A positive role of task-deactivated default mode network in narrative speech comprehension

Lanfang Liu [1,2], Hehui Li[3], Zhiting Ren[4], Qi Zhou[2], Yuxuan Zhang[2], Chunming Lu[2], Jiang Qiu[4], Hong Chen*[4], Guosheng Ding*[2]

[1]Department of Psychology, Sun Yat-sen University, Guangzhou, 510006, P.R. China;

[2]Center for Brain Disorders and Cognitive Sciences, Shenzhen University, Shenzhen, 518060, PR China

[3]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University & IDG/McGovern Institute for Brain Research, Beijing, 100875, P.R. China;

[4]Key Laboratory of Cognition and Personality (SWU), Ministry of Education & Department of Psychology, Southwest University, Chongqing, 400715, P.R. China

Corresponding author: either Guosheng Ding, State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, No. 19 Xinjiekouwai Street, Beijing 100875, China, E-mail: dinggsh@bnu.edu.cn, or Hong Chen, School of psychology, SouthWest University, No. 2 Tiansheng Street, Beibei, Chongqing, 400715, E-mail: chenhg@swu.edu.cn

Acknowledgments: This work was supported by grants from the National Natural Science Foundation of China (NSFC: 3190082, 31971036), China Postdoctoral Science Foundation (2019M653248) and the Open Research Fund of the State Key Laboratory of Cognitive Neuroscience and Learning from Beijing Normal University

(CNLYB1803).

Data availability statement: The data and codes used in this study are available from the corresponding author upon reasonable request.

Ethics approval statement: Written informed consent was obtained from all participants under the protocol approved by the Reviewer Board of Southwest University.

Conflict of interest disclosure: No conflict of interest is declared.

Abstract

Psychological theories have implicated an active role of the default mode network (DMN) in natural speech comprehension. However, as listeners need to keep tracking the external audio streams, the DMN is regularly de-activated and anticorrelated with externally-oriented networks. Such a pattern has been interpreted as the suppression of the DMN to support externally-oriented cognitive processes. The current study aims to resolve this seeming contradiction. Brain activities from a speaker telling autobiographical stories and a group of participants (N = 62) listening to the recordings were collected with fMRI. By analyzing the listeners' brains alone, we found the DMN was deactivated during speech listening relative to a fixation period and anticorrelated with the task-positive perisylvian language network (pLN). Dynamic Causal Modeling showed the pLN had inhibitory influence on the DMN, whereas the DMN had excitatory influence on the pLN. Further between-brain analyses revealed the activities of DMN in the listener's brain were tightly coupled with the activities of the homologous network in the speaker's brain. Significant interbrain couplings were also observed in the pLN, but were weaker and faded quicker. Moreover, listeners showing stronger coupling responses to the speaker in the DMN understood the speech better, and tended to exhibit more positive DMN → pLN effective connections. We conclude that the DMN may occupy an internal system that works cooperatively with the externally-oriented pLN to support narrative speech comprehension.

Keywords: default mode network, speech comprehension, neural coupling, fMRI

1. Introduction

In neurolinguistic studies, verbal communication has been usually approached as a signal transmission process, and the main attention has been paid to the neural responses bound to and triggered by the occurrence of speech signals (Stolk et al., 2016b). Within this signal-centered framework, the crucial role of a set of frontal and temporal regions (i.e., the perisylvian language network, pLN) in coding and decoding the sensory-motor, syntactic and semantic properties of speech signals has been well recognized. Nevertheless, in real-life situations, the speech signals we receive are usually impoverished, degraded or ambiguous. Yet, we can still decipher the meaning intended by a speaker accurately. By contrast, current state-of-art artificial agents, such as Apple's Siri or Microsoft's Cortana, often make communicative errors despite they have mastered the signal decoding rules. These observations indicate that successful speech understanding demands more than signal decoding.

Indeed, psychological models have proposed that, in addition to the perceptual and linguistic analyses, to successfully understand the speaker's mind during verbal communication further requires the listener to integrate the signal's meaning with his/her internal conceptual space and align his/her mental state with the speaker's (Pickering & Garrod, 2006; Berwick et al., 2013; Wheatley et al., 2019). These internally-directed processes have been frequently linked to the default mode network (DMN) (Spreng et al., 2008; Spreng & Grady, 2009; Li et al., 2014), an intrinsically organized functional system composed of the posterior cingulate cortex, precuneus,

101 medial prefrontal cortex, and lateral parietal cortex (Marcus E. Raichle et al., 2001).

102 Despite the theoretical expectation, direct and convincing evidence for an active
103 role of the DMN in auditory language comprehension is still lacking. Instead, as
104 auditory language processing requires the participants to keep tracking external audio
105 streams, the DMN has been repeatedly found to be deactivated relative to low-level
106 baselines (Wilson et al., 2007; Szaflarski et al., 2012; Rodriguez Moreno et al., 2014;
107 Horowitz-Kraus et al., 2017; Cuevas et al., 2019), and anticorrelated with networks
108 dedicated to externally-oriented processes (including the pLN and attention network)
109 (Fox et al., 2005; Uddin et al., 2009; Smith et al., 2012). The task-induced deactivation
110 and network anticorrelation seem to imply that the DMN was "suppressed" to support
111 externally-oriented cognitive activities (Anticevic et al., 2012; Gauffin et al., 2013;
112 Zhou et al., 2016).

113 By measuring the correlation of blood-oxygen-level-dependent (BOLD) signal
114 fluctuations across subjects (i.e., intersubject correlation, ISC) rather than the
115 magnitude of task-induced activation, several studies found significant ISC in the DMN
116 regions during narrative speech comprehension (Wilson et al., 2007; Lerner et al., 2011;
117 Simony et al., 2016; Yeshurun et al., 2017), which was modulated by the coherence of
118 the stimulus' temporal structure over minutes of time (Lerner et al., 2011; Simony et
119 al., 2016) and the manipulation of story belief (Yeshurun et al., 2017). These results are
120 considered to reflect the involvement of DMN in semantic and conceptual processing.
121 However, there is a possibility that the ISC (and its change) in the DMN might be an
122 epiphenomenon of task engagement. It is well known that the activities of DMN regions

123 are systematically modulated by task difficulty (Zwaan & Radvansky, 1998;
124 McKiernan et al., 2003; Greicius & Menon, 2004; Humphreys et al., 2015). If subjects
125 find the same part of the narrative more or less engaging, the DMN will be down- or
126 up-regulated similarly across subjects, which would result in the correlation of brain
127 signal fluctuations across subjects (Wilson et al., 2007).

128 The discrepancy between the theoretical prediction and empirical findings
129 regarding the role of DMN in auditory language comprehension may stem from two
130 methodological issues. First, by focusing on task-evoked brain activation or
131 deactivation, prior work has primarily captured the reflexive activity of the brain.
132 However, it has been proposed that our brain activities are mainly intrinsic, involving
133 "information processing for interpreting, responding to and predicting environmental
134 demand" (M. E. Raichle, 2010). While analyzing task-evoked brain responses has been
135 highly productive in delineating the functional role of pLN in linguistic processing, this
136 approach may not accurately capture the functional property of DMN. Second, in prior
137 work speech comprehension has been mainly analyzed as an independent process from
138 speech production within the boundaries of individual brains. However, by recording
139 the brain activities from both speakers and listeners during natural verbal
140 communication (the "two-brain" approach), recent work has demonstrated that neural
141 activities across the listener and speaker are tightly coupled (Stephens et al., 2010; Jiang
142 et al., 2012; Silbert et al., 2014; L. Liu et al., 2020). The listener-speaker neural coupling
143 is postulated to reflect the alignment between interlocutors at multiple levels of
144 linguistic and extra-linguistic representations (Hasson & Frith, 2016; Schoot et al., 2016)

which forms the basis of mutual understanding (Pickering & Garrod, 2006; Hasson et al., 2012; Stolk et al., 2016a). Thus, analyzing single individuals in isolation may fail to achieve a complete picture of the functional role of DMN in speech understanding.

