Language-selective brain regions track linguistic input more closely than domain-general regions

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

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

[*] Correspondence: iblank@mit.edu

Summary (250 words)

Language comprehension engages a cortical network of left frontal and temporal regions [1-6]. Activity in this network is sensitive to linguistic features such as lexical information, syntax and compositional semantics [7-10]. However, this network shows virtually no engagement in non-linguistic tasks [11-14] and is therefore language-selective. In addition, language comprehension engages a second network consisting of frontal, parietal, cingulate, and insular regions [15-18]. Activity in this "Multiple Demand (MD)" network [19] is sensitive to comprehension difficulty, increasing in the presence of e.g. ambiguity [20-26], infrequent words [27-33] and non-local syntactic dependencies [34-40]. However, this network similarly scales its activity with cognitive effort across a wide range of non-linguistic tasks [19, 41] and is therefore domain-general. Given the functional dissociation between the language and MD networks [42, 43], 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 presupposed that both networks track linguistic input closely, and in a manner consistent across individuals. Here, we used fMRI to test this assumption by comparing the BOLD signal time-courses in each network across different individuals listening to the same story [44-46]. 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.

Results

Correlations in network activity across individuals listening to the same story

To characterize the functional contributions of the language and MD networks to comprehension, we tested how closely each network tracks naturalistic linguistic input (stories) and how similar such tracking is across individuals. Our interest in naturalistic cognition is three-fold: first, some brain regions respond more reliably to richly structured natural input compared to experimentally controlled input [47]. Second, task-free natural language processing plausibly differs from task-based processing, especially

given that MD regions operate in a task-dependent manner by biasing representations in other cortical networks in favor of task-relevant features [48-50]. Third, naturalistic stories require all aspects of the linguistic input to be combined into a single rich representation, unlike experimental stimuli that focus on particular linguistic features and have lower ecological validity.

Prior to the story comprehension scan, language and MD regions were functionally identified in each individual participant. Language regions were localized using a reading task that contrasted sentences with pronounceable nonwords (Figure 1a). We localized 8 left-hemispheric (LH) regions (Figure 2a) as well as 8 right-hemispheric (RH) homologues, which are also activated during some aspects of language processing [1, 6, 51-56] but might differ from LH regions in their contribution to comprehension, as suggested by neuroimaging [57] and neuropsychological [58, 59] data. MD regions were functionally identified using a spatial working-memory task that contrasted a hard version with an easy version (Fig 1b). We localized nine regions in each hemisphere (Figure 2b) and, based on prior findings [60-64], grouped them into two functionally distinct sub-networks: fronto-parietal (MDfp) and cingulo-opercular (MDco) (similar results were obtained when regions were instead grouped by hemisphere).

Each participant (n=19) then listened to 1-4 stories (duration: 270s-364s) constructed from publicly available texts, each followed by a comprehension test. To ensure that the stories strongly engaged the MD network, they were edited to include frequent occurrences of linguistic phenomena that increase processing difficulty and have been demonstrated to recruit this network (Figure 1c) (such phenomena are not naturally frequent enough; [65-67]). Following [46], we reasoned that if a given brain region tracked linguistic input with little individual differences then its activity time-course would be similar across participants and would thus show high Inter-Subject Correlations (ISCs) [68]. Therefore, we recorded the BOLD signal time-course in each language and MD region during each story and computed regional ISCs. To ensure that ISCs reflected tracking of linguistic information and not low-level sensory information, signals were first regressed against time-courses from the auditory cortex (similar results were obtained without regression).

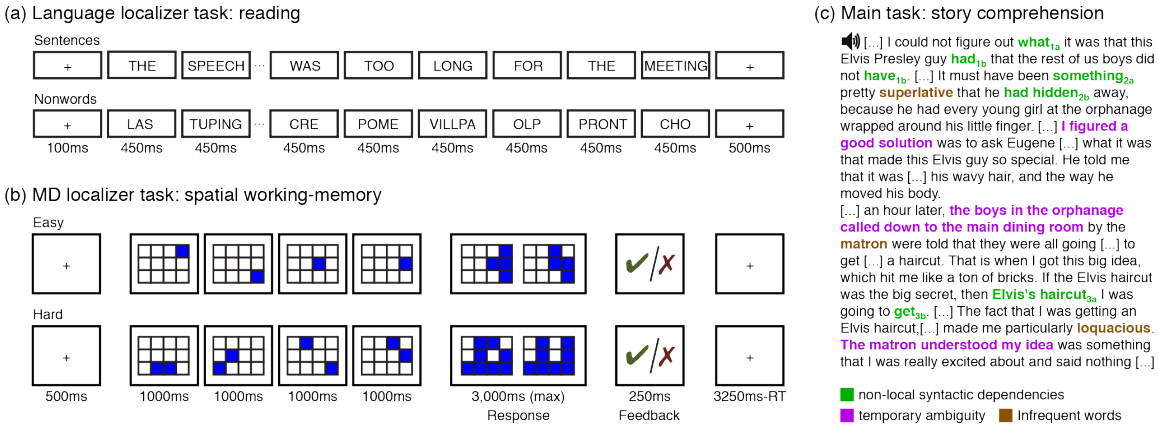

Figure 1. Experimental tasks. (a) The reading task used to localize language regions, based on the critical contrast sentences > nonwords. (b) The spatial working-memory task used to localize MD regions, based on the critical contrast hard > easy. (c) An excerpt from a story used in the main comprehension 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).

