

How a speaker herds the audience: multibrain neural convergence over time during naturalistic storytelling

Claire H.C. Chang^{1,2,*}, Samuel A. Nastase¹, Asieh Zadbood³, Uri Hasson¹

¹Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08540, United States

²Graduate Institute of Mind, Brain and Consciousness, Taipei Medical University, New Taipei City 235, Taiwan

³Department of Psychology, Columbia University, New York, NY 10027, United States

*Corresponding author. Graduate Institute of Mind, Brain and Consciousness, Taipei Medical University, 5F., No.301, Yuantong Rd., Zhonghe Dist., New Taipei City 235, Taiwan. E-mail: claire.hc.chang@tmu.edu.tw

Abstract

Storytelling—an ancient way for humans to share individual experiences with others—has been found to induce neural alignment among listeners. In exploring the dynamic fluctuations in listener–listener (LL) coupling throughout stories, we uncover a significant correlation between LL coupling and lagged speaker–listener (lag-SL) coupling over time. Using the analogy of neural pattern (dis)similarity as distances between participants, we term this phenomenon the “herding effect.” Like a shepherd guiding a group of sheep, the more closely listeners mirror the speaker’s preceding brain activity patterns (higher lag-SL similarity), the more tightly they cluster (higher LL similarity). This herding effect is particularly pronounced in brain regions where neural alignment among listeners tracks with moment-by-moment behavioral ratings of narrative content engagement. By integrating LL and SL neural coupling, this study reveals a dynamic, multibrain functional network between the speaker and the audience, with the unfolding narrative content playing a mediating role in network configuration.

Keywords: brain-to-brain coupling; fMRI; multibrain neural dynamics; narratives; naturalistic stimuli

Introduction

Humans use narratives to convey complex temporally structured sequences of thoughts to one another (Bruner 1986, Willems et al. 2020). This kind of communication is manifested through a process of neural “alignment” or “coupling” (Pickering and Garrod 2004, Hasson et al. 2012), whereby the speaker guides the listener(s) through a sequence of brain states to arrive at an understanding of the ideas or events the speaker intends to convey. Spoken stories have been found to drive neural alignment among listeners [listener–listener (LL) coupling] throughout the cortical language network and into higher-level areas of the default mode network (Chen et al. 2017). On the contrary, asymmetric, time-lagged coupling has been observed between the speaker and listener(s) [speaker–listener (SL) coupling] in an overlapping set of high-level cortical areas (Stephens et al. 2010, Zadbood et al. 2017, Liu et al. 2022).

The overall efficacy of a given narrative has been shown to vary across individuals. Both higher LL neural coupling and higher SL neural coupling have been separately associated with better behavioral estimates of speech comprehension across individuals (Stephens et al. 2010, Zadbood et al. 2017, Cohen et al. 2018,

Zheng et al. 2018, Davidesco et al. 2023, Liu et al. 2019, 2022, Pan et al. 2020, Meshulam et al. 2021, Nguyen et al. 2022, Zhang et al. 2022, 2022, Zhu et al. 2022, Chen et al. 2023). Individuals performing better in the post-test often showed higher neural coupling with the speaker or other listeners.

However, as the story dynamically evolves over time, it remains unclear how the SL coupling influences LL coupling. We aim to provide a novel, unified framework on how multibrain neural dynamics unfold between the speaker and listeners over time. We hypothesized that LL and SL neural coupling will tend to be correlated over time. We refer to this as the “herding” hypothesis throughout the article. Our rationale is as follows: if we consider the intersubject (dis)similarity of brain activity patterns within different cortical regions as the “distance” between subjects, SL dissimilarity reflects the distance between the speaker and listeners, while LL dissimilarity reflects the distance among listeners (Fig. 1). Under this framework, the herding hypothesis suggests that when listeners more closely follow the speaker, they tend to cluster more closely to each other; conversely, when they lose track of the speaker, they disperse in various directions (Finn et al. 2020).

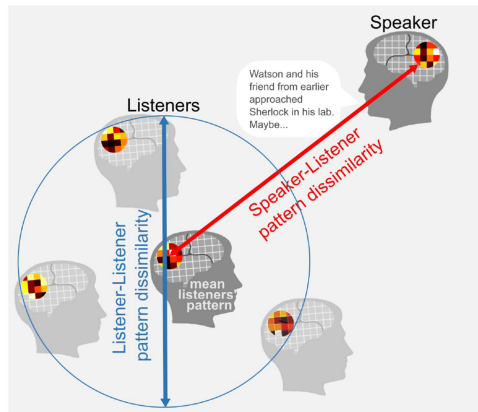
Received: 28 March 2024; Revised: 25 June 2024; Accepted: 31 August 2024

© The Author(s) 2024. Published by Oxford University Press.