To better illuminate the role of DMN in speech comprehension, we scanned participants engaged in natural verbal communication and analyzed both response profiles within the listeners' brains and neural couplings between the listeners and the speaker. A group Independent Component Analysis (ICA) combined with a "dual-regression" technique was applied to detect the DMN and pLN in individual brains. Replicating previous findings, we found task-induced negative activation in the DMN whereas positive activation in the pLN, and anticorrelation between the dynamics of the two networks. However, further two-brain analyses revealed significant listener-speaker neural couplings in both the DMN and pLN, with longer response lag in the DMN than in the pLN. When communication failed or when the content of story was not matched between the speaker and listener, the between-brain coupling in both the DMN and pLN vanished. Moreover, the level of coupling strength in both the DMN and pLN could predict the degree to which the listeners understood and memorize the speech. These results provide reliable evidence showing that the DMN plays an active in language comprehension, which may contribute in a different way and support distinct computations from that of the pLN.

2. Materials and Methods

A schema depicting the experimental procedure and data analyses in this study is shown

in Figure 1.

2.1 Participants and experimental procedure

A total of 67 Chinese college students (aged 19-27 years) participated in this study, including one female speaker and 66 listeners (35 females). We also recruited a native Mongolian speaker (a female college student) to serve for a control condition. All participants were right-handed, and reported no history of physiological or mental disorder. None of these listeners had learned Mongolian. Written informed consent was obtained from all participants under a protocol approved by the Reviewer Board of Southwest University in China.

The experimental protocol was identical to that used in our previously published study (L. Liu et al., 2020). We therefore only briefly described it here. In the first step, we had the speakers told stories based on their personal experience (10 min for each) while undergoing fMRI scanning. A noise-canceling microphone (FOMRI-III, Optoacoustics Ltd., Or-Yehuda, Israel) was used to simultaneously record the speech. After further off-line denoising by Adobe Audition 3.0 (Adobe Systems Inc., USA), the audio recordings were played back to each of the 66 listeners during scanning. A common set of visual stimuli were presented to the speakers and listeners, consisting of a fixation cross lasting for 20s and then an icon of a horn lasting until the end of the scanning run. The horn was designed to prompt the participant to start speaking or listening immediately.

Each participant listened to one of the three stories told by the Chinese speaker. To

make the listeners attend to the stories, we informed them beforehand that a test about the content of the story would be given after the scanning. To control for the effect of low-level acoustic processing, we also played the story told by the Mongolian speaker to each listener. Before the speech listening/speaking task, all participants underwent a 10-min resting-state scan during which their eyes were closed.

2.3 Behavioral assessment

To assess the degree to which the listeners understood the speech, we conducted an interview with each listener at the end of the scan. In the interview, the listeners were required first to retell the story with as much detail as possible. Next, experimenters asked the listeners several questions concerning the part of the contents not present in their free recalls. Two independent raters then scored the listeners based on the audio recordings of the interview. To make the scoring as objective as possible, we made a list of questions regarding important contents of the story (such as “what happened on her way to the hotel”). The correct answers included several critical points covering characters, places, time, and motivations et al. A score was given for each question on the list according to the information a listener provided about those key points. The scores for all questions were then summed up and divided by the full score. There was a high agreement between the assessments made by the two raters ($r_{(64)} = 0.80$, by Pearson’s correlation). The average score given by the two raters was used as the measurement for speech comprehension. To account for a potential effect of memory capacity on the story-retelling task, we measured participants’ memory span using a

digit memory test (Wechsler, 1987). The number of digits correctly repeated forward and backward were used as two covariates for the following brain-behavior analysis.

2.4 MRI acquisition and preprocessing

Imaging data were collected with a 3T Siemens Trio scanner in the MRI Center of the Southwest University of China. Functional images were acquired using a gradient echo-planar imaging sequence with the following parameters: repetition time = 2000 ms, echo time = 30 ms, flip angle = 90°, field of view = 220 mm^2, matrix size = 64 × 64, 32 interleaved slice, voxel size = 3.44 × 3.44 × 3.99 mm^3. T1 structural images were acquired using a MPRAGE sequence with the following parameters: repetition time = 2530 ms, echo time = 3.39 ms, flip angle = 7°, FOV = 256 mm^2, scan order = interleaved, matrix size = 256 × 256, and voxel size = 1.0 × 1.0 × 1.33 mm^3.

After discarding the first three volumes to allow for the equilibration of magnetic fields, a total of 310 volumes were acquired in each run. The DPABI toolkit (Yan, Wang, Zuo, & Zang, 2016) based on SPM12 (www.fil.ion.ucl.ac.uk/spm/) was utilized for image preprocessing. The preprocessing steps included slice-timing correction, spatial realignment, co-registration to individual subject's anatomical images, normalization to the Montreal Neurological Institute (MNI) space, resampling into a 3 × 3 × 3 mm^3 voxel size, and smoothing (FWHM = 7mm). The resultant images were further detrended, nuisance variable regressed, and high-pass filtered (1/128 Hz). The nuisance regression included the removal of five principal components of white matter and cerebrospinal fluid within individual subjects' T1 segmentation mask (Behzadi et al.,

2007), as well as Friston's 24 motion parameters (including each of the six motion parameters of the current and preceding volume, plus each of these values squared) (Friston et al., 1996). The datasets of four participants listening to the Chinese story, four participants listening to the Mongolian story, and two participants during the resting state were discarded due to excessive head movement (more than 3mm or 3 degrees).

2.5 Data analysis

2.5.1 Group Independent Component Analysis and component selection

To detect the DMN and pLN, a group spatial ICA was conducted using Group ICA of fMRI Toolbox (GIFT). The first 10 volumes corresponding to the 20s fixation were not included in this analysis. The dimensionality of data from each participant was first reduced to 30 dimensions using principal component analysis. Next, the dimension-reduced data (including both the speaker and all listeners) were temporally concatenated and a group dimension reduction was performed. Twenty independent sources were generated with the ICA using the infomax algorithm. Finally, spatial maps and associated time series for individual participants were reconstructed using aggregate components and original data via spatial-temporal regression (Beckmann et al., 2009). This dual-regression approach Since the spatial maps and time series of components have arbitrary units after the back-reconstruction step, they were scaled using Z-scores. These component time series were used in the further within- and between-brain connectivity analyses. A one-sample *t*-test was conducted on the scaled

maps to obtain the spatial distribution of each component.

The DMN and pLN were identified based on the spatial overlap of the component map with the meta-analytic maps for the DMN (using a search term "default mode network") and the pLN (using a search term "language") generated by the Neurosynth (https://neurosynth.org). For each component, its spatial map (i.e., the above *t*-statistic map) was thresholded using a set of *t*-values (ranges from 2.5 to 5.5, with an increment of 0.5) and then binarized. The degree of the spatial overlap between each binary component map and the thresholded binary meta-analytic map was then quantified by the widely-used Dice index (Dice, 1945), and the component showing the highest Dice score was selected. Upon identifying the components corresponding to the DMN and pLN, we thresholded the *t*-statistic maps with a $p < 0.05$ (corrected for multiple comparisons with family-wise error (FWE) correction) for visualization and making masks used in the following analysis.

2.5.2 BOLD signal change

To assess how the DMN and pLN responded to external stimulation, we calculated percent changes in BOLD signal amplitudes during speech listening (the task) relative to the 20s fixation period (the baseline). The BOLD signal changes were defined as [(*task-baseline*) / *baseline*], where *task* and *baseline* corresponded to the mean signal amplitudes of time points within the task phase and the baseline phase, respectively. We first extracted the BOLD signal change in every voxel across the whole brain and applied a voxel-wise *t*-test to assess whether the signal change differed from zero at the group level. Next, two masks were created using the thresholded *t*-maps for the DMN

and pLN components obtained above, and a histogram of *t*-values from all voxels lying within each mask was generated. The histogram is used to reflect the overall distribution of BOLD signal change within the two networks. In addition to the descriptive analysis, we averaged the percent signal changes over all voxels within the mask, and performed a one-tailed *t*-test to determine whether the network activity was significantly increased (in the pLN) or decreased (in the DMN).

2.5.3 Functional and effective connectivity analyses

To understand how the DMN may contribute to speech comprehension, we examined how this network interacted with the well-understood pLN. To compare this study with most studies in the literature, we first examined the (undirected) functional connectivity between the DMN and pLN. For each participant, a Pearson's correlation coefficient between the time series of DMN and pLN was computed. Then a *t*-test was applied to assess whether the correlation coefficients differed significantly from zero at the group level.