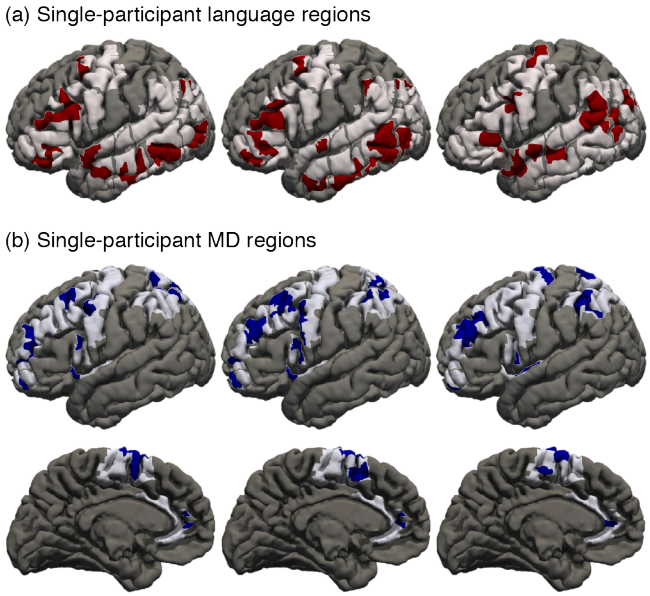

Figure 2. Functional regions of the language and MD networks. (a) LH language regions in 3 individual participants are shown in dark red. These regions were localized with a reading task (see Figure 1a). These regions were constrained to fall within eight broad areas where activations for this task are common across the population, shown in light pink. These areas were defined based on group-level data from a previous sample [1]. (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 [41].

We used linear, mixed-effect models to test whether the language and MD networks differed from each other in the degree of stimulus tracking, as estimated via their ISCs. Across stories, the LH language network showed the highest ISCs (Fisher transformed r=0.280), stronger than ISCs in the RH language network (r=0.210; Cohen's d=0.73, z=6.25, $p<10^{-9}$), the MDfp network (r=0.136; d=1.07, z=14.12, $p\approx0$) and the 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) (Figure 3; all p-values here and elsewhere are corrected for multiple comparisons using False-Discovery Rate (FDR) correction; [69]). The difference between the LH language network and the two MD networks was also observed for each story separately.

Next, we performed two replication experiments. In the first (n=13), we computed ISCs in a new participant group for two of the stories used above. In the second (n=19), we used a story recorded at a live event (duration: 420s; [45]) that did not undergo linguistic editing and was thus even more naturalistic than our previous stimuli. In both 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) (Figure 3).

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 people. 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 (defined anatomically) that track sensory input very closely [45]. 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 people; 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.