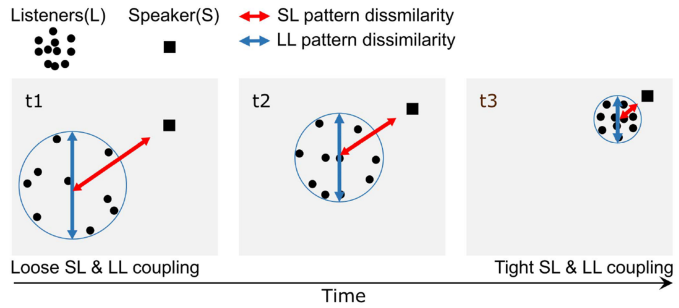
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

a). SL & LL neural pattern dissimilarity

- ↔ Speaker-Listener (SL) pattern dissimilarity
 ↔ Listener-Listener (LL) pattern dissimilarity



b). Herding hypothesis: the better the listeners follow the speaker, the closer the listeners cluster together



c). Prediction: Correlation between LL & SL (dis)similarity over time

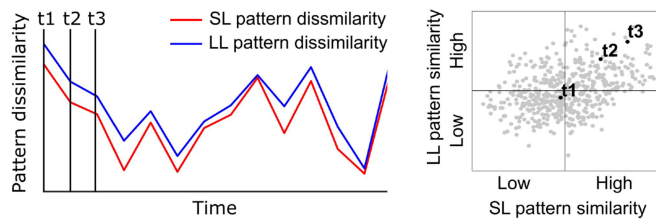


Figure 1. The herding hypothesis. (a) We use neural pattern dissimilarities to quantify the distances between the speaker and listeners (SL) and the distances among listeners (LL). (b) The herding hypothesis proposes that lag-SL pattern (dis)similarity will correlate with LL (dis)similarity over the course of a narrative. (c) In other words, over time, listeners will tend to cluster when they mirror the speaker closely, like a group of sheep guided by a shepherd. Gray points in the scatter plot correspond to pattern similarities at each time point in an example story. Note that SL coupling was computed using a lag of -10 to -1 TRs (speaker activity precedes by 15–1.5 s). See [Supplementary Fig. S1](#) for alternative hypotheses where SL and LL coupling converge and diverge in different ways.

We also hypothesized that, like a shepherd, the speaker first enacts the target brain states (capturing linguistic or narrative content). Then, the listeners re-enact similar brain states shortly after that. To focus on brain activation patterns that occur first in the speaker's brain and later in the audience's brains, we computed SL pattern dissimilarity with the speaker preceding the listeners across a window of lags ranging from -10 to -1 TR (repetition time) (1 TR = 1.5 s), based on previous studies showing that LL neural similarity peaks at lag 0, while SL neural similarity peaks within the selected lag window ([Stephens et al. 2010](#), [Dikker et al. 2014](#), [Silbert et al. 2014](#), [Zadbood et al. 2017](#), [Liu et al. 2022](#)). LL neural similarity tends to be synchronous because each listener's neural dynamics are similarly locked to the temporal structure of the speech ([Lerner et al. 2011](#), [Hasson et al. 2015](#)). On the contrary, in alignment with the flow of information in natural communication, we expected the listener's neural dynamics to lag after the speaker's neural dynamics ([Stephens et al. 2010](#)). In our analyses, we also validated the range of lags by examining peak neural similarity across an extended lag window of -10 to 10 TR ([Fig. 2](#)). The presence of a notable herding effect in this context would suggest that the listeners congregate around the trajectory outlined by the speaker's brain activity 15–1.5 s earlier in time.