Functional connectivities only capture the statistical dependencies between two signals (Friston, 2011).To better understand the observed dependencies, we further estimated effective connections (causal interactions) between the DMN and pLN using Dynamic Causal Modeling (DCM) toolbox in SPM12 (Friston, 2011). Three bilinear models without an input entering the system were constructed for each participant. Model 1 specified a connection from the pLN to DMN. Model 2 specified a connection from DMN to pLN. Model 3 specified bidirectional connections. All models included recurrent connections of each network. Since the speech signals presented during the

task were continuous stimuli, we fitted models based on the frequency-domain cross-spectral density of the time series, which is a technique originally developed for the modeling of resting-state fMRI data (Friston et al., 2014). Bayesian Model Selection (BMS) was used to identify the best-fitting model at the group level. Upon identifying the optimal model, Bayesian Parameter Averaging (BPA) was applied to estimate group-mean effective connectivity (EC) and to make statistical inference using Bayesian Parameter Averaging (Neumann & Lohmann, 2003; Kasess et al., 2010; Murta et al., 2012; Friston et al., 2014). In this method, individual subjects' parameters were weighted by their precision to compute the mean across subjects.

2.5.4 Measurement for listener-speaker neural coupling

The above analyses had focused on neural activities within the listeners' brains. We next examined whether the activities of DMN and pLN were coupled across the listener and speaker. Pearson's correlation coefficients between the ICA-derived network time series from each listener's brain and the time series of the homologous network from the speaker's brain were calculated separately for the DMN and pLN. Previous work has demonstrated that listeners need time to process the high-level information in speech in order to get aligned with the speaker (Stephens et al., 2010; Kuhlen et al., 2012; Y. Liu et al., 2017; L. Liu et al., 2020). To account for this effect, we repeated the inter-brain correlation analysis by shifting the time series of listener's network activity relative to that of the speaker's by 0-14s with 2s increments. At shift zero, the listener's network activity was time-locked to the speaker's vocalization. To assess the validity of using ICA-derived component time series to measure listener-speaker neural

316 coupling, we also performed the inter-brain correlation analysis using voxel-wise time
317 series (see supplementary material).

318 Two-tailed *t*-tests were used to assess whether the neural couplings differed
319 significantly from zero at the group level. We also examined whether the listener-
320 speaker neural coupling would differ between the pLN and DMN by using a paired *t*-
321 test. Multiple comparisons were corrected with an FDR q = 0.05 using the Benjamini–
322 Hochberg procedure (Benjamini & Hochberg, 1995).

323 **2.5.5 Control analyses**

324 To the best of our knowledge, this is the first study using network time series derived
325 from group ICA to assess interbrain neural coupling. It is unclear whether the interbrain
326 network couplings arose merely because the data from both the speaker and listeners
327 were pooled together during the ICA procedure. To test this possibility, we conducted
328 the group ICA for the same participants during resting states and assessed network
329 correlations between the speaker and each listener. In addition, we tested whether the
330 interbrain couplings arose simply because the speaker and listener received the same
331 speech signals. For this purpose, listener-speaker network correlations were analyzed
332 for the same participants listening to the Mongolian speaker telling a real-life story.
333 Finally, the listener-speaker neural coupling is assumed to reflect inter-brain alignment
334 at multiple levels of linguistic and conceptual representations (Menenti et al., 2012;
335 Schoot et al., 2016). However, interbrain coupling could also arise from a general effect
336 of arousal, focused attention, or memory shared by the listeners and the speaker. In

order to distinguish between the two possibilities, we examined the specificity of listener-speaker neural coupling to story contents using a Support Vector Machine (SVM) technique. Details for these analyses are provided in the supplementary material.

2.5.6 Brain-behavioral correlation

To determine the functional role of listener-speaker neural coupling in the DMN and pLN, we examined the correlation between coupling strength and listeners' comprehension scores. A Pearson's partial correlation analysis was applied, which controlled for the potential effect of memory span. To assess the generalizability of the brain-behavior relationship we may find, we further performed a within-data set cross-validation (Shen et al., 2017). Specifically, a linear regression model was built on the data from all participants but one, taking listener-speaker neural couplings in both the DMN and pLN as two predictors and listener's comprehension score as the outcome. Then the model was used to predict the comprehension score of the left-out participant based on his or her neural coupling with the speaker in the DMN and pLN. With this leave-one-out procedure, we obtained a predicted score for every participant. To quantitatively assess the predictive power of the model, a Pearson's correlation between the predicted scores and actual scores was calculated as a test statistic. Since the null distribution associated with this test statistic is unknown, we performed a permutation test (N = 1000) to determine whether the prediction (i.e., the correlation value) was significantly better than expected by chance. In each permutation, we randomized the comprehension scores across the training sample and re-conducted the prediction. A *p*-value was calculated as the percentage of permutations that showed a lower r-value than

359 the actual r-value.

360 **2.5.7 Analyses for potential associations of within-brain response profile with**
361 **between-brain neural coupling**

362 Finally, we explored whether the listener-speaker neural coupling was affected by (or
363 had an effect on) the task-induced (de)activation and network interactions within
364 listeners' brains. For this purpose, we examined the correlation between the strength of
365 interbrain coupling (for the peaking time) and the averaged BOLD signal change within
366 the network. In addition, we examined the correlation between the strength of interbrain
367 neural coupling and the effective connectivity between the DMN and pLN. As
368 exploratory analyses, results were not corrected for multiple companions.

369 **3. Results**

370 **3.1 Behavioral results**

371 For the story-recall task administered shortly after the scanning, participants scored on
372 average 91.58 ± 8.00%. There was no significant correlation between the
373 comprehension score and the digit-memory span ($p > 0.28$). For the Mongolian story
374 (the control condition), the interview showed that the participants had tried to
375 comprehend the speech but all failed.

3.2 ICA-based detection of the default mode network and the perisylvian language network

By measuring the spatial overlap of each ICA component with the meta-analytic maps derived from the Neurosynth, we identified two components corresponding to the DMN and pLN, which consistently showed the highest Dice score across a set of thresholds among the 20 components resulted from the group ICA analysis. The DMN component covered mainly the bilateral precuneus, inferior parietal lobe, and part of the medial frontal gyrus (Fig. 2a). The pLN component covered mainly the bilateral superior and middle temporal, precentral, and inferior frontal gyrus, as well as the supplementary motor area. Homologues of the two networks were found for participants listening to the unintelligible Mongolian speech and at the resting state (Fig. S1).

3.3 Opposite activation patterns between the DMN and pLN

To assess task-induced (de)activations in the pLN and DMN, we measured percent signal changes during speech listening relative to the fixation period preceding the task. In agreement with previous findings, a large portion of voxels (67.8%) within the DMN exhibited decreased activity, with the most significant effect found in the precuneus (Fig. 2b). In comparison, most voxels (78.8%) within the pLN exhibited increased activity, with the most significant effect shown in the bilateral temporal cortex. In line with the distribution pattern, the *t*-test on the BOLD signal change averaged across voxels within the network showed a marginally significant effect of deactivation in the DMN ($t_{(61)}$= -1.56, one-tailed p = 0.06), and a quite significant effect of activation in

the pLN ($t_{(61)}$= 7.09, one-tailed $p < 10^{-9}$) (Fig. 2b). These effects were more prominent when the part of voxels overlapped between the two networks were not included in the analysis (for DMN: $t_{(61)}$ = -2.01, p = 0.024; for pLN: $t_{(61)}$ = 7.40, $p < 10^{-9}$).

We note the baseline used to calculate the BOLD signal change might be too short. To assess this effect, we compared this method with a conventional General Linear Modeling (GLM) method which included longer baselines using an independent dataset. The results showed that the activation and deactivation patterns obtained by the two approaches were similar, with the one obtained from the GLM being statistically more significant (Fig. S2). Thus, a relatively short baseline does not compromise the findings of opposite activation patterns between the DMN and pLN.

3.4 Functional and effective connectivity between the DMN and pLN

In line with previous findings (Uddin et al., 2009; Smith et al., 2012), the functional connectivity analysis showed that the dynamics of DMN and pLN were significantly anticorrelated over time (mean r = -0.13, $t_{(61)}$= - 6.90, p < 10^{-8}) (Fig. 3a). We further examined the effective connectivity (causal interactions) between the two networks using DCM. The Bayesian Model Selection procedure reported the model with bidirectional connectivity to be the best-fitting, with model exceedance probabilities greater than 99.9%. The Bayesian Parameter Averaging produced a negative value of group-mean connection from the pLN to the DMN (EC = -0.10, Bayesian probability =1.0), indicating that the activity of pLN had an inhibitory effect on the activity of DMN. The value of group-mean connection from the DMN to the pLN was positive

(EC = 0.07, Bayesian probability = 1.0), indicating the excitatory influence of the DMN's activity over the activity of pLN (Fig. 3b). By comparison, the connection between the pLN and DMN in both directions were negative when comprehension failed (pLN → DMN: -0.020; DMN → pLN: -0.0002) or when participants were at resting states (pLN → DMN: -0.005; DMN → pLN: -0.011).