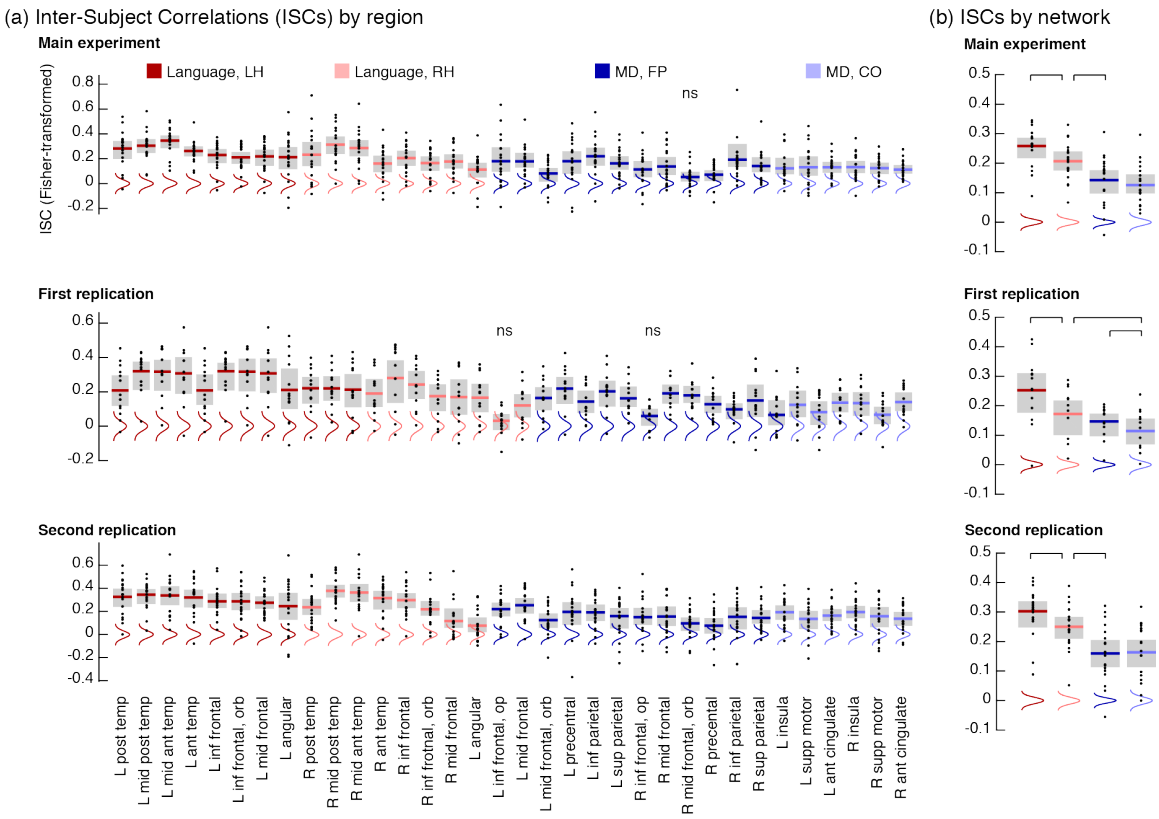

Figure 3. ISCs during story comprehension in the language and MD networks. (a) ISC (Fisher-transformed) for each brain region. Black dots are individual data points. Thick, colored horizontal lines show the average ISCs across participants. Gray rectangles show 95% confidence intervals of these average 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 [70]: 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 then computed Within-Subject Correlations (WSCs) for each network across the two instances. One group of participants (n=7) heard the stories repeatedly within the same scanning session (approximately one hour apart); another group (n=8) heard the stories in two sessions that were 6.5-21.5 months apart. These two groups did not differ from each other in their network WSCs, so their data were combined. 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).

These WSCs are lower than the ISCs reported above; this effect was expected because WSCs are measured by correlating noisy signals from two single trials, whereas ISCs are measured by correlating a signal from one participant with an average (i.e., noise reduced) signal across all other participants. To better compare WSCs and ISCs, we thus re-computed ISCs by correlating signal time-courses across pairs of individual participants (Figure 4b). Now, ISCs appeared weaker than WSCs (i.e., signals across participants were less similar than signals within a participant), but both measures patterned similarly in terms of between-network differences (for all comparisons between WSCs and ISCs, p>0.52). 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.

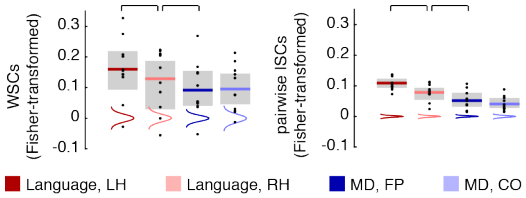

Figure 4. WSCs (left) and pairwise-ISCs (right) during story comprehension in the language and MD networks. Same conventions as in Figure 3.

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 differences 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: 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.

Critically, characterizing the respective contributions of the language and MD networks to comprehension was methodologically possible due to localization of these networks using functional contrasts, individually for each participant. First, identifying networks functionally allows us to tie our findings to a wealth of prior literature characterizing the response profiles of those networks. Second, our approach takes into account inter-individual variability in the mapping of function onto anatomy by comparing functional regions across participants even when those regions do not align well spatially. Such variability, evident in the temporal cortex [71-73] and especially in the frontal cortex [74, 75] (where language and MD regions lie side by side; [43]), renders anatomical localization precarious [76-79].

Indeed, pioneering studies of inter-subject correlations during language processing [44-46] computed ISCs for anatomical locations, assuming that the same location had a common function across participants. These studies revealed that broad cortical swaths show significant ISCs during comprehension, proposing a neural correlate of "shared understanding" across individuals [80] yet offering no principled way to relate those regions to known functional divisions in the cortex. This issue was further complicated because studies had not directly contrasted regions to each other, and had usually reported only *p*-values but not the sizes of the correlations. By augmenting the ISC methodology with a single-participant functional localization approach, the present study provides one key characterization of the functional topology of ISCs, distinguishing between 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 [20-40], we demonstrate that they track natural language relatively weakly even when it includes frequent occurrences of challenging linguistic features. Reconciling our data with past findings is thus challenging. Moreover, prior evidence suggests that MD regions track other naturalistic stimuli, such as audiovisual movies, with experiential features like "suspense" modulating MD activity similarly across individuals [81], possibly by influencing the frequency of attentional disengagement [82]. 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 [83, 84].

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. Finally, activity in MD regions may reflect internal fluctuations in domain-general attention or "focus" [85, 86] that may co-vary with the emotional manipulations in movies [87] 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.

Conclusion

Using a combination of task-based functional localization in individual participants and a naturalistic cognition paradigm for comparing brain activity across participants, we characterize 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 linguistic input as closely. These findings thus 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.

Experimental Procedures

The following methodological details have been previously reported (see Supplementary Materials): the design, materials and procedure for the language and MD localizer tasks [1, 41]; the stories used in the main experiments [42, 45]; data acquisition parameters [42]; spatial [1] and temporal [42] preprocessing streams; modeling of the localizer data [9]; and definition of language and MD regions [1, 41].

Participants

Forty-five participants (30 females) between the ages of 18 and 50, recruited from the MIT student body and the surrounding community, were paid for participation. 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).

ISCs and WSCs

For each participant and functional region, 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 we first temporally *z*-scored in each voxel and then averaged across voxels. Next, those signals were regressed against signals extracted from low-level auditory regions (defined anatomically around the postero-medial and antero-lateral sections of Heschl's gyrus bilaterally). Finally, for each

participant and region, we computed Pearson's moment correlation coefficient between the residual time-course and the corresponding average residual time-course across the remaining participants [45].

For each participant who listened to the same story on two occasions, we correlated the residual time-course in each region across the two trials. Because these WSCs are based on two single-trial signals, we also re-computed ISCs in a comparable manner; namely, for each participant and region, we correlated the residual time-course with the corresponding, individual residual time-course of each of the other participants, and averaged the resulting values.

ISCs/WSCs were Fisher-transformed prior to averaging and statistical testing in order to improve normality.