From a mathematical standpoint, the fluctuating LL and lag-SL time series could be positively correlated, negatively correlated, or uncorrelated ([Supplementary Fig. S1](#)). Two related research threads motivated our hypothesis about the relationship between LL coupling and SL coupling. First, prior research has indicated that increased LL neural similarity reflects listeners converging on a shared understanding of a speech stimulus ([Cohen et al. 2018](#), [Davidesco et al. 2023](#), [Meshulam et al. 2021](#), [Zhang et al. 2022](#), [Chen et al. 2023](#)). Second, increased lag-SL neural similarity signifies that the listener's internal representation of the speech is aligning with that of the speaker ([Stephens et al. 2010](#), [Zadbood](#)

[et al. 2017](#), [Zheng et al. 2018](#), [Davidesco et al. 2023](#), [Liu et al. 2022](#), [Nguyen et al. 2022](#), [Zhang et al. 2022](#), [Zhu et al. 2022](#)). We hypothesized that these two scenarios would tend to co-occur, resulting in a positive correlation between LL and lag-SL over the course of a narrative. Alternative scenarios to the predicted positive correlation between LL and lag-SL across time points include the absence of correlation between LL and lag-SL, instances of high LL accompanied by low lag-SL, or vice versa ([Supplementary Fig. S1](#)). The discussion will address whether these scenarios are observed and explore their potential interpretations.

Aiming to understand better the multibrain neural dynamics underlying storytelling, we first verify the herding hypothesis: over time, the listeners tend to cluster when they more closely mirror the speaker and disperse in different directions when they deviate from the speaker's neural trajectory. We then use a behavioral assessment of narrative engagement to illustrate that this herding effect is strongest for brain regions where listeners are most synchronized for the more compelling moments of the story.

Materials and methods

fMRI datasets

This study employed two openly accessible auditory storytelling datasets from the “Narratives” collection (available on OpenNeuro: <https://openneuro.org/datasets/ds002245>; [Nastase et al. 2021](#)), namely, the “Sherlock” and “Merlin” datasets. The data were initially reported by [Zadbood et al. \(2017\)](#). The speaker data were obtained from a single participant across two separate experiments. The participant viewed an ~25-min movie in each experiment and was informed beforehand about a subsequent verbal recall task. After viewing the movie stimulus, they were instructed to verbally describe as many scenes as possible without any prompts. Voice recordings and functional magnetic