3.5 Listener-speaker neural coupling in the DMN and pLN

The task-negative activity and the anticorrelation with the pLN derived from the above within-brain analyses seem to indicate a negative role of the DMN in speech comprehension. However, the between-brain analyses revealed that the network dynamics of DMN in the listener's brain were significantly coupled (temporally correlated) with the dynamics of the homologous network in the speaker's brain (Fig. 3A). We also found significant listener-speaker neural coupling in the pLN. The temporal profile differed between the two networks: in the pLN, the listener's network activity lagged the speaker's network activity by 0-8s and peaked at a lag of 4s; in the DMN, the listener's activity lagged by 2-12s and peaked at a lag of 6s. Overall, the listener-speaker neural coupling in the DMN was stronger than that in the pLN. An exception is at lag 0s, wherein listener-speaker neural coupling in the pLN was significantly stronger than that in the DMN. A similar pattern was obtained from the analysis based on inter-brain voxel-to-voxel correlation (Fig. 4a and Fig.S3).

A SVM classifier which took the interbrain couplings in the DMN and pLN as two features was built to identify which story (out of three) told by the Chinese speaker was

played to each listener. The SVM classifier achieved an average accuracy of 89.25% (sensitivity: 77.42%, specificity: 95.16%), which was higher than the best performance (70.97%) of the 1000 permutations. The area under the receiver-operating characteristic (ROC) curve was 0.90, demonstrating that the model discriminated well between the content-matched and -unmatched pairs (Fig. 4b). These results suggest that the interbrain couplings likely arose from semantic and conceptual representations shared by the speaker and listeners, rather than from joint attention, arousal or working memory processes.

When communication failed or when participants were at the resting state, no significant neural coupling between the listener and speaker was found in either the DMN or pLN (Fig. 4c). This result suggests that the coupling of network dynamics was not driven by a potentially shared acoustic representation between the listener and speaker. It also mitigates a methodological concern that the network time series correlated between the listeners and the speaker was because that the data from both sides had been pooled together to identify the network component during the group ICA analysis.

3.6 Listener-speaker neural coupling predicts speech comprehension

To determine the behavioral significance of listener-speaker neural coupling, we examined its correlation with the listener's level of speech understanding. A significant positive correlation was found between the comprehension score and the strength of the listener's neural coupling with the speaker in the DMN (partial $r_{(60)} = 0.32$, $p = 0.012$,

with memory span regressed out). There was also a positive correlation between coupling strength in the pLN and comprehension scores (partial $r_{(60)} = 0.31$, $p = 0.015$, with memory span regressed out) (Fig. 5).

A leave-one-out cross-validation procedure was performed to evaluate the generalizability of the brain-behavioral relation. In each fold of cross-validation, a regression model was built which took the listener-speaker neural coupling in both the DMN and pLN as inputs and generated predictions of comprehension scores in a novel participant. The predicted scores were positively correlated with actual scores, with predictive power significantly better than the results from 1000 iterations of permutations testing ($r_{(60)} = 0.29$, $p = 0.009$) (Fig. 5). The predictive power was better than using the data of DMN or pLN alone (DMN: $r = 0.18$; pLN: $r = 0.23$). These findings suggest both the DMN and pLN contribute to speech comprehension.

3.7 Covarying of within-brain connectivity and between-brain neural coupling

The analyses exploring the potential association of within-brain BOLD signal changes with between-brain couplings showed no significant results for either the DMN or pLN (both $p > 0.1$), suggesting interbrain neural alignments were (partly) independent from within-brain neural responses to external stimulation. Interestingly, we found that the DMN → pLN effective connection within the listeners' brains tended to correlate positively with the listener-speaker neural couplings in the DMN ($r_{(60)} = 0.245$, $p = 0.055$, by Pearson's correlation). In addition, the increase of the DMN → pLN connection during successful speech comprehension relative to the failed condition also

positively correlated with the increase in listener-speaker neural coupling in the DMN ($r_{(57)} = 0.26$, $p = 0.047$) (Fig. 6). These results may reflect the top-down modulation of higher-level conceptual processing implemented by the DMN on lower-level signal-decoding processes implemented by the pLN. No significant correlation was found between the pLN → DMN connection and listener-speaker neural coupling in the pLN (both $p > 0.1$).

4. **Discussion**

In this study, we investigated the functional properties of DMN during natural verbal communication and examined both within-brain response profiles and between-brain interactions. Replicating previous findings (Wirth et al., 2011; Rodriguez Moreno et al., 2014; Humphreys et al., 2015; Jackson et al., 2019), most DMN areas in the listener's brain exhibited reduced activities during speech comprehension relative to the low-level baseline, and the network dynamics of DMN were anticorrelated with the task-positive pLN. Dynamic Causal Modeling showed the pLN had inhibitory influence on the DMN, whereas the DMN had excitatory influence on the pLN. Further two-brain analyses revealed the activities of the DMN in the listener's brain were tightly coupled with the activities of homologous network in the speaker's brain. Significant listener-speaker neural coupling was also observed in the pLN, but was weaker and peaked earlier than that in the DMN. Notably, the interbrain network couplings were not driven by low-level perceptual processing, neither by general attention or memory processes: when communication failed, the interbrain coupling vanished; moreover, the coupling

pattern can be used to discriminate which story (out of three) was processed by the listener with a high accuracy. Finally, the strength of inter-brain neural coupling in both the DMN and pLN predicted the degree to which listeners comprehended and memorized the speech.

Currently, whether and how the DMN contributes to auditory language comprehension remains elusive. To explore the potential role of DMN in language processing, existing studies primarily rely on the manipulation of task stimuli or demands to induce changes in neural activities within individual brains. However, the DMN seems to be only weakly modulated by the task-related change, or even de-activated during task processing relative to low-level baselines. Several studies measured the similarity (temporal correlations) of brain activities across subjects receiving the same language inputs, and observed significant ISC and task-induced changes of ISC in the DMN (Wilson et al., 2007; Simony et al., 2016; Yeshurun et al., 2017). These findings are suggested to reflect the active role of DMN in the semantic and conceptual processing of language. However, the ISC could also arise from the shared level of task engagement across subjects. In this study, by examining participants in communication rather than in isolation, we found significant listener-speaker neural coupling in the DMN and a positive correlation of this coupling with behavioral measurements. These findings suggest the information represented in the DMN was aligned between the listener and speaker, which facilitated information transfer from the speaker to the listener. Notably, the inter-brain coupling was unlikely an epiphenomenon of task engagement, since the level of task engagement was not

equivalent across the speaker and listener over the course of communication. For instance, when encountering difficulties in lexical retrieval or organizing thoughts, the speaker would articulate slower than usual; however, for the listener, the slowly-presented utterances are easier to comprehend.

