-specific language network to fall within areas where activations for this contrast are relatively likely across the population. These masks are based on a group-level representation of the contrast obtained from a previous sample. We used 8 such masks in the left-hemisphere, including regions in the posterior, mid-posterior, mid-anterior and anterior temporal lobe, as well as in the middle frontal gyrus, the inferior frontal gyrus and its orbital part (Figure 2a). These masks were mirror-projected onto the right-hemisphere to create 8 homologous masks (the masks cover significant parts of the cortex, so their mirrored version is likely to encompass the right-hemisphere homologue of the left-hemisphere language network, despite possible hemispheric asymmetries in their precise locations). In each of the resulting 16 masks, a participant-specific language fROI was defined as the top 10% of voxels with the highest contrast values. This top *n*% approach ensures that fROIs can be defined in every participant and that their sizes are

the same across participants, allowing for generalizable results (Nieto-Castañón and Fedorenko, 2012).

fROIs in the MD network were similarly defined (using the "top 10%" approach) based on the hard > easy contrast in the spatial working-memory game. Here, instead of using binary masks based on group-level functional data, we used anatomical masks (Tzourio-Mazoyer et al., 2002; see Fedorenko et al., 2013; Blank et al., 2014). Nine masks were used in each hemisphere, including regions in the middle frontal gyrus and its orbital part, the opercular part of the inferior frontal gyrus, the precental gyrus, the posterior and inferior parts of the partietal lobe, the insula, and supplementary motor area and the cingulate cortex (Figure 2b). Based on prior findings (Dosenbach et al., 2006; Dosenbach et al., 2007; Nomura et al., 2010; Power et al., 2011; Mantini et al., 2013), we grouped the resulting fROIs into two functionally distinct sub-networks: fronto-parietal (first 5 masks) and cingulo-opercular (last 3 masks). Similar results were obtained when fROIs were instead grouped by hemisphere. (We note that functional masks derived for the MD network based on 197 participants significantly overlapped with the anatomical masks; we chose to use the anatomical masks in order to maintain comparability between our functional data and data from previous studies that have used these masks).

Any voxels that were identified by both the language and the MD localizer were excluded from analysis (this procedure did not influence the results). Language fROIs had a median overlap of 0 voxels with the MD network (inter-quartile range: 2.1%). MD fROIs also had a median overlap of 0 voxels with the language network (inter-quartile range: 1.6%). The resulting fROIs had an average size of 247±77 voxels in the language network, and 212±111 voxels in the MD network.

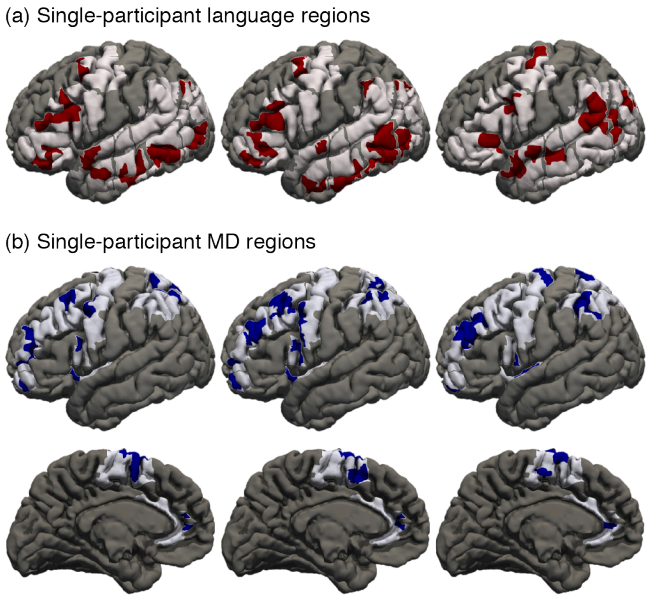

Figure 2. Functional regions of the language and MD networks. (a) LH language regions in 3 individual participants are shown in dark red. These regions were localized with a reading task (see Figure 1a). These regions were constrained to fall within eight broad areas where activations for this task are common across the population, shown in light pink. These areas were defined based on group-level data from a previous sample (Fedorenko et al., 2010). (b) LH MD regions of the same 3 participants are shown in dark blue. These regions were localized with a spatial working-memory task (see Figure 1b). These regions were constrained to fall within nine broad areas where activations for this localizer are common across the population, shown in light blue. These areas were anatomically defined (Fedorenko et al., 2013). Apparent overlap between language and MD fROIs is illusory and due to projection onto the cortical surface.

Critical analysis: inter-subject correlations (ISCs)

Computing ISCs. For each participant and fROI, BOLD signal time-courses recorded during story comprehension were extracted from each voxel beginning 6 seconds following the onset of the story (to exclude an initial rise in the hemodynamic response relative to fixation, which could increase ISCs). These time-courses were first temporally *z*-scored in each voxel and then averaged across voxels. Next, to ensure that the resulting signal time-course reflected the tracking of high-level linguistic information and not low-level sensory information, we removed from it any variance that was explained by activity in the auditory cortex. Specifically, the signal was regressed against signals extracted from anatomically defined regions around the postero-medial and antero-lateral sections of Heschl's gyrus bilaterally (Tzourio-Mazoyer et al., 2002) (this regression did not affect the pattern of results reported here; see Table 2c). Finally, for each story, participant, and fROI we computed an ISC value, namely, Pearson's moment correlation coefficient between the residual time-course and the corresponding average residual time-course across the remaining participants (Lerner et al., 2011). ISCs were Fisher-transformed prior to statistical testing in order to improve normality (Silver and Dunlap, 1987).

Statistical testing. In each fROI, ISCs were then tested for significance against an empirical null distribution based on 1,000 simulated signal time-courses that were generated by phase-randomization of the original data (Theiler et al., 1992). Namely, we generated null distributions for individual participants, fit each distribution with a Gaussian, and analytically combined the resulting parameters across participants. The true ISCs, also averaged across participants, were then *z*-scored relative to these empirical parameters and converted to one-tailed *p*-values.

Critically, ISCs were compared across networks using a linear, mixed-effects regression (Barr et al., 2013) implemented with the "lme4" package in R. In each experiment, ISCs across all fROIs, participants and stories were modeled with a fixed effect of fROI and random intercepts for participant and story. The fixed effect estimates were combined across fROIs within each functional network (LH language, RH language, fronto-parietal MD, and cingulo-opercular MD) and were pairwise compared to each other using the "multcomp" package in R. Hypotheses were two-tailed for the first experiment and one-tailed for the replications and control analyses. In each experiment, *p*-values are reported following False Discovery Rate (FDR) correction for multiple comparisons (Benjamini and Yekutieli, 2001).