Statistical tests

In each region, ISCs/WSCs were 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 [88]. Individual distributions were each fit with a Gaussian and the resulting parameters were analytically combined across participants. The original ISCs/WSCs, also averaged across participants, were then *z*-scored relative to these parameters and converted to one-tailed *p*-values.

ISCs/WSCs were compared across networks using a linear, mixed-effects regression [89] implemented with the "lme4" package in R. In each experiment, ISCs/WSCs across all brain regions, participants and stories were modeled with a fixed effect of region and random intercepts for participant and story. The fixed effect estimates were combined across regions within each functional network (LH language, RH language, MDfp and MDco) and were pairwise compared to each other using the "multcomp" package in R. Hypotheses were two-tailed for the first experiment and one-tailed afterwards. For more information, see Supplementary Materials.

In each experiment, *p*-values are reported following False Discovery Rate (FDR) correction for multiple comparisons [69].

Acknowledgements

We thank Alexander Paunov and Zach Mineroff for their help with data collection, Anastasia Vishnevetsky for her help with constructing the stories, Nancy Kanwisher and Ted Gibson for recording the stories, and Uri Hasson for providing the story for the second replication. We also thank Nancy Kanwisher, Ted Gibson, John Duncan, 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.

Supplementary Materials

Language localizer task

Regions in the language network were localized using a reading task contrasting sentences and lists of unconnected, pronounceable nonwords (ref), in a standard blocked design with a counterbalanced order across runs (for timing parameters, see Table S1).

Stimuli were presented one word / nonword at a time (see Figure 1). 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. For half of these participants, the localizer included an additional condition of unconnected word lists, for purposes of another experiment. The remaining 35 participants instead read the materials passively (we included a button-pressing event at the end of each trial, to help these participants remain alert and focused). Note that in the former version nonwords are more engaging than sentences because their memorization is harder, whereas in the latter version sentences are more engaging than nonwords because they are meaningful. Importantly, this localizer has been shown to generalize across such manipulations, as the language network robustly and reliably shows a sentences > nonwords effect regardless of the task [1]. This localizer also generalizes across both visual and auditory presentations [90-92].

MD localizer task

Regions in the MD network were localized with a spatial working-memory game [11] contrasting a hard version with an easy version. On each trial (8s), participants saw a 3x4 grid and kept track of eight (hard version) or four (easy version) locations that were sequentially flashed two at a time or one at a time, respectively (1s per flash, 4s total). Then, participants indicated their memory for these locations in a 2-Alternative, Forced-Choice (2AFC) paradigm via a button press (3s total). Feedback was immediately provided upon choice (or lack thereof) (250ms). Trials began and ended with brief fixations (500ms and 250ms, respectively). 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 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, namely, the 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 [41].

Story comprehension task

Each subject listened to 1-4 stories over scanner-safe headphones (Sensimetrics, Malden, MA). In the main experiment and the first replication, stories were constructed based on publicly available fairy tales and short stories. These stories were edited to include a variety of linguistic phenomena that have been shown to increase processing difficulty and recruit the MD network, but do not occur with sufficiently high frequency in natural texts (see main text; for a sample text, see Appendix 1). 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). Each story started an ended with 16s seconds of music and fixation that were not analyzed.

After each story, participants answered 6-12 comprehension questions that required attentive listening (i.e., could not have been answered correctly based on common sense). For the main experiment and the first replication, participants answered 2AFC questions via a button press while in the scanner. For the second replication, participants filled in a 4AFC questionnaire once they got outside the scanner. For eight

participants, answers to these questions were not collected. The remaining 37 participants demonstrated very good comprehension of the stories, with a negatively skewed accuracy distribution (mode=100%, median=87.5%, semi-interquartile range=12.85%).

Data acquisition and preprocessing

Structural and functional data were collected on a whole-body 3 Tesla Siemens Trio scanner with a 32-channel head coil at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT. 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 4.4mm 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.

Data preprocessing was carried out with SPM5 and 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, normalization, resampling into 2mm isotropic voxels, smoothing with a 4mm FWHM Gaussian filter and high-pass filtering at 200s.

Additional temporal preprocessing of data from the story comprehension runs was carried out using the CONN toolbox [93] 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 [94].

Modeling localizer data

For each localizer task, 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: sentences > nonwords for the language localizer, and hard > easy for the MD localizer.

Defining participant-specific language and MD regions

Language and MD regions 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, regions in the language network were defined using group-constrained, participant-specific localization [1]. 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 LH, 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. These masks were mirror-projected onto the RH to create 8 homologous masks (the masks cover significant parts of the cortex, so their mirrored version is likely to encompass the RH homologue of the LH language network, despite possible hemispheric asymmetries in their precise locations). In each of the resulting 16 masks, a participant-specific language region was defined as the top 10% of voxels with the highest contrast values. This top *n*% approach ensures that functional regions can be defined in every participant and that their sizes are the same across participants, allowing for generalizable results [95].

Regions in the MD network were similarly defined based on the hard > easy contrast in the spatial working-memory game. Here, instead of using binary masks based on group-level data, we used anatomical masks ([96]; see [41, 42]). 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 partieal lobe, the insula, and supplementary motor area and the cingulate cortex. The first five masks constitute the fronto-parietal MD sub-network, and the last three constitute the cingulo-opercular sub-network.

Table S1. 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
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

Statistical tests

Statistical tests on WSC data were run on a sample including both participants who listened to the same story twice within the same scanning session and those who listened to the same story across two sessions. 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 WSCs with a fixed effect of the interaction between brain region 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.