While most DMN areas exhibited reduced activities during speech listening relative to the fixation period, most pLN areas showed increased activities. Moreover, the dynamics of the two networks were anticorrelated. Taking the perspective from non-language studies, such a pattern seems to indicate the DMN's activity interferes with the ongoing speech processing and needs to be suppressed in order to support the externally-directed cognitive processes (Anticevic et al., 2012; Gauffin et al., 2013; Zhou et al., 2016). However, further two-brain analyses indicate a division of labor rather than simply an antagonistic relationship between the DMN and the pLN. Like the DMN, the activities in the pLN were coupled across the listener and speaker, and the interbrain coupling was also positively correlated with speech comprehension. Notably, the strength and temporal profile of inter-brain coupling differed between the two networks. In the pLN, listeners' coupling responses to the speaker first occurred without a temporal lag (i.e., time-locked to the speaker's vocalization) and then reached the peak at a lag of 4s; in the DMN, the coupling responses first occurred at a lag of 2s and reached the peak at a lag of 6s. Similarly delayed coupling responses in the listener's brain to the speaker's brain have been reported in previous studies applying fNIRS (Y. Liu et al., 2017), EEG (Kuhlen et al., 2012; Perez et al., 2017) and fMRI (Stephens et al., 2010; Silbert et al., 2014) techniques. The response lag is likely related

546 to the hierarchical nature of language production and comprehension. During natural
547 communication, the speaker develops a concept or message, retrieves the lexical-
548 syntactic forms, organizes them into sentences based on grammar principles, and finally
549 produce utterances; the listener analyzes the sounds, maps utterances into meaning,
550 integrates words into sentences, and builds a situation model. The high-level
551 information takes relatively long time for the speaker to produce and for the listener to
552 comprehend, which may result in the delay in interbrain coupling. The differences in
553 the response lag thus suggest the DMN was dominantly involved in the processing of
554 high-level information integrated over a larger timescale (about 2-12s), while the pLN
555 was involved in processing lower-level information integrated over a shorter timescale
556 (about 0-8s). We also observed the overall interbrain couplings in the DMN were
557 stronger than in the pLN. This may because the listener and speaker obtained greater
558 alignments on high-level representations than on low-level representations (such as
559 auditory-motor information).

560 The proposed division of labor between the DMN and pLN is in line with an
561 influential theory proposed by Berwick and colleagues (Berwick et al., 2013).
562 According to this theory, the basic design for language faculty is composed of a basic
563 language system that represents syntactic rules and lexical items, a sensory-motor
564 system that connects mental expressions generated by the basic language system to the
565 external world, and an internal conceptual-intentional system which connects these
566 mental expressions to activities in the internalized mental world. Based on the results
567 from this study and prior knowledge about the DMN under resting states, we speculate

568 the DMN likely occupies (part of) the internal conceptual-intentional system, and the
569 pLN occupies (part of) the systems for linguistic representation and sensory-motor
570 conversion. This internal versus external functional division would explain the opposite
571 response profiles of the DMN and pLN to external stimulations while both networks
572 play an important role in language comprehension. If the assumption that DMN
573 occupies the internal conceptual-intentional system holds, it would call for a re-
574 interpretation of the activity of DMN in several circumstances. For instance, the
575 overactivation of DMN in older adults during task processing may unnecessarily reflect
576 a failure to suppress task-unrelated thoughts, but may reflect a shift from a cognitive
577 style dominated by externally-directed processing to a style with more engagement of
578 internally-directed processing.

579 To conclude, this study revealed that, despite being deactivated and anticorrelated
580 with the task-positive pLN, the DMN plays an active role in narrative speech
581 comprehension, as demonstrated by the tight listener-speaker neural coupling in this
582 network and the positive correlation of network coupling with speech comprehension.
583 We infer the DMN may occupy an internal conceptual system that works cooperatively
584 with the externally-oriented pLN to subserve language comprehension. These findings
585 shed new light on the functional property of DMN during task processing and extend
586 our understanding about how language is processed in the brain. This study also
587 highlights the importance of taking the two-brain approach to gain a more complete
588 picture of the brain.

589

Acknowledgments: This work was supported by grants from the National Natural Science Foundation of China (NSFC:31900802, 31971036), China Postdoctoral Science Foundation (2019M653248) and the Open Research Fund of the State Key Laboratory of Cognitive Neuroscience and Learning (CNLYB1803). No conflict of interest is declared.

Credit Author Statement

Guosheng Ding & Hong Chen: Conceptualization, Writing-Review & Editing, Supervision; Lanfang Liu: Conceptualization, Formal analysis, Writing-Original Draft, Project administration; Zhiting Ren: Project administration, Writing-Review & Editing; Hehui Li, Yuxuan Zhang, Jiang Qiu & Chunming Lu: Conceptualization, Writing - Review & Editing.

5. Reference

Anticevic, A., Cole, M. W., Murray, J. D., Corlett, P. R., Wang, X.-J., & Krystal, J. H. (2012). The role of default network deactivation in cognition and disease. *Trends in Cognitive Sciences, 16*(12), 584-592. doi:https://doi.org/10.1016/j.tics.2012.10.008

Beckmann, C. F., Mackay, C. E., Filippini, N., & Smith, S. M. (2009). Group comparison of resting-state FMRI data using multi-subject ICA and dual regression. *NeuroImage, 47*(Suppl 1), S148.

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage,*

37(1), 90-101. doi:https://doi.org/10.1016/j.neuroimage.2007.04.042

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289-300. doi:10.1111/j.2517-6161.1995.tb02031.x

Berwick, R. C., Friederici, A. D., Chomsky, N., & Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends in Cognitive Sciences, 17*(2), 89-98. doi:https://doi.org/10.1016/j.tics.2012.12.002

Cuevas, P., Steines, M., He, Y., Nagels, A., Culham, J., & Straube, B. (2019). The facilitative effect of gestures on the neural processing of semantic complexity in a continuous narrative. *NeuroImage, 195*, 38-47. doi:https://doi.org/10.1016/j.neuroimage.2019.03.054

Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology, 26*(3), 297-302. doi:10.2307/1932409

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America, 102*(27), 9673-9678. doi:10.1073/pnas.0504136102

Friston, K. J. (2011). Functional and Effective Connectivity: A Review. *Brain Connectivity, 1*(1), 13-36. doi:10.1089/brain.2011.0008

Friston, K. J., Kahan, J., Biswal, B., & Razi, A. (2014). A DCM for resting state fMRI. *NeuroImage, 94*, 396-407. doi:https://doi.org/10.1016/j.neuroimage.2013.12.009

633 Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., & Turner, R. (1996).
634 Movement-Related effects in fMRI time-series. *Magnetic Resonance in Medicine,*
635 *35*(3), 346-355. doi:10.1002/mrm.1910350312
636 Gauffin, H., van Ettinger-Veenstra, H., Landtblom, A.-M., Ulrici, D., McAllister, A.,
637 Karlsson, T., & Engström, M. (2013). Impaired language function in generalized
638 epilepsy: Inadequate suppression of the default mode network. *Epilepsy &*
639 *Behavior, 28*(1), 26-35. doi:https://doi.org/10.1016/j.yebeh.2013.04.001
640 Greicius, M. D., & Menon, V. (2004). Default-Mode Activity during a Passive Sensory
641 Task: Uncoupled from Deactivation but Impacting Activation. *Journal of Cognitive*
642 *Neuroscience, 16*(9), 1484-1492. doi:10.1162/0898929042568532
643 Hasson, U., & Frith, C. D. (2016). Mirroring and beyond: coupled dynamics as a
644 generalized framework for modelling social interactions. *Philosophical*
645 *Transactions of the Royal Society of London. Series B: Biological Sciences,*
646 *371*(1693). doi:10.1098/rstb.2015.0366
647 Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-
648 to-brain coupling: a mechanism for creating and sharing a social world. *Trends in*
649 *Cognitive Science, 16*(2), 114-121. doi:10.1016/j.tics.2011.12.007
650 Horowitz-Kraus, T., Farah, R., Hajinazarian, A., Eaton, K., Rajagopal, A., Schmithorst,
651 V. J., . . . Holland, S. K. (2017). Maturation of Brain Regions Related to the Default
652 Mode Network during Adolescence Facilitates Narrative Comprehension. *Journal*
653 *of child and adolescent behavior, 5*(1), 328. doi:10.4172/2375-4494.1000328
654 Humphreys, G. F., Hoffman, P., Visser, M., Binney, R. J., & Lambon Ralph, M. A.

(2015). Establishing task- and modality-dependent dissociations between the semantic and default mode networks. *Proceedings of the National Academy of Sciences, 112*(25), 7857. doi:10.1073/pnas.1422760112

Jackson, R. L., Cloutman, L. L., & Lambon Ralph, M. A. (2019). Exploring distinct default mode and semantic networks using a systematic ICA approach. *Cortex, 113*, 279-297. doi:https://doi.org/10.1016/j.cortex.2018.12.019

Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., & Lu, C. (2012). Neural synchronization during face-to-face communication. *Journal of Neuroscience, 32*(45), 16064-16069. doi:10.1523/JNEUROSCI.2926-12.2012

Kasess, C. H., Stephan, K. E., Weissenbacher, A., Pezawas, L., Moser, E., & Windischberger, C. (2010). Multi-subject analyses with dynamic causal modeling. *NeuroImage, 49*(4), 3065-3074. doi:https://doi.org/10.1016/j.neuroimage.2009.11.037

Kuhlen, A. K., Allefeld, C., & Haynes, J.-D. (2012). Content-specific coordination of listeners' to speakers' EEG during communication. *Frontiers in Human Neuroscience, 6*, 266. doi: https://doi.org/10.3389/fnhum.2012.00266

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

Li, W., Mai, X., & Liu, C. (2014). The default mode network and social understanding of others: what do brain connectivity studies tell us. *8*(74). doi:10.3389/fnhum.2014.00074

Liu, L., Zhang, Y., Zhou, Q., Garrett, D. D., Lu, C., Chen, A., . . . Ding, G. (2020).