For all findings based on linear, mixed-effects regression analyses, similar results were obtained when data for each participant were first averaged across fROIs within each network and pairwise network comparisons (across participants) were then tested using exact permutation tests (Gill, 2007). Therefore, our results are independent of assumptions regarding data normality.

Control analysis: within-subject correlations (WSCs)

Computing WSCs. For each participant who listened to the same story on two separate trials, we computed a within-subject correlation (WSC) value for each fROI by correlating the signal time-courses across the two trials. The resulting correlations were Fisher-transformed.

Note that unlike ISCs, which compare the signal from one participant to an average signal across all other participants, WSCs compare two single-trial signals. Consequently, the two measures are not directly comparable: despite the fact that WSCs are not contaminated by inter-individual variability and should thus be higher than ISCs, ISCs will *de facto* be higher because signal averaging removes a lot of noise from the data. In order to make ISCs comparable to WSCs we therefore computed "pairwise ISCs": for each participant and fROI, we correlated the signal time-course separately with each of the corresponding, individual signal time-courses of the other participants, Fisher-transformed the resulting correlation values, and averaged them.

Statistical tests. Prior to these analyses, we tested whether WSCs in the within-session and across-session datasets differed from each other. To this end, we performed a linear, mixed-effects regression analysis that modeled individual WSCs for all fROIs, participants and stories with a fixed effect of the interaction between fROI and dataset, random intercepts for participant and story, and a random slope for dataset varying by participant (this model was chosen because a fuller model failed to converge). Pairwise contrasts tested whether WSCs in each network were stronger across sessions than within a session. These two groups did not differ from each other in their network WSCs. Therefore, these two sets of data were modeled together in the critical analyses: here, WSCs were compared across networks using the same model that was used to test ISCs, modeling individual WSCs for all fROIs, participants and stories

A similar approach was used for comparing WSCs to pairwise-ISCs. Here, contrasts tested whether pairwise differences between networks observed with WSCs were distinct from those observed with ISCs.

Results

Correlations of network activity across individuals listening to the same story

ISC data are presented in Figure 3. Across stories in the main experiment, the left-hemisphere (LH) language network showed the highest ISCs (Fisher transformed r=0.280), stronger than ISCs in the right-hemisphere (RH) language network (r=0.210; Cohen's d=0.73, z=6.25, $p<10^{-9}$), the fronto-parietal MD (MDfp) network (r=0.136; d=1.07, z=14.12, $p\approx0$) and the cingulo-opercular MD (MDco) network (r=0.117; d=1.32, z=13.51, $p\approx0$). The RH language network, in turn, showed higher ISCs than both the

MDfp network (d=1.07, z=7.27, $p<10^{-11}$) and the MDco network (d=1.04, z=7.72, $p<10^{-13}$). The two MD networks did not differ from each other (d=1.80, z=1.70, p=0.218). The difference between the LH language network and the two MD networks was also observed for each story separately.

In both replication experiments, we again found that ISCs in the LH language network (replication 1: r=0.252; replication 2: r=0.303) were stronger than in the RH language network (r=0.172, d=0.90, z=5.62, $p<10^{-7}$; r=0.250, d=0.77, z=3.35, p=0.001), the MDfp network (r=0.147, d=1.06, z=8.09, $p<10^{-15}$; r=0.160, d=1.29, z=9.95, $p\approx0$) and the MDco network (r=0.114, d=1.33, z=8.95, $p\approx0$; r=0.163, d=1.34, z=8.20, $p<10^{-15}$). ISCs in the RH language network were somewhat stronger than ISCs in the MDfp network (d=0.46, z=1.93, p=0.066; d=0.82, z=6.28, $p<10^{-9}$) and stronger than ISCs in the MDco network (d=0.70, z=3.74, p<0.001; d=0.83, z=5.10, $p<10^{-7}$). The two latter networks reliably differed from each other only in the first replication (d=0.53, z=2.28, p<0.033).

Across these three experiments, we find that signals in the language and MD networks differ in their ISCs and, thus, in the percentage of variance they share across individuals. To further interpret these findings we computed an "upper bound" on ISCs, reflecting the highest values that could be expected in our measurements; namely, we computed ISCs in low-level auditory regions (see Materials and Methods) that track sensory input very closely (Lerner et al., 2011). Combining data across experiments, these auditory ISCs are estimated at r=0.450. Thus, signals in the LH language network (r=0.287) share 40.8% of this "maximum shareable variance" across individuals; signals in the RH language network (r=0.216) share 23%, whereas signals in the MDfp network (r=0.153) and MDco network (r=0.134) share only 11.6% and 8.8%, respectively. Importantly, however, almost all ISCs – even those in MD regions – are significantly greater than expected by chance (Figure 3). Therefore, even domain-general MD regions track stories to a non-trivial extent in spite of doing so substantially and reliably more weakly than the language regions.

Is it possible that other sub-regions of the MD network, not identified by our localizer, track the stories more strongly? To test this possibility, we computed traditional, voxel-based ISCs (based on anatomical alignment of individual brains) and identified, within each mask of the MD network, the voxels that showed the highest ISCs during one story. These voxels served as "alternative fROIs", and we estimated the strength of their ISCs using independent data from another story. The resulting ISCs were weaker than those reported above (Table 3), and the same finding held in "alternative fROIs" identified in the language network. Critically, compared to the original fROIs, the alternative fROIs responded less robustly to the language and MD localizers (responses in the original fROIs were obtained from runs of the localizers that were held-out during fROI definition). For instance, alternative fROIs in the MDco network did not respond

differentially to the hard and easy versions of the spatial working memory task; and alternative fROIs in the RH language network did not respond differentially to sentences and nonwords (Table 3). These decreased functional signatures are likely caused by inter-individual variability in the precise anatomical locations of the language and MD regions, such that a given voxel might belong to a certain network in some participants but not others. Therefore, with no means for establishing functional (rather than anatomical) correspondence across individual brains in areas that lie outside of our localizer-defined fROIs, we do not find evidence for close linguistic tracking of input anywhere in the MD network.