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.

For all findings based on linear, mixed-effects regression analyses, similar results were obtained when ISCs/WSCs for each participant were first averaged across regions within each network and pairwise network comparisons (across participants) were then tested using exact permutation tests [97]. Therefore, our results are independent of assumptions regarding data normality.

Appendix 1: A sample story and comprehension questions

At ten years old, I could not figure out what it was that this Elvis Presley guy had that the rest of us boys did not have. He seemed to be no different from the rest of us. He was simply a man who had a head, two arms and two legs. It must have been something pretty superlative that he had hidden away, because he had every young girl at the orphanage wrapped around his little finger.

At about nine o'clock on Saturday morning, I figured a good solution was to ask Eugene Correthers, who was one of the older and smarter boys, what it was that made this Elvis guy so special. He told me that it was not anything about Elvis's personality, but his wavy hair, and the way he moved his body. About a half an hour later, the boys in the orphanage called down to the main dining room by the matron were told that they were all going to downtown Jacksonville, Florida to get a new pair of Buster Brown shoes and a haircut. That is when I got this big idea, which hit me like a ton of bricks. If the Elvis haircut was the big secret, then Elvis's haircut I was going to get.

I was going to have my day in the sun, and all the way to town that was all I talked about. The fact that I was getting an Elvis haircut, not just the simple fact that we were getting out of the orphanage, made me particularly loquacious. I told everybody, including the orphanage matron I normally feared, that I was going to look just like Elvis Presley and that I would learn to move around just like he did and that I would be rich and famous one day, just like him. The matron understood my idea was something that I was really excited about and said nothing.

When I got my new Buster Brown shoes, I was smiling from ear to ear. Those shoes, they shined really brightly, and I liked looking at the bones in my feet, which I had

never seen before, through a special x-ray machine they had in the shoe store that made the bones in your feet look green. I was now almost ready to go back to the orphanage and practice being like the man who all the girls loved, since I had my new Buster Brown shoes. It was the new haircut, though, that I needed to complete my new look.

We finally arrived at the unassuming, unembellished barbershop, where they cut our hair for free because we were orphans. Even though we were supposed to slowly wait to be called, I ran straight up to one of the barber chairs and climbed up onto the board the barber placed across the arms to make me sit up higher. I looked at the man and said, with a beaming smile on my face, "I want an Elvis haircut. Can you make my hair like Elvis's?" I asked. The barber, who was a genial young man, grinned back at me and said that he would try his best.

I was so happy when he started to cut my hair, but just as he started to cut, the matron, who had been watching me and had a look as cold as ice, motioned for him to come over to where she was standing. She whispered something into his ear that caused the barber to shake his head, like he was telling her, "No". In response, the matron walked over to a little man sitting in an office chair that squeaked as it rolled around the floor and spoke to him. It was the little man who then walked over and said something to the man who was cutting my hair. The next thing I knew, the man who was cutting my hair told me that he was no longer allowed to give me an Elvis cut.

"Why not?" I cried desperately.

The kindly barber stopped by the matron did not answer, but from his expression, I could tell that he wished he could cut it as I had asked.

Within a few minutes, it wasn't an Elvis haircut, but a short buzz cut that the barber had given me. When he finished shaving off all my hair and made me smell real good with his powder, the barber handed me a nickel and told me to go outside to the snack machine and buy myself a candy bar. I handed him the nickel back and told him that I was not hungry. "I'm so sorry, baby," he said, as I climbed out of his barber chair. "I am not a baby," I said, as I wiped the tears from my eyes.

I then sat down on the floor and brushed away the hair that had accumulated on my shiny new Buster Brown shoes. My head was no longer in the clouds, and I got up off the floor, brushed off my short pants, and walked sullenly towards the door.

The matron was smiling at me sort of funny like.

The barber upset by the matron said to her, "You are just a damn bitch, lady."

She yelled back at him at the top of her lungs, before walking toward the office, as fast as she could.

To show his anger, the man hit the wall with his hand and then walked outside where he stood against the brick wall, smoking a cigarette. I understood right there my haircut was something that had been out of the power of the barber and then I slowly walked outside to join the man. He looked down, smiled at me, then he patted me on the top of my bald as a coot head. It was a fact of my life that I was not gonna have hair that was anything like Elvis's anytime soon. I then looked up at the barber with my wet red eyes and asked, "Do you know if Elvis Presley has green bones?"

1. Why was the boy interested in Elvis?
   A. <u>Girls at the orphanage liked Elvis</u>
   B. Elvis had a lot of money

2. What made Elvis special, in the opinion of Eugene Correthers?

A. Elvis's personality
B. Wavy hair

3. On the bus, how did the boy behave?
   A. Talked a lot
   B. Was very quiet

4. What was the barber's initial reaction to the boy's request?
   A. Said he couldn't do it
   B. Said he'd try his best