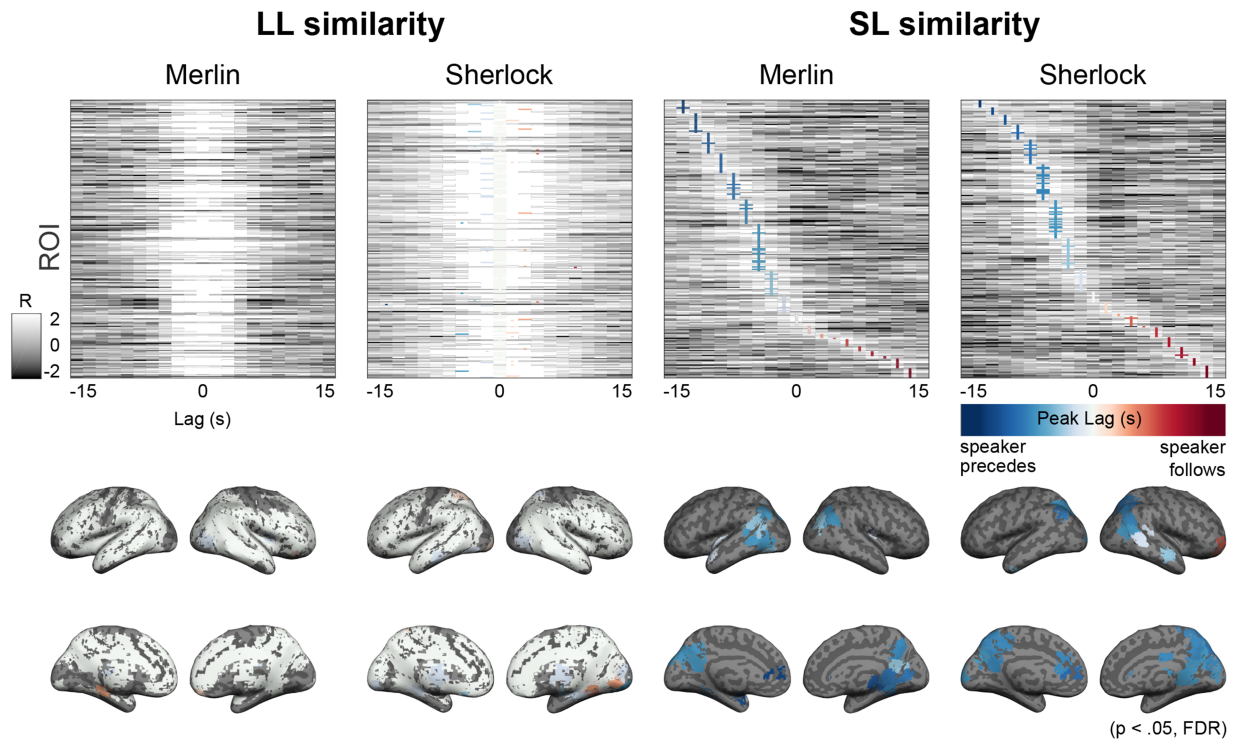


Figure 2. LL and SL neural pattern similarities. Full intersubject similarity matrixes (upper) where each row shows the neural pattern similarities in each brain region at varying lags across columns. Brain regions are ordered by their peak lags. Intersubject pattern similarities for each TR were averaged across all TRs in the story. Lags with the peak correlation values are color-coded. Significant peak lags are marked with wide (horizontal) colored bars ($P < .05$, FDR correction). Nonsignificant peak lags are marked with narrow colored bars. The correlation values are normalized with Fisher's transformation and then z-scored across lags. Region of interest (ROI)s with significant peak lags are plotted on the brain (lower). LL similarities peak at lag 0 in most brain regions, reflecting that the listeners are synchronized. On the other hand, SL similarities often peak at negative lags, indicating that the speaker precedes the listeners.

resonance imaging (fMRI) signals were simultaneously collected from the speaker. The recorded spoken recall of “Sherlock” lasts 18 min and that of “Merlin” lasts 15 min. A sample of 18 participants (11 female participants) were recruited to listen to these spoken stories. Each listener was scanned with the sole task of focusing on the provided stimuli.

All participants reported fluency in English and were 18–40 years of age. The criteria for participant exclusion have been described by Zadbood et al. (2017). All participants provided written informed consent, and the institutional review board of Princeton University approved the experimental protocol.

fMRI preprocessing

fMRI data were preprocessed using FSL the FMRIB Software Library (<https://fsl.fmrib.ox.ac.uk/>), including slice time correction, volume registration, and high-pass filtering (140s cutoff). All data were aligned to standard $3 \times 3 \times 4$ mm Montreal Neurological Institute space (MNI152). A gray matter mask was applied. The first 25 and last 20 volumes of fMRI data were discarded to remove large signal fluctuations at the beginning and end of the time course to account for signal stabilization and stimulus onset/offset before computing intersubject dissimilarities (Nastase et al. 2019). The time series in each voxel was mean centered before pattern similarity analyses. The mean responses of each ROI were computed at every time point for each participant and subtracted accordingly (Murphy et al. 2008, Garrido et al. 2013).

ROI masks

We used 238 functional ROIs defined independently by Shen et al. (2013) based on whole-brain parcellation of resting-state fMRI

data. ROIs with <10 voxels based on the coverage of our BOLD acquisition were excluded from further analyses.

SL and LL neural similarities

We computed intersubject pattern correlations in each ROI at each time point of the story, i.e. TR by TR spatial pattern similarities (Fig. 1), using the leave-one-participant-out method (Nastase et al. 2019). For LL similarity, we computed the correlation between one listener's activation pattern and the average pattern of the other 17 listeners. Similarly, SL similarity was computed between the speaker and the average pattern of 17 listeners, excluding each listener. It can be noted that quantifying SL coupling in this way entails that SL coupling can be high while LL coupling is low, i.e. listeners may be widely dispersed but roughly centered on the speaker. We also recomputed SL coupling by first computing the similarities between the speaker and each listener and then averaging these similarities. This analysis yielded qualitatively similar results. According to the literature, the speaker and listener activation patterns are not necessarily temporally synchronized (Stephens et al. 2010, Dikker et al. 2014, Silbert et al. 2014, Zadbood et al. 2017, Liu et al. 2022). Therefore, we also computed the neural similarities at varying lags.

The Pearson correlation was used to estimate pattern similarity. Time-lagged neural similarities were computed by circularly shifting the time series such that the nonoverlapping edge of the shifted time series was concatenated to the beginning or end. The resulting correlation values were normalized with the Fisher's z transformation before further statistical analyses.