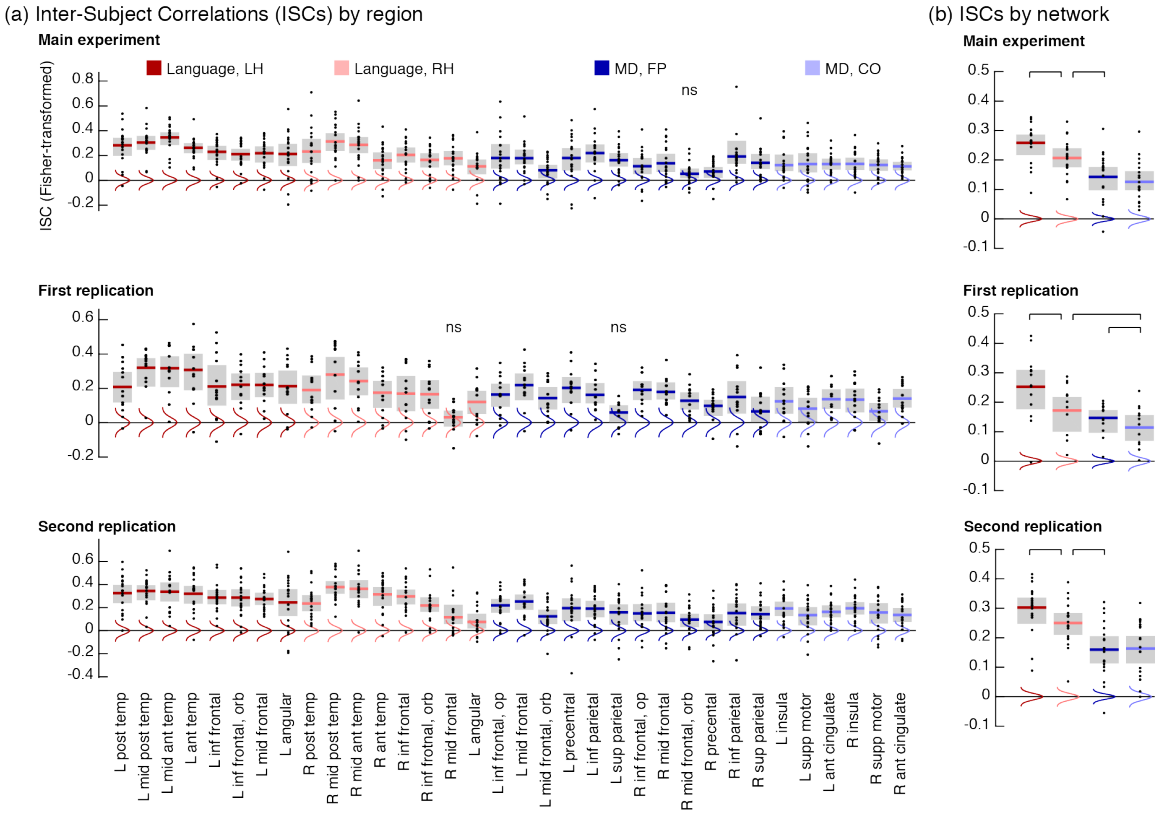

Figure 3. ISCs during story comprehension in the language and MD networks. (a) ISC (Fisher-transformed) for each brain region. Black dots are individual data points. Thick, colored horizontal lines show the average ISCs across participants. Gray rectangles show 95% confidence intervals of these average ISCs (empirically derived using 1,000 permutations). Colored vertical curves show Gaussian fits to empirical null distributions against which average ISCs can be tested (ns, non-significant results at a threshold of 0.05; FDR-corrected). Regions are grouped into 4 functional networks, indicated by color. Across experiments, a replicable pattern emerges where ISCs are stronger in language regions (red) than in MD regions (blue). (b) Mean ISCs within each functional network, same conventions as in (a). Black, horizontal lines connect pairs of networks that significantly differ from one another (in each pair, the left ISC is greater than the right ISCs and all ISCs that are further to the right). L – left; R – right; Post – posterior; Temp – temporal; Mid – middle; Ant – anterior; Inf – inferior; Orb – orbital; Op – opercular; Sup – superior; Supp – supplementary.

Correlations of network activity within individuals listening to a story twice

The relatively low ISCs in MD regions could be interpreted in two ways: on the one hand, MD regions might closely track linguistic input but do so in an idiosyncratic fashion across individuals. For example, if different people find different sections of the story difficult to comprehend, they might each recruit their MD network at respectively different times. In this case, MD activity time-courses would be stimulus-locked for each individual but would differ across individuals. Alternatively, activity in the MD regions might not be closely linked to the linguistic input at all. These two interpretations can be distinguished by correlating signal time-courses within a given individual who is listening to the same story twice (Hasson et al., 2009): if MD activity tracks the story in an idiosyncratic manner across individuals, then it should still be similar across two instances of the same story within an individual; however, if MD activity does not track the story, then it should not exhibit reliable time-courses even within an individual.

Therefore, we scanned several participants listening to stories twice and computed within-subject correlations (WSCs). In line with our findings above, WSCs in the LH language network (r=0.160) were stronger than in the RH language network (r=0.129; d=0.33, z=3.66, p<0.001), the MDfp network (r=0.083; d=0.83, z=8.5, p≈0) and the MDco network (r=0.097; d=1.25, z=6.05, $p<10^{-8}$). WSCs in the RH language network were stronger than those in the MDfp network (d=0.30, z=4.48, $p<10^{-4}$) and the MDco network (d=0.32, z=2.66, p=0.012), but the two latter networks did not differ (Figure 4a). When we directly contrasted WSCs to ISCs (the latter re-computed as "pairwise-ISCs" to be directly comparable to the former; see Materials and Methods) we found that the patterns of results were indistinguishable across the two measures (for all comparisons between WSCs and pairwise-ISCs, p>0.52) (Figure 4b). Therefore, even across story repetitions within a given individual, MD network activity is significantly less reliable than language network activity, indicating that the former, but not the latter, tracks linguistic input closely.

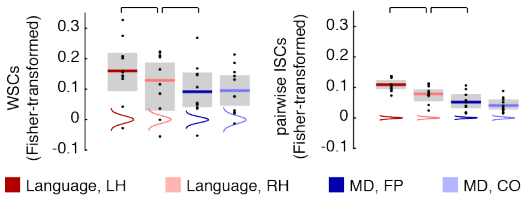

Figure 4. WSCs (left) and pairwise-ISCs (right) during story comprehension in the language and MD networks. Same conventions as in Figure 3.