5. What did the barber hand to the boy when he finished the haircut?
   A. A candy
   B. A nickel

6. What did the barber do to show his anger?
   A. Pushed the matron aside
   B. Hit the wall

References

1. Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., and Kanwisher, N. (2010). New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. Journal of Neurophysiology *104*, 1177-1194.
2. Menenti, L., Gierhan, S.M., Segaert, K., and 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.
3. Bates, E., Wilson, S.M., Saygin, A.P., Dick, F., Sereno, M.I., Knight, R.T., and Dronkers, N.F. (2003). Voxel-based lesion–symptom mapping. Nature neuroscience *6*, 448-450.
4. Mesulam, M.-M., Thompson, C.K., Weintraub, S., and Rogalski, E.J. (2015). The Wernicke conundrum and the anatomy of language comprehension in primary progressive aphasia. Brain *138*, 2423-2437.
5. Mirman, D., Chen, Q., Zhang, Y., Wang, Z., Faseyitan, O.K., Coslett, H.B., and Schwartz, M.F. (2015). Neural organization of spoken language revealed by lesion-symptom mapping. Nature communications *6*.
6. Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. Trends in Cognitive Sciences *9*, 512-518.
7. Keller, T.A., Carpenter, P.A., and Just, M.A. (2001). The neural bases of sentence comprehension: A fMRI examination of syntactic and lexical processing. Cerebral cortex *11*, 223-237.
8. Fedorenko, E., Nieto-Castañón, A., and Kanwisher, N. (2012). Lexical and syntactic representations in the brain: an fMRI investigation with multi-voxel pattern analyses. Neuropsychologia *50*, 499-513.
9. Blank, I., Balewski, Z., Mahowald, K., and Fedorenko, E. (2016). Syntactic processing is distributed across the language system. NeuroImage *127*, 307-323.
10. Bautista, A., and Wilson, S.M. (2016). Neural responses to grammatically and lexically degraded speech. Language, Cognition and Neuroscience *31*, 567-574.
11. Fedorenko, E., Behr, M.K., and Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. Proceedings of the National Academy of Sciences *108*, 16428-16433.
12. Monti, M.M., Parsons, L.M., and Osherson, D.N. (2009). The boundaries of language and thought in deductive inference. Proceedings of the National Academy of Sciences *106*, 12554-12559.
13. Monti, M.M., Parsons, L.M., and Osherson, D.N. (2012). Thought beyond language neural dissociation of algebra and natural language. Psychological Science, 0956797612437427.
14. Monti, M.M., and Osherson, D.N. (2012). Logic, language and the brain. Brain research *1428*, 33-42.

15. Bedny, M., McGill, M., and Thompson-Schill, S.L. (2008). Semantic adaptation and competition during word comprehension. Cerebral Cortex *18*, 2574-2585.
16. Hindy, N.C., Altmann, G.T., Kalenik, E., and Thompson-Schill, S.L. (2012). The effect of object state-changes on event processing: do objects compete with themselves? The journal of neuroscience *32*, 5795-5803.
17. Hindy, N.C., Solomon, S.H., Altmann, G.T., and Thompson-Schill, S.L. (2015). A cortical network for the encoding of object change. Cerebral Cortex *25*, 884-894.
18. Thompson-Schill, S.L., Bedny, M., and Goldberg, R.F. (2005). The frontal lobes and the regulation of mental activity. Current opinion in neurobiology *15*, 219-224.
19. Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. Trends in Cognitive Sciences *14*, 172-179.
20. Rodd, J.M., Davis, M.H., and Johnsrude, I.S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. Cerebral Cortex *15*, 1261-1269.
21. Novais-Santos, S., Gee, J., Shah, M., Troiani, V., Work, M., and Grossman, M. (2007). Resolving sentence ambiguity with planning and working memory resources: Evidence from fMRI. NeuroImage *37*, 361-378.
22. Nieuwland, M.S., Petersson, K.M., and Van Berkum, J.J. (2007). On sense and reference: Examining the functional neuroanatomy of referential processing. NeuroImage *37*, 993-1004.
23. Ye, Z., and Zhou, X. (2009). Conflict control during sentence comprehension: fMRI evidence. NeuroImage *48*, 280-290.
24. January, D., Trueswell, J.C., and Thompson-Schill, S.L. (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.
25. McMillan, C.T., Clark, R., Gunawardena, D., Ryant, N., and Grossman, M. (2012). fMRI evidence for strategic decision-making during resolution of pronoun reference. Neuropsychologia *50*, 674-687.
26. McMillan, C.T., Coleman, D., Clark, R., Liang, T.-W., Gross, R.G., and Grossman, M. (2013). Converging evidence for the processing costs associated with ambiguous quantifier comprehension. Frontiers in psychology *4*, 153.
27. Fiez, J.A., Balota, D.A., Raichle, M.E., and Petersen, S.E. (1999). Effects of lexicality, frequency, and spelling-to-sound consistency on the functional anatomy of reading. Neuron *24*, 205-218.
28. Fiebach, C.J., Friederici, A.D., Müller, K., and Von Cramon, D.Y. (2002). fMRI evidence for dual routes to the mental lexicon in visual word recognition. Journal of cognitive neuroscience *14*, 11-23.
29. Chee, M.W., Westphal, C., Goh, J., Graham, S., and Song, A.W. (2003). Word frequency and subsequent memory effects studied using event-related fMRI. NeuroImage *20*, 1042-1051.
30. Nakic, M., Smith, B.W., Busis, S., Vythilingam, M., and Blair, R.J.R. (2006). The impact of affect and frequency on lexical decision: the role of the amygdala and inferior frontal cortex. Neuroimage *31*, 1752-1761.
31. Yarkoni, T., Speer, N.K., Balota, D.A., McAvoy, M.P., and Zacks, J.M. (2008). Pictures of a thousand words: Investigating the neural mechanisms of reading with extremely rapid event-related fMRI. NeuroImage *42*, 973-987.
32. Hauk, O., Davis, M.H., and 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.
33. Carreiras, M., Riba, J., Vergara, M., Heldmann, M., and Münte, T.F. (2009). Syllable congruency and word frequency effects on brain activation. Human Brain Mapping *30*, 3079-3088.
34. Stromswold, K., Caplan, D., Alpert, N., and Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. Brain and language *52*, 452-473.
35. Stowe, L.A., Broere, C.A., Paans, A.M., Wijers, A.A., Mulder, G., Vaalburg, W., and Zwarts, F. (1998). Localizing components of a complex task: sentence processing and working memory. Neuroreport *9*, 2995-2999.
36. Caplan, D., Alpert, N., and Waters, G. (1999). PET studies of syntactic processing with auditory sentence presentation. NeuroImage *9*, 343-351.