677 Auditory–Articulatory Neural Alignment between Listener and Speaker during
678 Verbal Communication. *Cerebral Cortex, 30*(3), 942-951.
679 doi:https://doi.org/10.1093/cercor/bhz138

680 Liu, Y., Piazza, E. A., Simony, E., Shewokis, P. A., Onaral, B., Hasson, U., & Ayaz, H.
681 (2017). Measuring speaker-listener neural coupling with functional near infrared
682 spectroscopy. *Scientific Reports, 7*, 43293. doi:10.1038/srep43293

683 McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J., & Binder, J. R. (2003). A
684 Parametric Manipulation of Factors Affecting Task-induced Deactivation in
685 Functional Neuroimaging. *Journal of Cognitive Neuroscience, 15*(3), 394-408.
686 doi:10.1162/089892903321593117

687 Menenti, L., Garrod, S., & Pickering, M. (2012). Toward a neural basis of interactive
688 alignment in conversation. *Frontiers in Human Neuroscience, 6*(185).
689 doi:10.3389/fnhum.2012.00185

690 Murta, T., Leal, A., Garrido, M. I., & Figueiredo, P. (2012). Dynamic Causal Modelling
691 of epileptic seizure propagation pathways: A combined EEG–fMRI study.
692 *NeuroImage, 62*(3), 1634-1642.
693 doi:https://doi.org/10.1016/j.neuroimage.2012.05.053

694 Neumann, J., & Lohmann, G. (2003). Bayesian second-level analysis of functional
695 magnetic resonance images. *NeuroImage, 20*(2), 1346-1355.
696 doi:https://doi.org/10.1016/S1053-8119(03)00443-9

697 Perez, A., Carreiras, M., & Dunabeitia, J. A. (2017). Brain-to-brain entrainment: EEG
698 interbrain synchronization while speaking and listening. *Scientific Reports, 7*(1),

699 4190. doi:10.1038/s41598-017-04464-4

700 Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful
701 communication. *Research on Language & Computation, 4*(2), 203-228.
702 doi:https://doi.org/10.1007/s11168-006-9004-0

703 Raichle, M. E. (2010). Two views of brain function. *Trends in Cognitive Science, 14*(4),
704 180-190. doi:10.1016/j.tics.2010.01.008

705 Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., &
706 Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the*
707 *National Academy of Sciences, 98*(2), 676-682. doi:10.1073/pnas.98.2.676

708 Rodriguez Moreno, D., Schiff, N. D., & Hirsch, J. (2014). Negative Blood Oxygen
709 Level Dependent Signals During Speech Comprehension. *Brain Connectivity, 5*(4),
710 232-244. doi:10.1089/brain.2014.0272

711 Schoot, L., Hagoort, P., & Segaert, K. (2016). What can we learn from a two-brain
712 approach to verbal interaction? *Neuroscience & Biobehavioral Reviews, 68*, 454-
713 459. doi:http://dx.doi.org/10.1016/j.neubiorev.2016.06.009

714 Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X.,
715 & Constable, R. T. (2017). Using connectome-based predictive modeling to predict
716 individual behavior from brain connectivity. *Nature Protocols, 12*(3), 506-518.
717 doi:10.1038/nprot.2016.178

718 Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled
719 neural systems underlie the production and comprehension of naturalistic narrative
720 speech. *Proceedings of the National Academy of Sciences, 111*(43), E4687-E4696.

721 doi:https://doi.org/10.1073/pnas.1323812111

722 Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., & Hasson, U.
723 (2016). Dynamic reconfiguration of the default mode network during narrative
724 comprehension. *Nature communications, 7*(1), 12141. doi:10.1038/ncomms12141

725 Smith, S. M., Miller, K. L., Moeller, S., Xu, J., Auerbach, E. J., Woolrich, M. W., . . .
726 Ugurbil, K. (2012). Temporally-independent functional modes of spontaneous
727 brain activity. *Proceedings of the National Academy of Sciences, 109*(8), 3131-
728 3136. doi:10.1073/pnas.1121329109

729 Spreng, R. N., & Grady, C. L. (2009). Patterns of Brain Activity Supporting
730 Autobiographical Memory, Prospection, and Theory of Mind, and Their
731 Relationship to the Default Mode Network. *Journal of Cognitive Neuroscience,*
732 *22*(6), 1112-1123. doi:10.1162/jocn.2009.21282

733 Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2008). The Common Neural Basis of
734 Autobiographical Memory, Prospection, Navigation, Theory of Mind, and the
735 Default Mode: A Quantitative Meta-analysis. *Journal of Cognitive Neuroscience,*
736 *21*(3), 489-510. doi:10.1162/jocn.2008.21029

737 Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling
738 underlies successful communication. *Proceedings of the National Academy of*
739 *Sciences, 107*(32), 14425-14430. doi:https://doi.org/10.1073/pnas.1323812111

740 Stolk, A., Verhagen, L., & Toni, I. (2016a). Conceptual Alignment: How Brains Achieve
741 Mutual Understanding. *Trends Cogn Sci, 20*(3), 180-191.
742 doi:10.1016/j.tics.2015.11.007

743 Stolk, A., Verhagen, L., & Toni, I. (2016b). Conceptual Alignment: How Brains
744 Achieve Mutual Understanding. *Trends in Cognitive Science, 20*(3), 180-191.
745 doi:10.1016/j.tics.2015.11.007
746 Szaflarski, J. P., Altaye, M., Rajagopal, A., Eaton, K., Meng, X., Plante, E., & Holland,
747 S. K. (2012). A 10-year longitudinal fMRI study of narrative comprehension in
748 children and adolescents. *NeuroImage, 63*(3), 1188-1195.
749 doi:https://doi.org/10.1016/j.neuroimage.2012.08.049
750 Uddin, L. Q., Clare Kelly, A. M., Biswal, B. B., Xavier Castellanos, F., & Milham, M.
751 P. (2009). Functional connectivity of default mode network components:
752 Correlation, anticorrelation, and causality. *Human Brain Mapping, 30*(2), 625-637.
753 doi:10.1002/hbm.20531
754 Wechsler, D. (1987). *WMS-R: Wechsler memory scale-revised*: Psychological
755 Corporation.
756 Wheatley, T., Boncz, A., Toni, I., & Stolk, A. (2019). Beyond the Isolated Brain: The
757 Promise and Challenge of Interacting Minds. *Neuron, 103*(2), 186-188.
758 doi:10.1016/j.neuron.2019.05.009
759 Wilson, S. M., Molnar-Szakacs, I., & Iacoboni, M. (2007). Beyond Superior Temporal
760 Cortex: Intersubject Correlations in Narrative Speech Comprehension. *Cerebral*
761 *Cortex, 18*(1), 230-242. doi:10.1093/cercor/bhm049
762 Wirth, M., Jann, K., Dierks, T., Federspiel, A., Wiest, R., & Horn, H. (2011). Semantic
763 memory involvement in the default mode network: a functional neuroimaging
764 study using independent component analysis. *NeuroImage, 54*(4), 3057-3066.