Table 2. The patterns of story tracking across the language and MD networks generalize across changes in the experimental setup and analysis stream[1,2]

(a) Testing whether ISCs are modulated by task characteristics[3]

Critical dataset	Dataset size	Stories	Comparison dataset	ISC ~ Network×Dataset + Network + Dataset + (1 | ID) + (1 | Story)	
				χ^2	*p*
Participants not administered a comprehension test	*n* = 8	(i) (ii) (iii) (iv)	For each participant: other participants in the same experiment	Interaction: $\chi^2_{(3)}$=2.01	0.57
				Dataset: $\chi^2_{(1)}$=1.44	0.23
Participants with MD fROIs defined as nonwords > sentences	*n* = 6			Interaction: $\chi^2_{(3)}$=0.26	0.97
				Dataset: $\chi^2_{(3)}$=0.23	0.63

(b) Visual presentation of the stories[4,5] (word-by-word, timing matched to the auditory version)

	DV ~ fROI + (1 | ID) + (1 | Story)					
	DV: Inter-subject correlations			DV: Within-subject correlations		
	ISC	*Z*	*p*	WSC	*Z*	*p*
Language LH	0.24	7.57	**<10^{-12}**	0.20	5.58	**<10^{-6}**
Language RH	0.20	6.16	**<10^{-8}**	0.17	4.77	**<10^{-5}**
MDfp	0.15	4.77	**<10^{-5}**	0.10	3.01	**0.006**
MDco	0.08	2.58	**0.015**	0.03	0.85	0.58
Language LH > RH		2.59	**0.016**		1.13	0.42
Language LH > MDfp		5.87	**<10^{-7}**		4.12	**0.0001**
Language LH > MDco		8.39	**≈0**		6.05	**<10^{-7}**
Language RH > MDfp		3.04	**0.004**		2.88	**0.008**
Language RH > MDco		5.99	**<10^{-8}**		5.01	**<10^{-5}**
MDfp > MDco		3.70	**0.0005**		2.78	**0.01**

(c) ISCs computed without regressing out time-series from the auditory cortex[6]

	Critical data[5] ISC ~ fROI + (1 | ID) + (1 | Story)			**Critical data vs. original data[7]** ISC ~ fROI + fROI×Dataset + Dataset + (1 | ID) + (1 | Story)		
	ISC	*Z*	*p*	$ISC_{original}$	$Z_{datasets}$	*p*
Language LH	0.33	10.96	**≈0**	0.28	4.90	**<10^{-4}**
Language RH	0.24	7.91	**<10^{-14}**	0.21	2.83	**0.017**
MDfp	0.14	4.85	**<10^{-5}**	0.13	1.18	0.44
MDco	0.09	3.84	**0.0002**	0.12	0.16	1
Language LH > RH		8.31	**≈0**		1.46	0.35
Language LH > MDfp		18.41	**≈0**		3.05	**0.011**
Language LH > MDco		17.85	**≈0**		3.08	**0.015**
Language RH > MDfp		9.31	**≈0**		1.45	0.31
Language RH > MDco		10.16	**≈0**		1.73	0.24
MDfp > MDco		2.47	**0.02**		0.55	0.95

[1]Within each sub-table and column, *p*-values are FDR-corrected for multiple comparisons.
[2]All results are from linear, mixed-effect regressions with models as specified (ID = participant).
[3] For each participant in the critical dataset, ISCs were computed relative to average data across participants in the respective comparison dataset. ISCs were then averaged across fROIs within each network.
[4] Based on stories (ii) and (iv). Sample size: *n*=11 for ISCs, *n*=7 for WSCs.
[5] Unlike the main experiment, here we test ISCs and WSCs against zero (using a *Z* statistic) and not against an empirical null distribution. These two tests provide similar results.
[6] Data from the main experiment, based on stories (i), (ii), (iii), (iv). Sample size: *n*=19 participants.
[7] Here, ISCs are reported for the original data; the *Z* statistic compares them to ISCs in the critical dataset.

Table 3. Functional profiles of "alternative" fROIs defined as the top 10% of voxels in each mask showing the highest ISCs (computed based on anatomical alignment across individual brains)[1]

	Within dataset (new / original): DV ~ fROI + (1 | ID) Across datasets: DV ~ fROI + Dataset + fROI×Dataset + (1 + Dataset | ID)							
	Language LH		Language RH		MDfp		MDco	
DV	new	original	new	original	new	original	new	original
ISC[2]	0.22 Z=16.2 **p≈0**	0.28 Z=13.4 **p≈0**	0.18 Z=13.5 **p≈0**	0.21 Z=10.1 **p≈0**	0.11 Z=8.2 **$p<10^{-15}$**	0.15 Z=7.6 **$p<10^{-13}$**	0.09 Z=6.3 **$p<10^{-9}$**	0.11 Z=5.0 **$<10^{-5}$**
	Z=4.78, $p<10^{-5}$		**Z=2.38, p=0.03**		**Z=4.05, $p<10^{-4}$**		Z=1.5, p=0.21	
Language localizer[3]: sentences > nonwords	0.17 Z=4.4 **$p<10^{-4}$**	0.63 Z=12.0 **p≈0**	-0.01 Z=-0.4 p=1	0.21 Z=4.0 **$p=10^{-4}$**	-0.24 Z=-7.0 **$p<10^{-11}$**	-0.38 Z=-7.7 **$p<10^{-12}$**	-0.03 Z=-0.7 p=0.77	-0.13 Z=-2.4 **p=0.03**
	Z=11.9, p≈0		**Z=5.9, $p<10^{-7}$**		**Z=-3.9, $p<10^{-3}$**		**Z=-2.3, p=0.04**	
MD localizer[3]: hard > easy	-0.03 Z=-0.4 p=1	-0.12 Z=-1.4 p=0.48	0.06 Z=0.8 p=0.97	0.01 Z=0.09 p=1	0.37 Z=5.22 **$p<10^{-6}$**	0.95 Z=10.5 **p≈0**	0.04 Z=0.5 p=1	0.52 Z=5.3 **$p<10^{-6}$**
	Z=1.4, p=0.42		Z=0.7, p=0.94		**Z=8.4, p≈0**		**Z=5.9, $p<10^{-7}$**	

[1] We compare the data of the first replication reported in the manuscript ("original" dataset) to data derived from the "alternative" fROIs ("new" dataset). The first replication was chosen because it had a sufficient number of participants (n=13) who listened to the same two stories, namely, (i) and (iii).

[2] For the new dataset, one story was used for defining "alternative" fROIs, and the held out story was used to estimate their ISCs independently of the criteria used to define them. The process was then repeated with the two stories in reversed roles, and the resulting two estimates for each fROI were averaged.

[3] For language (MD) fROIs in the original dataset, one run of the language (MD) localizer was used to define fROIs and the second run was then used to estimate their responses independently of the criteria used to define them. The process was then repeated with the two runs in reversed roles, and the resulting two estimates for each fROI were averaged.

Discussion

During story comprehension, a robust and reliable difference in neural activity distinguished between the language network and the MD network. The language network, particularly in the LH, showed relatively little individual variation in activity (high ISCs) due to close tracking of the story (high WSCs). In contrast, MD network activity was more idiosyncratic across individuals (low ISCs), showing weaker tracking of the story (low WSCs). These findings suggest a novel typology of mental processes contributing to language comprehension: it is not only a question of which linguistic features are tracked by different mechanisms, but of whether (and to what extent) these mechanisms track linguistic input. Thus, some processes implemented in the language network are stimulus-related and consistent across individuals; other processes, implemented in the MD network, are less tightly coupled to the input and appear more idiosyncratic.