We statistically evaluated the SL and LL neural similarities separately before examining the herding effect (Fig. 2). We generated surrogates with the same mean and autocorrelation as the original time series by time shifting and time reversing the functional data before computing the intersubject similarities. We computed the correlation between the original seed and time-shifted/time-reversed target time series. All possible time shifts were used to generate the null distribution. After averaging across time points and participants, the resulting correlation values were compiled into null distributions. One-tailed z-tests were applied to compare neural similarities within the window of lag -10 to $+10$ TRs against this null distribution. We corrected for multiple comparisons across lags and ROIs by controlling the false discovery rate (FDR) at $q < 0.05$ (Benjamini and Hochberg 1995).

Computing the herding metric

We defined the herding metric as the correlation between LL neural similarity at lag 0 and SL neural similarity at lags within the window of -10 to -1 TRs (i.e. the speaker precedes the listeners by 1.5–15 s). A significant herding effect indicates that the listeners are more synchronized when they echo the speaker's activation pattern. Two statistical tests were applied.

First, to verify that only the speaker showed the herding effect, we replaced the actual speaker with each of the 18 listeners to serve as the pseudo-speaker (Supplementary Fig. S2). LL similarity was computed among the remaining 17 listeners, excluding the pseudo-speaker, using the leave-one-out method, i.e. the correlation between the activation pattern from one listener and the averaged pattern of all the other 16 listeners. SL similarity was computed between the actual speaker and the average pattern of the 17 listeners. Pseudo-SL similarity was computed using the same method as the SL similarity, except that the pseudo-speaker replaced the actual speaker. We computed the herding metric with the real and pseudo-SL similarity and compared the real and pseudo-herding effects using a two-sample, one-tailed *t*-test ($N = 18$). We corrected for multiple comparisons across lags and ROIs by controlling the FDR at $q < 0.05$. Only the ROI \times SL lag combinations that passed this test were included for the second statistical test.

Second, the speaker must precede the listeners to “herd” them. Therefore, we tested the real herding effect against correlation values between LL at lag 0 and SL at all the possible lags outside of the chosen lag window (-10 to -1 TR) using one-tailed z-tests. We circularly shifted the original time series to obtain a time-lagged time series. The number of possible lags equals the number of time points. The FDR method was used to control for multiple comparisons (ROI \times SL lag; $q < 0.05$). Only ROIs that passed both statistical tests are considered to show a significant herding effect.

To quantify the amplitude of the herding effect, we extracted the peak LL–SL correlation value within the -10 to -1 TR SL lag window. We required that the peak value be larger than the absolute value of any negative peak and excluded any peaks occurring at the edges of the window. In other words, maximum correlation values occurring at lag -10 or -1 TR were not recognized as peaks.

Behavioral engagement

Engagement ratings

Behavioral assessments of dynamic engagement were acquired in another group of participants recruited via Amazon Mechanical Turk. Participants with < 20 unique rating scores (i.e. effectively flat ratings across the story) were excluded. Thirty-three raters were included for “Merlin” (15 females). A separate sample of 34 raters was included for “Sherlock” (14 females). All participants

reported fluency in English and were 25–71 years of age. All participants provided written informed consent, and the institutional review board of Princeton University approved the experimental protocol.

The participants were instructed to indicate “how engaging the current event is” while listening to the stories by moving a slider continuously. We presented the stories and collected the data using the web-based tool DANTE (Dimensional Annotation Tool for Emotions) (<https://github.com/phuselab/DANTE>) (Boccignone et al. 2017). The rating scores were acquired with a resolution of 0.04 s and then downsampled to 1.5 s ($= 1$ TR).

The engagement scores were z-scored across time, detrended, and averaged across raters.

Correlation between engagement and LL neural similarity

To quantify the relationship between time-point-by-time-point engagement ratings and LL neural similarity, we computed the Pearson correlation between the engagement scores and the LL similarity over time within ROIs, showing a significant herding effect (one-tailed). We corrected for multiple comparisons across ROIs by controlling the FDR at $q < 0.05$.

Correlation between the engagement effect and herding effect across ROIs

To quantify the relationship between engagement ratings and group-level herding, we computed the Pearson correlation between the herding effect and engagement effect across all ROIs ($P < .05$). It can be noted that since the number of ROIs is fixed, a significant *P*-value associated with this correlation does not indicate generalization to other regions (and does not support population-level inference).