765 doi:https://doi.org/10.1016/j.neuroimage.2010.10.039

766 Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., & Hasson,
767 U. (2017). Same Story, Different Story:The Neural Representation of Interpretive
768 Frameworks. *28*(3), 307-319. doi:10.1177/0956797616682029

769 Zhou, L., Pu, W., Wang, J., Liu, H., Wu, G., Liu, C., . . . Liu, Z. (2016). Inefficient DMN
770 Suppression in Schizophrenia Patients with Impaired Cognitive Function but not
771 Patients with Preserved Cognitive Function. *Scientific Reports, 6*(1), 21657.
772 doi:10.1038/srep21657

773 Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension
774 and memory. *Psychological Bulletin, 123*(2), 162-185. doi:10.1037/0033-
775 2909.123.2.162

776

777

778

779

780

6. **Figures and captions**

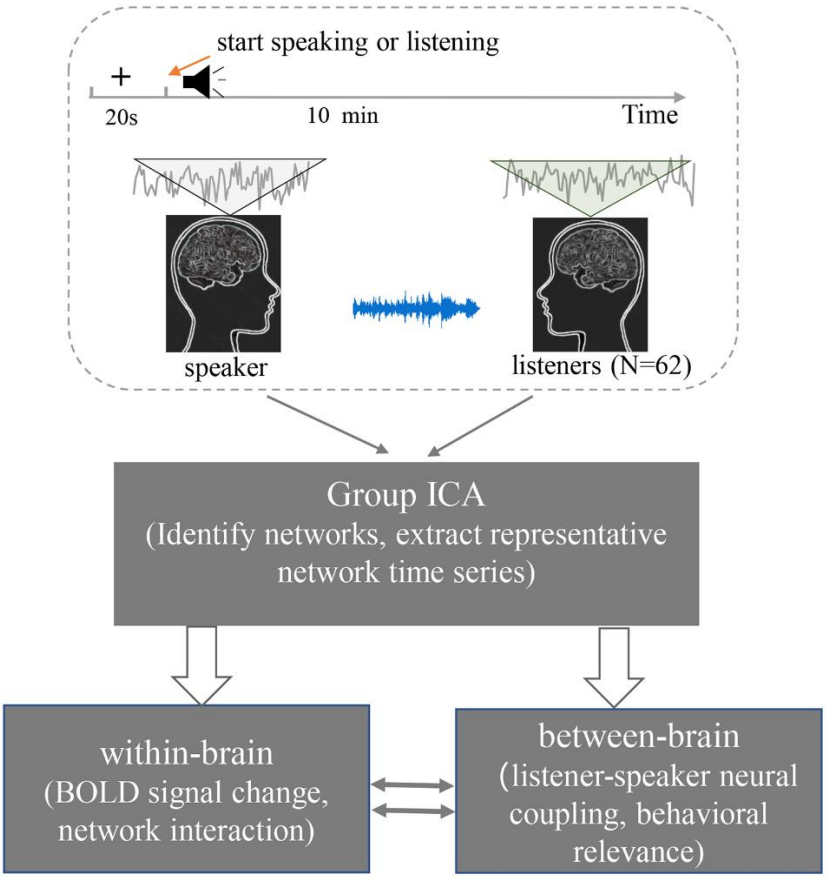


Figure 1. Shema for the experimental design and data analyses. We recorded a speaker telling real-life stories during fMRI scanning, and then played the recording to a group of listeners during scanning. A group ICA was conducted on the pooled data to detect the default mode network (DMN) and perisylvian language network (pLN). We first examined BOLD signal changes and network interactions within single brains. Next, we examined neural couplings between the speaker and each listener. Finally, we explored whether the between-brain coupling was affected by (or had an effect on) the task-induced (de)activation and network interactions within the listeners' brains.

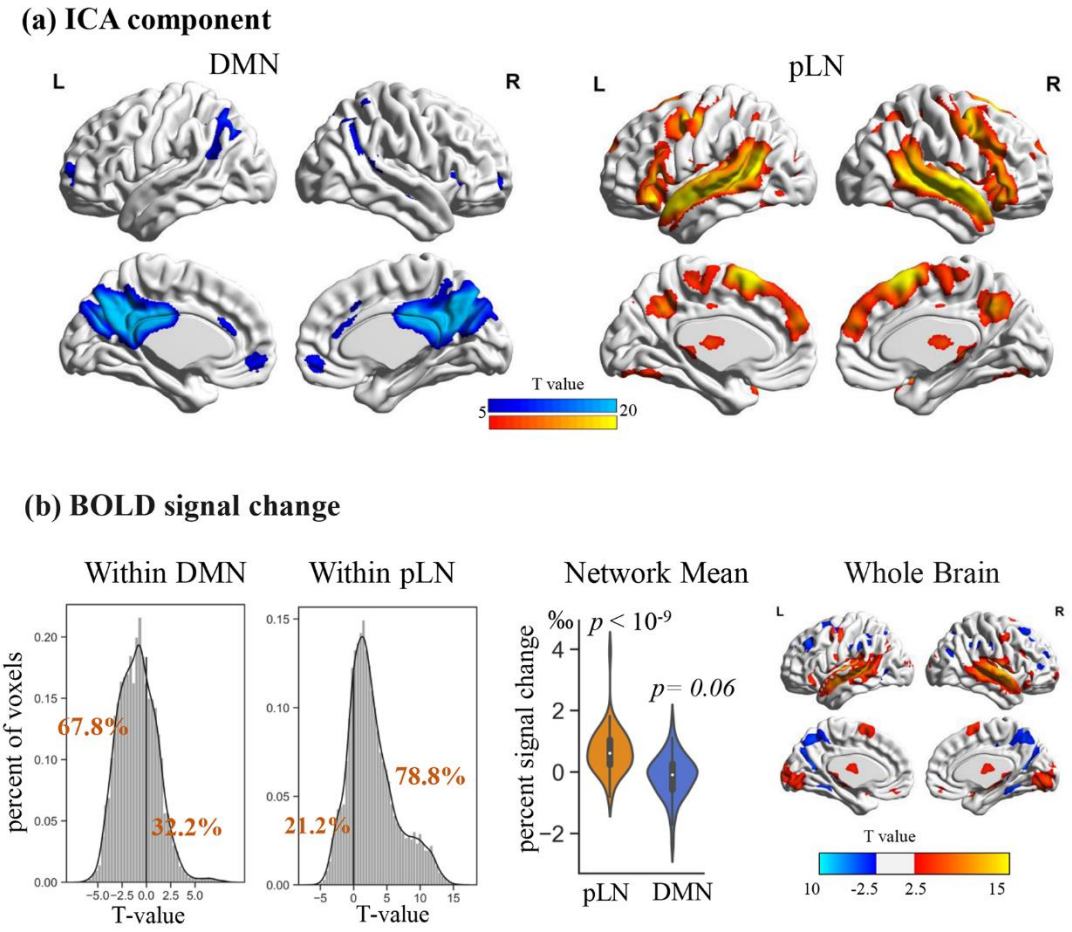


Figure 2. The default mode network (DMN) and the perisylvian language network (pLN) and their activation patterns. (a). The spatial distribution of the DMN and the pLN identified by the group ICA. Presented are the *t*-statistic maps thresholded at $p < 0.05$ using FWE correction. (b). Task-induced activations in the pLN and deactivations in the DMN, assessed by percent signal changes from the baseline (a fixation period) to the speech listening task. The histogram presents the distribution of *t*-values across all voxels within the network. The violin plot shows the distribution of percent signal change averaged over voxels within the network. The whole-brain analysis shows that significantly activated and deactivated voxels mainly fall into the pLN areas and DMN areas, respectively.