37. Constable, R.T., Pugh, K.R., Berroya, E., Mencl, W.E., Westerveld, M., Ni, W., and Shankweiler, D. (2004). Sentence complexity and input modality effects in sentence comprehension: an fMRI study. Neuroimage *22*, 11-21.
38. Chen, E., West, W.C., Waters, G., and Caplan, D. (2006). Determinants of BOLD signal correlates of processing object-extracted relative clauses. Cortex *42*, 591-604.
39. Peelle, J.E., Troiani, V., Wingfield, A., and Grossman, M. (2009). Neural processing during older adults' comprehension of spoken sentences: age differences in resource allocation and connectivity. Cerebral Cortex, bhp142.
40. Barde, L.H., Yeatman, J.D., Lee, E.S., Glover, G., and Feldman, H.M. (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.
41. Fedorenko, E., Duncan, J., and Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. Proceedings of the National Academy of Sciences *110*, 16616-16621.
42. Blank, I., Kanwisher, N., and 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.
43. Fedorenko, E., Duncan, J., and Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. Current Biology *22*, 2059-2062.
44. Wilson, S.M., Molnar-Szakacs, I., and Iacoboni, M. (2008). Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. Cerebral Cortex *18*, 230-242.
45. Lerner, Y., Honey, C.J., Silbert, L.J., and Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. The Journal of Neuroscience *31*, 2906-2915.
46. Honey, C.J., Thompson, C.R., Lerner, Y., and Hasson, U. (2012). Not lost in translation: neural responses shared across languages. The Journal of Neuroscience *32*, 15277-15283.
47. Hasson, U., Malach, R., and Heeger, D.J. (2010). Reliability of cortical activity during natural stimulation. Trends in Cognitive Sciences *14*, 40-48.
48. Miller, E.K., and Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. Annual review of neuroscience *24*, 167-202.
49. Sreenivasan, K.K., Curtis, C.E., and D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. Trends in cognitive sciences *18*, 82-89.
50. D'Esposito, M., and Postle, B.R. (2015). The cognitive neuroscience of working memory. Annual Review of Psychology *66*, 115-142.
51. George, M.S., Parekh, P.I., Rosinsky, N., Ketter, T.A., Kimbrell, T.A., Heilman, K.M., Herscovitch, P., and Post, R.M. (1996). Understanding emotional prosody activates right hemisphere regions. Archives of neurology *53*, 665-670.
52. Yang, J. (2014). The role of the right hemisphere in metaphor comprehension: A meta‐analysis of functional magnetic resonance imaging studies. Human Brain Mapping *35*, 107-122.
53. Bozic, M., Fonteneau, E., Su, L., and Marslen‐Wilson, W.D. (2015). Grammatical analysis as a distributed neurobiological function. Human Brain Mapping *36*, 1190-1201.
54. Binder, J.R., Frost, J.A., Hammeke, T.A., Cox, R.W., Rao, S.M., and Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. The Journal of Neuroscience *17*, 353-362.
55. Chiarello, C., Banich, M., and Mack, M. (2003). Parallel systems for processing language: Hemispheric complementarity in the normal brain. Mind, Brain, and Language: Multidisciplinary Perspectives, 229-247.
56. Price, C.J. (2012). A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage *62*, 816-847.
57. Fedorenko, E., and Thompson-Schill, S.L. (2014). Reworking the language network. Trends in Cognitive Sciences *18*, 120-126.
58. Geschwind, N. (1970). The organization of language and the brain. In Science. (Citeseer).
59. Damasio, A.R. (1992). Aphasia. New England Journal of Medicine *326*, 531-539.
60. Dosenbach, N.U., Visscher, K.M., Palmer, E.D., Miezin, F.M., Wenger, K.K., Kang, H.C., Burgund, E.D., Grimes, A.L., Schlaggar, B.L., and Petersen, S.E. (2006). A core system for the implementation of task sets. Neuron *50*, 799-812.