This distinction importantly constrains cognitive models of language processing: it narrows the space of domain-general processes that can be implemented in the MD network to those processes that do not require continuous access to the input. This conclusion is inconsistent with the assumption of close input tracking, which implicitly underlies existing interpretations of MD network activity in task-based neuroimaging studies of comprehension. It might also be inconsistent with current psycholinguistic models describing how domain-general working-memory resources contribute to incremental, moment-to-moment language processing along with language-specific knowledge (for a review, see: Levy, 2013).

Characterizing the respective contributions of the language and MD networks to comprehension was methodologically possible due to the localization of these networks using functional contrasts, individually for each participant. This method accounts for inter-individual variability in the mapping of function onto cortical anatomy (Jones and Powell, 1970; Gloor, 1997; Amunts et al., 1999; Tomaiuolo et al., 1999; Wise et al., 2001; Chein et al., 2002), conferring high functional resolution (Nieto-Castañón and Fedorenko, 2012) that is unobtainable if regions of interest are instead defined based on anatomical criteria or group analyses of functional data (Juch et al., 2005; Poldrack, 2006; Saxe et al., 2006; Fischl et al., 2008; Tahmasebi et al., 2011; Frost and Goebel, 2012). Consequently, single-participant functional localization provides a principled way of relating our ISC data to known functional divisions in the cortex. This method thus augments the ISC approach, allowing us to provide a novel key characterization of the functional topology of ISCs based on the distinction between the language and MD networks.

Within this topology, the role of MD regions in language comprehension is particularly interesting. Whereas task-based studies have demonstrated that MD regions scale their activity with increasing comprehension difficulty in numerous contexts (Stromswold et al., 1996; Stowe et al., 1998; Caplan et al., 1999; Fiez et al., 1999; Fiebach et al., 2002; Chee et al., 2003; Constable et al., 2004; Rodd et al., 2005; Chen et al., 2006; Nakic et al., 2006; Nieuwland et al., 2007; Novais-Santos et al., 2007; Hauk et al., 2008; Yarkoni et al., 2008; Carreiras et al., 2009; January et al., 2009; Peelle et al., 2009; Ye and Zhou, 2009; Barde et al., 2012; McMillan et al., 2012; McMillan et al., 2013), we demonstrate that they track natural language relatively weakly. One might suggest that the domain-general operations of the MD network are only recruited when linguistic labor is sufficiently high and burdens the language network beyond its capacities; as long as this threshold is not crossed, the working-memory resources that aid in comprehension might be domain-specific and implemented in the language network. However, we find such an interpretation unlikely given that our story stimuli contain frequent occurrences of challenging linguistic phenomena that correlate with behavioral measures of online comprehension difficulty (Shain et al., 2016).

Our finding that the MD network tracks linguistic stimuli relatively weakly also appears to disagree with prior evidence that this network tracks other naturalistic stimuli that are not purely linguistic. Specifically, in audiovisual movies, experiential features like "suspense" modulate MD activity similarly across individuals (Naci et al., 2014), possibly by influencing the frequency of attentional disengagement (Nakano et al., 2013). Does the domain-general MD network play a different role in language comprehension compared to its role in processing other naturalistic stimuli?

Perhaps MD regions are biased towards visual information (or audio-visual integration) in movies compared to the auditory information of stories (Michalka et al., 2015; Braga et al., 2016). Alternatively, MD regions may track both movies and stories, but fluctuations in MD activity during movie viewing could simply be slower, and thus more reliably measured, compared to the fast fluctuations during story comprehension. Therefore, evidence of stimulus tracking by MD regions during story comprehension might only be evident at high frequencies that cannot be measured with the temporally slow BOLD signal of fMRI. Still, we note that the temporal resolution of fMRI was sufficient to capture story tracking in language regions, so the argument above only holds if the MD network tracks stories on a faster time-scale than the language network. Finally, activity in MD regions may reflect internal fluctuations in domain-general attention or "focus" (Norman and Shallice, 1986; Chun et al., 2011) that may co-vary with the emotional manipulations in movies (Williams et al., 2016) but be relatively independent of input processing difficulty during natural language comprehension. This account is also consistent with previous findings of greater MD activity with increased linguistic demands in experimentally designed tasks, insofar as such tasks control the focus of participants more explicitly than naturalistic stories.

To conclude, our study synergistically combines task-based functional localization in individual participants and a naturalistic cognition paradigm for comparing brain activity across participants to characterize the distinct contributions of the language network and MD network to story comprehension. Whereas activity in the language network is similar across individuals and closely tracks stories, activity in the MD network is more idiosyncratic and does not track linguistic input as closely. These findings suggest a novel distinction between different mechanisms that underlie language processing based on individual differences in their processing patterns and their coupling to the linguistic input.

Acknowledgements

We would like to acknowledge the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT, and its support team (Steve Shannon, Atsushi Takahashi, and Sheeba Arnold). We thank MIT affiliates Alexander Paunov and

Zach Mineroff for their help with data collection, Anastasia Vishnevetsky for her help with constructing the stories and Nancy Kanwisher and Ted Gibson for recording the stories. We thank Uri Hasson (Princeton University) for providing the materials for the second replication. We also thank Nancy Kanwisher, Ted Gibson, John Duncan (Medical Research Council, Cognition and Brain Science Unit), and the audience at the CUNY Sentence Processing conference in San Diego for comments on earlier versions of this work. E.F. was supported by a K99/R00 award HD 057522 from NICHD and by a grant from the Simons Foundation to the Simons Center for the Social Brain at MIT. This work was also partially supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via Air Force Research Laboratory (AFRL), under contract FA8650-14-C-7358. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, AFRL, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

References

Amici S, Ogar J, Brambati SM, Miller BL, Neuhaus J, Dronkers NL, Gorno-Tempini ML (2007) Performance in specific language tasks correlates with regional volume changes in progressive aphasia. Cognitive and Behavioral Neurology 20:203-211.

Amunts K, Schleicher A, Bürgel U, Mohlberg H, Uylings H, Zilles K (1999) Broca's region revisited: Cytoarchitecture and intersubject variability. Journal of Comparative Neurology 412:319-341.

Barde LH, Yeatman JD, Lee ES, Glover G, Feldman HM (2012) Differences in neural activation between preterm and full term born adolescents on a sentence comprehension task: Implications for educational accommodations. Developmental cognitive neuroscience 2:S114-S128.

Barr DJ, Levy R, Scheepers C, Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language 68:255-278.

Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, Dronkers NF (2003) Voxel-based lesion–symptom mapping. Nature Neuroscience 6:448-450.

Bautista A, Wilson SM (2016) Neural responses to grammatically and lexically degraded speech. Language, Cognition and Neuroscience 31:567-574.

Bedny M, Pascual-Leone A, Dodell-Feder D, Fedorenko E, Saxe R (2011) Language processing in the occipital cortex of congenitally blind adults. Proceedings of the National Academy of Sciences 108:4429-4434.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Annals of statistics:1165-1188.

Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T (1997) Human brain language areas identified by functional magnetic resonance imaging. The Journal of Neuroscience 17:353-362.

Blank I, Balewski Z, Mahowald K, Fedorenko E (2016) Syntactic processing is distributed across the language system. Neuroimage 127:307-323.

Blank IA, Kanwisher N, Fedorenko E (2014) A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. Journal of Neurophysiology 112:1105-1118.
Braga RM, Hellyer PJ, Wise RJ, Leech R (2016) Auditory and visual connectivity gradients in frontoparietal cortex. Human Brain Mapping.
Braver TS, Reynolds JR, Donaldson DI (2003) Neural mechanisms of transient and sustained cognitive control during task switching. Neuron 39:713-726.
Braze D, Mencl WE, Tabor W, Pugh KR, Constable RT, Fulbright RK, Magnuson JS, Van Dyke JA, Shankweiler DP (2011) Unification of sentence processing via ear and eye: An fMRI study. Cortex 47:416-431.
Broca P (1861/2006) Comments regarding the seat of the faculty of spoken language, followed by an observation of aphemia (loss of speech). In: Broca's Region (Grodzinsky Y, Amunts K, eds), pp 291-304. New York, NY: Oxford University Press.
Buchweitz A, Mason RA, Tomitch L, Just MA (2009) Brain activation for reading and listening comprehension: An fMRI study of modality effects and individual differences in language comprehension. Psychology & neuroscience 2:111.
Caplan D, Alpert N, Waters G (1999) PET studies of syntactic processing with auditory sentence presentation. NeuroImage 9:343-351.
Carreiras M, Riba J, Vergara M, Heldmann M, Münte TF (2009) Syllable congruency and word frequency effects on brain activation. Human Brain Mapping 30:3079-3088.
Chee MW, O'Craven KM, Bergida R, Rosen BR, Savoy RL (1999) Auditory and visual word processing studied with fMRI. Human brain mapping 7:15-28.
Chee MW, Westphal C, Goh J, Graham S, Song AW (2003) Word frequency and subsequent memory effects studied using event-related fMRI. NeuroImage 20:1042-1051.
Chein J, Fissell K, Jacobs S, Fiez JA (2002) Functional heterogeneity within Broca's area during verbal working memory. Physiology & Behavior 77:635-639.
Chen E, West WC, Waters G, Caplan D (2006) Determinants of BOLD signal correlates of processing object-extracted relative clauses. Cortex 42:591-604.
Chun MM, Golomb JD, Turk-Browne NB (2011) A taxonomy of external and internal attention. Annual review of psychology 62:73-101.
Cole MW, Schneider W (2007) The cognitive control network: Integrated cortical regions with dissociable functions. NeuroImage 37:343-360.
Collins MJ (1996) A new statistical parser based on bigram lexical dependencies. In: Proceedings of the 34th Annual Meeting on Association for Computational Linguistics, pp 184-191. Stroudsburg, PA: Association for Computational Linguistics.
Constable RT, Pugh KR, Berroya E, Mencl WE, Westerveld M, Ni W, Shankweiler D (2004) Sentence complexity and input modality effects in sentence comprehension: an fMRI study. Neuroimage 22:11-21.
Cordes D, Haughton VM, Arfanakis K, Carew JD, Turski PA, Moritz CH, Quigley MA, Meyerand ME (2001) Frequencies contributing to functional connectivity in the cerebral cortex in “resting-state” data. American Journal of Neuroradiology 22:1326-1333.
D’Esposito M, Postle BR (2015) The cognitive neuroscience of working memory. Annual Review of Psychology 66:115-142.
Dax M (1863) Observations tendant à prouver la coïncidence constante des dérangements de la parole avec une lésion de l'hémisphère gauche du cerveau. Comptes rendus de l'Académie des Sciences 56:536.
Dosenbach NU, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, Burgund ED, Grimes AL, Schlaggar BL, Petersen SE (2006) A core system for the implementation of task sets. Neuron 50:799-812.

Dosenbach NU, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RA, Fox MD, Snyder AZ, Vincent JL, Raichle ME (2007) Distinct brain networks for adaptive and stable task control in humans. Proceedings of the National Academy of Sciences 104:11073-11078.
Duncan J (2010) The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. Trends in Cognitive Sciences 14:172-179.
Duncan J, Owen AM (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. Trends in Neurosciences 23:475-483.
Fedorenko E (2014) The role of domain-general cognitive control in language comprehension. Frontiers in Psychology 5:335.
Fedorenko E, Kanwisher N (2009) Neuroimaging of language: Why hasn't a clearer picture emerged? Language and Linguistics Compass 3:839-865.
Fedorenko E, Varley R (2016) Language and thought are not the same thing: Evidence from neuroimaging and neurological patients. Annals of the New York Academy of Sciences 1369:132-153.
Fedorenko E, Behr MK, Kanwisher N (2011) Functional specificity for high-level linguistic processing in the human brain. Proceedings of the National Academy of Sciences 108:16428-16433.
Fedorenko E, Duncan J, Kanwisher N (2012a) Language-selective and domain-general regions lie side by side within Broca's area. Current Biology 22:2059-2062.
Fedorenko E, Nieto-Castañón A, Kanwisher N (2012b) Lexical and syntactic representations in the brain: an fMRI investigation with multi-voxel pattern analyses. Neuropsychologia 50:499-513.
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.
Fedorenko E, Mineroff Z, Siegelman M, Blank I (2017) The distinction between lexico-semantic and syntactic processing is not an organizing dimension of the human language system. In: 30th CUNY Conference on Human Sentence Processing. Cambridge, MA.
Fedorenko E, Hsieh P-J, Nieto-Castañón A, Whitfield-Gabrieli S, Kanwisher N (2010) New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. Journal of Neurophysiology 104:1177-1194.
Fedorenko E, Scott TL, Brunner P, Coon WG, Pritchett B, Schalk G, Kanwisher N (2016) Neural correlate of the construction of sentence meaning. Proceedings of the National Academy of Sciences 113:E6256-E6262.
Ferreira F, Henderson JM, Anes MD, Weeks PA, McFarlane DK (1996) Effects of lexical frequency and syntactic complexity in spoken-language comprehension: Evidence from the auditory moving-window technique. Journal of Experimental Psychology: Learning, Memory, and Cognition 22:324-335.
Fiebach CJ, Friederici AD, Müller K, Von Cramon DY (2002) fMRI evidence for dual routes to the mental lexicon in visual word recognition. Journal of cognitive neuroscience 14:11-23.
Fiez JA, Balota DA, Raichle ME, Petersen SE (1999) Effects of lexicality, frequency, and spelling-to-sound consistency on the functional anatomy of reading. Neuron 24:205-218.
Fischl B, Rajendran N, Busa E, Augustinack J, Hinds O, Yeo BT, Mohlberg H, Amunts K, Zilles K (2008) Cortical folding patterns and predicting cytoarchitecture. Cerebral cortex 18:1973-1980.
Fridriksson J, Nettles C, Davis M, Morrow L, Montgomery A (2006) Functional communication and executive function in aphasia. Clinical Linguistics & Phonetics 20:401-410.
Friederici AD (2012) The cortical language circuit: From auditory perception to sentence comprehension. Trends in Cognitive Sciences 16:262-268.
Frost MA, Goebel R (2012) Measuring structural–functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. Neuroimage 59:1369-1381.