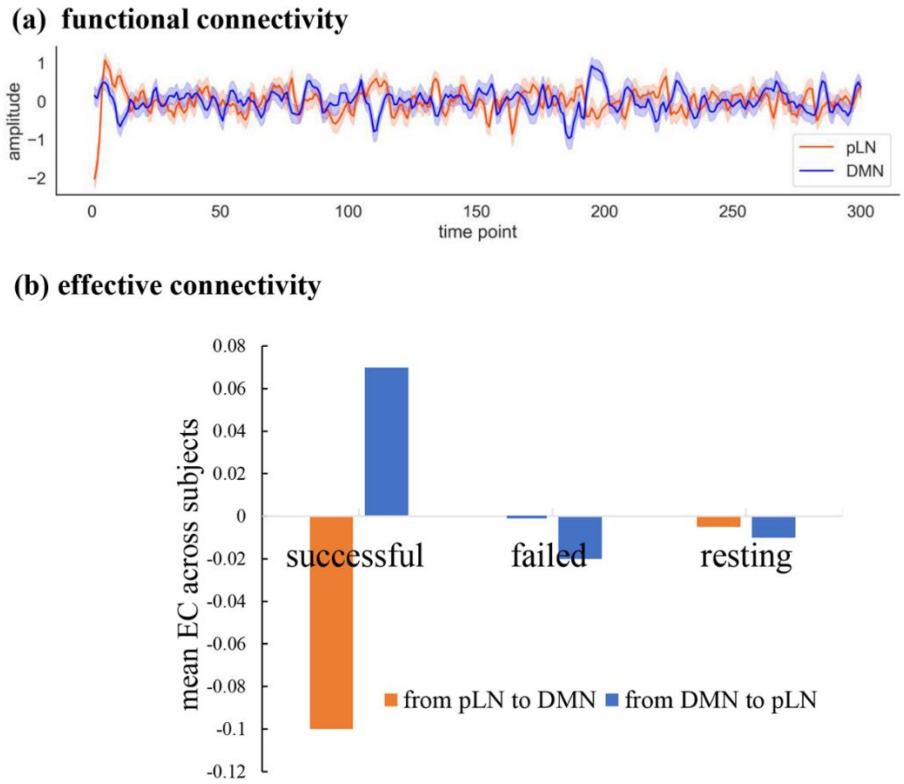


Figure 3. Within-brain functional and effective connectivity between the DMN and pLN. (a) Anticorrelation between the dynamics of DMN and pLN. The plot shows the mean and 95% confidence interval (*shaded area*) of network time series across 62 participants. (b) Effective connectivity between the two networks revealed by Dynamic Causal Modeling. During successful speech comprehension, there was a negative (indicating inhibitory) connection from the pLN to the DMN, but a positive (indicating excitatory) connection from the DMN to the pLN. This excitatory effect was absent when comprehension failed or when participants were at rest.

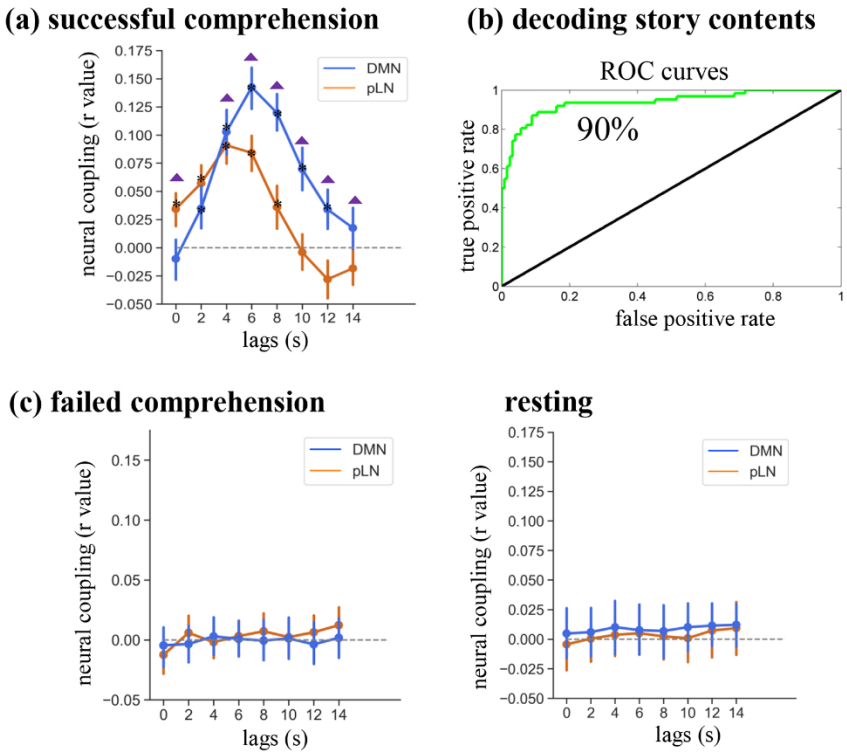


Figure 4. Neural couplings between the listener and speaker as a function of temporal lags. (a) During successful communication, the listener's network activity was coupled with the speaker's network activity with several delays. Note, the coupling in DMN was stronger than that in the pLN (except for the zero-lag activity). (b) The coupling pattern of the two networks can be used to decode which story (out of three) was played to the listener. (c) When listeners could not understand the speaker or were at a resting state, the interbrain coupling vanished. Plotted are the mean and 95% confidence interval of the neural couplings. The asterisk denotes significant differences from zero (by one-sample *t*-tests); the triangle denotes significant differences between the DMN and pLN (by paired *t*-tests).

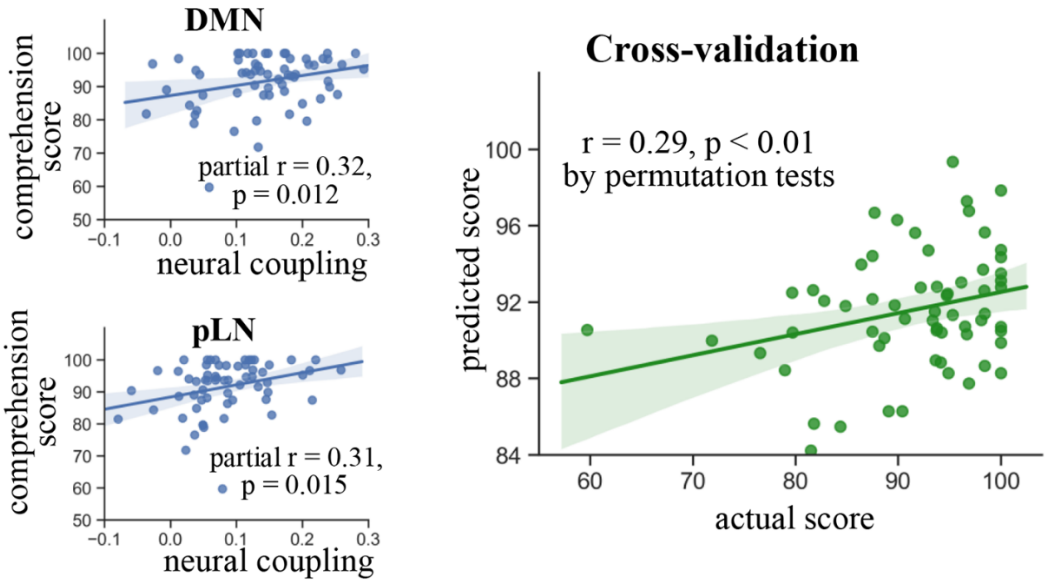
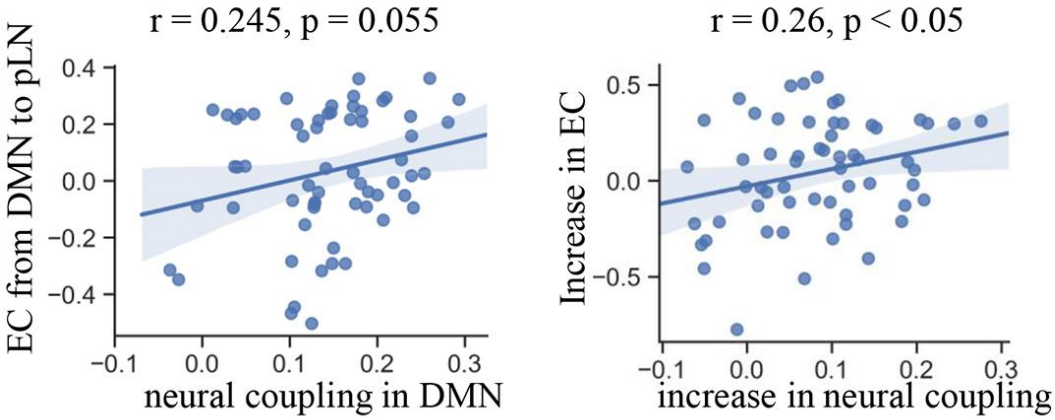


Figure 5. The strength of listener-speaker neural coupling predicts speech comprehension. Left: Correlations between speech understanding and listener-speaker neural couplings in the DMN and pLN, with memory capacity controlled. Right: Cross-validation for the brain-behavior relationship using a level-one-out procedure. In each fold of cross-validation, a regression model was built based on the data from all participants (N = 62) but one. The model was then used to predict the comprehension scores for the left-out participant based on his/her neural couplings with the speaker in the DMN and pLN.

866
867



868

869 Figure 6. Within-brain effective connection covaries with between-brain neural
870 coupling. Listeners more tightly coupled with the speaker in the DMN tended to exhibit
871 more positive DMN → pLN effective connections. Besides, the increase in DMN
872 coupling during successful relative to failed speech comprehension was positively
873 correlated with the increase in DMN → pLN connection.