Results

The herding hypothesis predicts that listeners will more closely cluster at moments of the story where they more accurately follow the speaker (Fig. 1). We quantify the distance between the speaker and listeners by computing the moment-by-moment intersubject (dis)similarity of neural activity patterns. The resulting time series of dynamic LL coupling and SL coupling indicate how tightly the listeners are clustered and aligned to the speaker, respectively. We calculate the SL dissimilarity at different lags from -15 to -1.5 s. We first verify that listener activity patterns echo those of the speaker with a lag of several seconds, i.e. SL similarities peak at negative lags (Fig. 2). We then reveal a significant herding effect in which the LL coupling correlates with the strength of lag-SL coupling in the default mode network (DMN) and language network (Fig. 3). Finally, we show higher LL coupling at moments of the story behaviorally reported as more engaging (Fig. 4a). This effect is more robust in brain regions showing a higher herding effect, such as DMN (Fig. 4b), providing behavioral evidence that the more “herded” brain regions are more synchronized by engaging story moments than the other regions.

SL and LL neural similarities: listeners follow the speaker's brain activity patterns

To visualize the relationship between SL and LL neural similarities, we first plot the total ROI \times lag intersubject similarity matrices (Fig. 2). In agreement with previous studies (Stephens et al. 2010, Dikker et al. 2014, Silbert et al. 2014, Zadbood et al. 2017, Liu et al. 2022), SL dynamics are markedly different from LL dynamics. LL similarities peak at lag 0 in most regions: listener

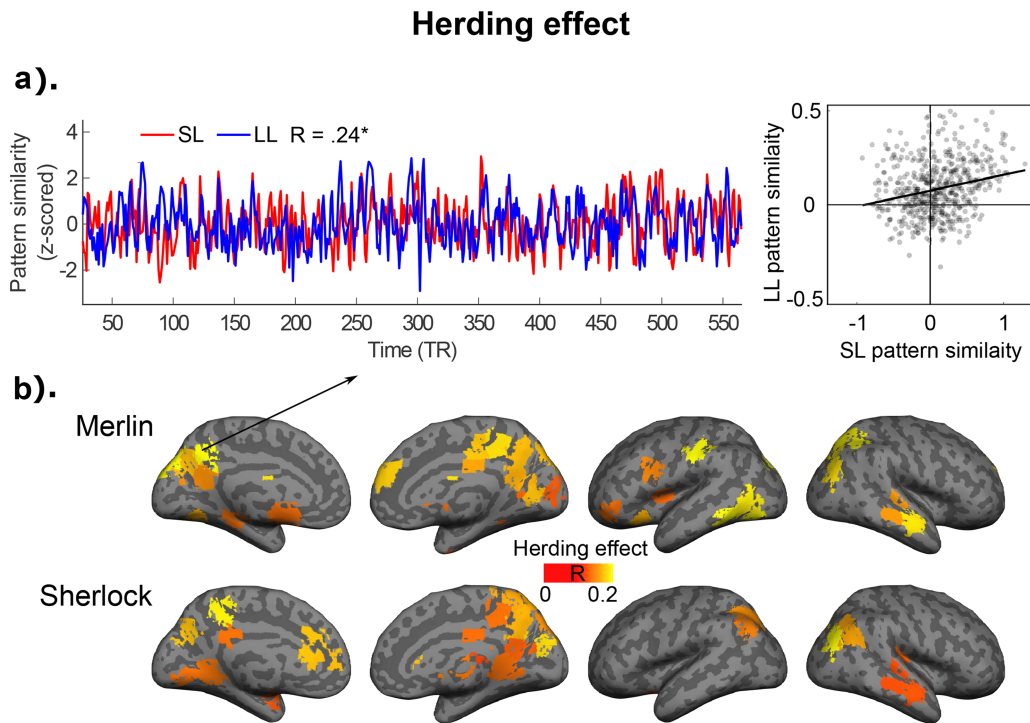


Figure 3. Cortical areas with a significant herding effect. (a) A left precuneus ROI showing a significant correlation between lag-SL and LL coupling over the course of “Merlin,” namely, a significant herding effect. (b) All ROIs with a significant herding effect. They are colored according to the amplitude of the correlation between lag-SL and LL similarity ($P < .05$, FDR correction). Lag-SL and LL similarities were normalized with Fisher’s transformation and z-scored across time before computing the herding effect.

brain activity patterns are synchronized in processing the story’s content. In contrast, significant SL similarities mainly occur at negative lags: listener activity patterns echo the speaker activity patterns with delays of several seconds. These second-long lags align with previous studies on SL coupling (Stephens et al. 2010, Dikker et al. 2014, Silbert et al. 2014, Zadbood et al. 2017, Liu et al. 2022).