Fuster JM (1989) The prefrontal cortex: Anatomy, physiology and neuropsychology of the frontal lobe, 2 Edition. New York: Raven Press.
Futrell R, Mahowald K, Gibson E (2015) Large-scale evidence of dependency length minimization in 37 languages. Proceedings of the National Academy of Sciences 112:10336-10341.
Geranmayeh F, Brownsett SL, Wise RJ (2014) Task-induced brain activity in aphasic stroke patients: What is driving recovery? Brain 137:2632-2648.
Geschwind N (1970) The organization of language and the brain. Science 170:940-944.
Gill PMW (2007) Efficient calculation of p-values in linear-statistic permutation significance tests. Journal of Statistical Computation and Simulation 77:55-61.
Gloor P (1997) The temporal lobe and limbic system. New York, NY: Oxford University Press.
Hagoort P (2013) MUC (memory, unification, control) and beyond. Frontiers in Psychology 4:416.
Hasson U, Malach R, Heeger DJ (2010) Reliability of cortical activity during natural stimulation. Trends in Cognitive Sciences 14:40-48.
Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject synchronization of cortical activity during natural vision. science 303:1634-1640.
Hasson U, Ghazanfar AA, Galantucci B, Garrod S, Keysers C (2012) Brain-to-brain coupling: a mechanism for creating and sharing a social world. Trends in cognitive sciences 16:114-121.
Hasson U, Avidan G, Gelbard H, Vallines I, Harel M, Minshew N, Behrmann M (2009) Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real‐life viewing conditions. Autism Research 2:220-231.
Hauk O, Davis MH, Pulvermüller F (2008) Modulation of brain activity by multiple lexical and word form variables in visual word recognition: A parametric fMRI study. Neuroimage 42:1185-1195.
Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nature Reviews Neuroscience 8:393-402.
Honey CJ, Thompson CR, Lerner Y, Hasson U (2012) Not lost in translation: Neural responses shared across languages. The Journal of Neuroscience 32:15277-15283.
Hugdahl K, Raichle ME, Mitra A, Specht K (2015) On the existence of a generalized non-specific task-dependent network. Frontiers in Human Neuroscience 9:430.
January D, Trueswell JC, Thompson-Schill SL (2009) Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. Journal of Cognitive Neuroscience 21:2434-2444.
Jones E, Powell T (1970) An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. Brain: a journal of neurology 93:793.
Juch H, Zimine I, Seghier ML, Lazeyras F, Fasel JH (2005) Anatomical variability of the lateral frontal lobe surface: Implication for intersubject variability in language neuroimaging. NeuroImage 24:504-514.
Jung-Beeman M (2005) Bilateral brain processes for comprehending natural language. Trends in Cognitive Sciences 9:512-518.
Keller TA, Carpenter PA, Just MA (2001) The neural bases of sentence comprehension: A fMRI examination of syntactic and lexical processing. Cerebral cortex 11:223-237.
Lerner Y, Honey CJ, Silbert LJ, Hasson U (2011) Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. The Journal of Neuroscience 31:2906-2915.
Levy RP (2013) Memory and surprisal in human sentence comprehension. In: Sentence Processing, pp 78-114. Hove: Psychology Press.
Luria AR (1966/2012) Higher cortical functions in man: Springer Science & Business Media.
Mantini D, Corbetta M, Romani GL, Orban GA, Vanduffel W (2013) Evolutionarily novel functional networks in the human brain? The Journal of Neuroscience 33:3259-3275.

McMillan CT, Clark R, Gunawardena D, Ryant N, Grossman M (2012) fMRI evidence for strategic decision-making during resolution of pronoun reference. Neuropsychologia 50:674-687.
McMillan CT, Coleman D, Clark R, Liang T-W, Gross RG, Grossman M (2013) Converging evidence for the processing costs associated with ambiguous quantifier comprehension. Frontiers in psychology 4:153.
Menenti L, Gierhan SME, Segaert K, Hagoort P (2011) Shared language overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. Psychological Science 22:1173-1182.
Michalka SW, Kong L, Rosen ML, Shinn-Cunningham BG, Somers DC (2015) Short-term memory for space and time flexibly recruit complementary sensory-biased frontal lobe attention networks. Neuron 87:882-892.
Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annual Review of Neuroscience 24:167-202.
Murray LL (2012) Attention and other cognitive deficits in aphasia: Presence and relation to language and communication measures. American Journal of Speech-Language Pathology 21:S51-S64.
Naci L, Cusack R, Anello M, Owen AM (2014) A common neural code for similar conscious experiences in different individuals. Proceedings of the National Academy of Sciences 111:14277-14282.
Nakano T, Kato M, Morito Y, Itoi S, Kitazawa S (2013) Blink-related momentary activation of the default mode network while viewing videos. Proceedings of the National Academy of Sciences 110:702-706.
Nakic M, Smith BW, Busis S, Vythilingam M, Blair RJR (2006) The impact of affect and frequency on lexical decision: the role of the amygdala and inferior frontal cortex. Neuroimage 31:1752-1761.
Neville HJ, Bavelier D, Corina D, Rauschecker J, Karni A, Lalwani A, Braun A, Clark V, Jezzard P, Turner R (1998) Cerebral organization for language in deaf and hearing subjects: biological constraints and effects of experience. Proceedings of the National Academy of Sciences 95:922-929.
Nieto-Castañón A, Fedorenko E (2012) Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. NeuroImage 63:1646-1669.
Nieuwenhuis S, Forstmann BU, Wagenmakers E-J (2011) Erroneous analyses of interactions in neuroscience: A problem of significance. Nature Neuroscience 14:1105-1107.
Nieuwland MS, Petersson KM, Van Berkum JJ (2007) On sense and reference: Examining the functional neuroanatomy of referential processing. NeuroImage 37:993-1004.
Nomura EM, Gratton C, Visser RM, Kayser A, Perez F, D'Esposito M (2010) Double dissociation of two cognitive control networks in patients with focal brain lesions. Proceedings of the National Academy of Sciences 107:12017-12022.
Norman DA, Shallice T (1986) Attention to action. In: Consciousness and self-regulation, pp 1-18: Springer.
Novais-Santos S, Gee J, Shah M, Troiani V, Work M, Grossman M (2007) Resolving sentence ambiguity with planning and working memory resources: Evidence from fMRI. NeuroImage 37:361-378.
Novick JM, Trueswell JC, Thompson-Schill SL (2005) Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. Cognitive, Affective, & Behavioral Neuroscience 5:263-281.
Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97-113.