61. Dosenbach, N.U., Fair, D.A., Miezin, F.M., Cohen, A.L., Wenger, K.K., Dosenbach, R.A., Fox, M.D., Snyder, A.Z., Vincent, J.L., and Raichle, M.E. (2007). Distinct brain networks for adaptive and stable task control in humans. Proceedings of the National Academy of Sciences *104*, 11073-11078.
62. Nomura, E.M., Gratton, C., Visser, R.M., Kayser, A., Perez, F., and 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.
63. Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., and Schlaggar, B.L. (2011). Functional network organization of the human brain. Neuron *72*, 665-678.
64. Mantini, D., Corbetta, M., Romani, G.L., Orban, G.A., and Vanduffel, W. (2013). Evolutionarily novel functional networks in the human brain? The Journal of Neuroscience *33*, 3259-3275.
65. Roland, D., Dick, F., and Elman, J.L. (2007). Frequency of basic English grammatical structures: A corpus analysis. Journal of Memory and Language *57*, 348-379.
66. Collins, M.J. (1996). A new statistical parser based on bigram lexical dependencies. In Proceedings of the 34th Annual Meeting on Association for Computational Linguistics. (Stroudsburg, PA: Association for Computational Linguistics), pp. 184-191.
67. Futrell, R., Mahowald, K., and Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. Proceedings of the National Academy of Sciences *112*, 10336-10341.
68. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. science *303*, 1634-1640.
69. Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Annals of statistics, 1165-1188.
70. Hasson, U., Avidan, G., Gelbard, H., Vallines, I., Harel, M., Minshew, N., and Behrmann, M. (2009). Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real‐life viewing conditions. Autism Research *2*, 220-231.
71. Jones, E., and Powell, T. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. Brain: a journal of neurology *93*, 793.
72. Gloor, P. (1997). The temporal lobe and limbic system, (New York, NY: Oxford University Press).
73. Wise, R.J., Scott, S.K., Blank, S.C., Mummery, C.J., Murphy, K., and Warburton, E.A. (2001). Separate neural subsystems withinWernicke's area'. Brain *124*, 83-95.
74. Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H., and Zilles, K. (1999). Broca's region revisited: Cytoarchitecture and intersubject variability. Journal of Comparative Neurology *412*, 319-341.
75. Chein, J., Fissell, K., Jacobs, S., and Fiez, J.A. (2002). Functional heterogeneity within Broca's area during verbal working memory. Physiology & Behavior *77*, 635-639.
76. Poldrack, R.A. (2006). Can cognitive processes be inferred from neuroimaging data? Trends in Cognitive Sciences *10*, 59-63.
77. Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B.T., Mohlberg, H., Amunts, K., and Zilles, K. (2008). Cortical folding patterns and predicting cytoarchitecture. Cerebral cortex *18*, 1973-1980.
78. Tahmasebi, A.M., Davis, M.H., Wild, C.J., Rodd, J.M., Hakyemez, H., Abolmaesumi, P., and Johnsrude, I.S. (2011). Is the link between anatomical structure and function equally strong at all cognitive levels of processing? Cerebral Cortex, bhr205.
79. Frost, M.A., and Goebel, R. (2012). Measuring structural–functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. Neuroimage *59*, 1369-1381.
80. Hasson, U., Ghazanfar, A.A., Galantucci, B., Garrod, S., and Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. Trends in cognitive sciences *16*, 114-121.
81. Naci, L., Cusack, R., Anello, M., and Owen, A.M. (2014). A common neural code for similar conscious experiences in different individuals. Proceedings of the National Academy of Sciences *111*, 14277-14282.

82. Nakano, T., Kato, M., Morito, Y., Itoi, S., and 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.
83. Michalka, S.W., Kong, L., Rosen, M.L., Shinn-Cunningham, B.G., and Somers, D.C. (2015). Short-term memory for space and time flexibly recruit complementary sensory-biased frontal lobe attention networks. Neuron *87*, 882-892.
84. Braga, R.M., Hellyer, P.J., Wise, R.J., and Leech, R. (2016). Auditory and visual connectivity gradients in frontoparietal cortex. Human Brain Mapping.
85. Norman, D.A., and Shallice, T. (1986). Attention to action. In Consciousness and self-regulation. (Springer), pp. 1-18.
86. Chun, M.M., Golomb, J.D., and Turk-Browne, N.B. (2011). A taxonomy of external and internal attention. Annual review of psychology *62*, 73-101.
87. Williams, J., Stönner, C., Wicker, J., Krauter, N., Derstroff, B., Bourtsoukidis, E., Klüpfel, T., and Kramer, S. (2016). Cinema audiences reproducibly vary the chemical composition of air during films, by broadcasting scene specific emissions on breath. Scientific reports *6*.
88. Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Farmer, J.D. (1992). Testing for nonlinearity in time series: the method of surrogate data. Physica D: Nonlinear Phenomena *58*, 77-94.
89. Barr, D.J., Levy, R., Scheepers, C., and Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language *68*, 255-278.
90. Braze, D., Mencl, W.E., Tabor, W., Pugh, K.R., Constable, R.T., Fulbright, R.K., Magnuson, J.S., Van Dyke, J.A., and Shankweiler, D.P. (2011). Unification of sentence processing via ear and eye: an fMRI study. cortex *47*, 416-431.
91. Vagharchakian, L., Dehaene-Lambertz, G., Pallier, C., and Dehaene, S. (2012). A temporal bottleneck in the language comprehension network. The Journal of Neuroscience *32*, 9089-9102.
92. Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. Frontiers in Psychology *5*, 335.
93. Whitfield-Gabrieli, S., and Nieto-Castanon, A. (2012). Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain connectivity *2*, 125-141.
94. Cordes, D., Haughton, V.M., Arfanakis, K., Carew, J.D., Turski, P.A., Moritz, C.H., Quigley, M.A., and Meyerand, M.E. (2001). Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. American Journal of Neuroradiology *22*, 1326-1333.
95. Nieto-Castañón, A., and Fedorenko, E. (2012). Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. NeuroImage *63*, 1646-1669.
96. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and 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.
97. Gill, P.M. (2007). Efficient calculation of p-values in linear-statistic permutation significance tests. Journal of statistical computation and simulation *77*, 55-61.