Peelle JE, Troiani V, Wingfield A, Grossman M (2009) Neural processing during older adults' comprehension of spoken sentences: age differences in resource allocation and connectivity. Cerebral Cortex:bhp142.
Poldrack RA (2006) Can cognitive processes be inferred from neuroimaging data? Trends in Cognitive Sciences 10:59-63.
Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL (2011) Functional network organization of the human brain. Neuron 72:665-678.
Regev M, Honey CJ, Simony E, Hasson U (2013) Selective and invariant neural responses to spoken and written narratives. The Journal of Neuroscience 33:15978-15988.
Rodd JM, Davis MH, Johnsrude IS (2005) The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. Cerebral Cortex 15:1261-1269.
Roland D, Dick F, Elman JL (2007) Frequency of basic English grammatical structures: A corpus analysis. Journal of Memory and Language 57:348-379.
Saxe R, Brett M, Kanwisher N (2006) Divide and conquer: A defense of functional localizers. NeuroImage 30:1088-1096.
Schmälzle R, Häcker FE, Honey CJ, Hasson U (2015) Engaged listeners: shared neural processing of powerful political speeches. Social, Cognitive and Affective Neuroscience:1137-1143.
Scott TL, Gallée J, Fedorenko E (2016) A new fun and robust version of an fMRI localizer for the frontotemporal language system. Cognitive Neuroscience.
Sebastian R, Laird AR, Kiran S (2011) Meta-analysis of the neural representation of first language and second language. Applied psycholinguistics 32:799.
Shain C, van Schijndel M, Futrell R, Gibson E, Schuler W (2016) Memory access during incremental sentence processing causes reading time latency. In: 26th Confernece on Computational Linguistics, p 49. Osaka, Japan.
Silbert LJ, Honey CJ, Simony E, Poeppel D, Hasson U (2014) Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. Proceedings of the National Academy of Sciences 111:E4687-E4696.
Silver NC, Dunlap WP (1987) Averaging correlation coefficients: Should Fisher's z transformation be used? Journal of Applied Psychology 72:146.
Sreenivasan KK, Curtis CE, D'Esposito M (2014) Revisiting the role of persistent neural activity during working memory. Trends in cognitive sciences 18:82-89.
Stowe LA, Broere CA, Paans AM, Wijers AA, Mulder G, Vaalburg W, Zwarts F (1998) Localizing components of a complex task: sentence processing and working memory. Neuroreport 9:2995-2999.
Stromswold K, Caplan D, Alpert N, Rauch S (1996) Localization of syntactic comprehension by positron emission tomography. Brain and language 52:452-473.
Tahmasebi AM, Davis MH, Wild CJ, Rodd JM, Hakyemez H, Abolmaesumi P, Johnsrude IS (2011) Is the link between anatomical structure and function equally strong at all cognitive levels of processing? Cerebral Cortex:bhr205.
Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD (1992) Testing for nonlinearity in time series: the method of surrogate data. Physica D: Nonlinear Phenomena 58:77-94.
Thompson-Schill SL, Bedny M, Goldberg RF (2005) The frontal lobes and the regulation of mental activity. Current Opinion in Neurobiology 15:219-224.
Tomaiuolo F, MacDonald JD, Caramanos Z, Posner G, Chiavaras M, Evans AC, Petrides M (1999) Morphology, morphometry and probability mapping of the pars opercularis of the inferior frontal gyrus: An in vivo MRI analysis. European Journal of Neuroscience 11:3033-3046.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15:273-289.
Vagharchakian L, Dehaene-Lambertz G, Pallier C, Dehaene S (2012) A temporal bottleneck in the language comprehension network. The Journal of Neuroscience 32:9089-9102.
Waters GS, Caplan D (2001) Age, working memory, and on-line syntactic processing in sentence comprehension. Psychology and Aging 16:128-144.
Wernicke C (1874/1969) The symptom complex of aphasia: A psychological study on an anatomical basis. In: Boston Studies in the Philosophy of Science (Cohen RS, Wartofsky MW, eds), pp 34-97. Dordrecht, The Netherlands: D. Reidel Publishing.
Whitfield-Gabrieli S, Nieto-Castanon A (2012) Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain Connectivity 2:125-141.
Wiener D, Tabor Connor L, Obler L (2004) Inhibition and auditory comprehension in Wernicke's aphasia. Aphasiology 18:599-609.
Willems RM, Van der Haegen L, Fisher SE, Francks C (2014) On the other hand: including left-handers in cognitive neuroscience and neurogenetics. Nature Reviews Neuroscience 15:193-201.
Williams J, Stönner C, Wicker J, Krauter N, Derstroff B, Bourtsoukidis E, Klüpfel T, Kramer S (2016) Cinema audiences reproducibly vary the chemical composition of air during films, by broadcasting scene specific emissions on breath. Scientific reports 6.
Wilson SM, Molnar-Szakacs I, Iacoboni M (2008) Beyond superior temporal cortex: Intersubject correlations in narrative speech comprehension. Cerebral Cortex 18:230-242.
Wise RJ, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA (2001) Separate neural subsystems withinWernicke's area'. Brain 124:83-95.
Woolgar A, Parr A, Cusack R, Thompson R, Nimmo-Smith I, Torralva T, Roca M, Antoun N, Manes F, Duncan J (2010) Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. Proceedings of the National Academy of Sciences 107:14899-14902.
Yarkoni T, Speer NK, Balota DA, McAvoy MP, Zacks JM (2008) Pictures of a thousand words: Investigating the neural mechanisms of reading with extremely rapid event-related fMRI. NeuroImage 42:973-987.
Ye Z, Zhou X (2009) Conflict control during sentence comprehension: fMRI evidence. NeuroImage 48:280-290.