Herding effect: the more accurately the listeners follow the speaker, the more tightly the listeners cluster.

The herding hypothesis predicts that the more accurately the listeners follow the speaker, the more closely the listeners cluster. We quantified the herding effect by computing the correlation between moment-by-moment LL coupling and lagged SL coupling as the narrative unfolds over time. Statistical significance for the herding effect was assessed using permutation procedures based on two surrogate datasets, one generated by replacing the speaker with a “pseudo-speaker” sampled from the listeners (Supplementary Fig. S2) and the other by applying unreasonable SL lags (i.e. the speaker precedes the listeners by >15 s or the speaker does not precede the listeners). Only ROIs in which the SL herding effect was greater for speakers rather than pseudo-speakers and with reasonable rather than unreasonable SL lags were deemed statistically significant. Given the directionality of our hypothesis, we tested the herding hypothesis with one-tailed tests.

Our results reveal a significant herding effect for both stories in the precuneus, posterior cingulate cortex, cuneus, superior and middle temporal gyrus, and superior/middle occipital gyrus (Fig. 3); many of these regions have been implicated in representing high-level events and narrative features (Chen et al. 2017, Baldassano et al. 2017, 2018, Chang et al. 2021). Using lag-SL

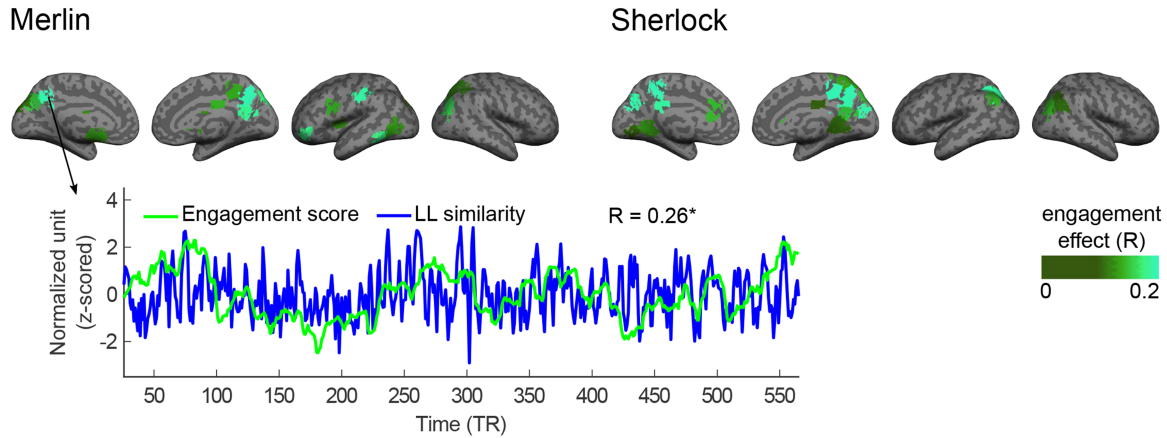
instead of SL at lag 0 constrains our analysis to neural patterns that occur first in the speaker’s brain and are observed later in the listeners’ brains. A significant lag-SL correlation aligns with the directionality of information flow from the speaker to the audience during natural communication (Fig. 2). We also compared the herding metrics based on SL coupling at lag 0 and found no significant effect (Supplementary Fig. S3). It can be noted that a significant herding effect does not provide evidence for a causal relationship between lag-SL and LL coupling.

Supplementary Fig. S4a displays an ROI with a negative correlation between LL and lag-SL revealed by *ad hoc* two-tailed tests. Supplementary Fig. S5 shows an exemplar ROI with significant LL coupling but a nonsignificant correlation between LL and the lag-SL herding effect. In addition, an alternative SL coupling metric, in which we averaged the pattern similarity between the speaker and each listener instead of averaging the pattern similarity between the speaker and the averaged listener pattern (Fig. 3), revealed a similar herding effect (Supplementary Fig. S6).

More “herded” brain regions show greater LL similarity at engaging moments of the story

To test the hypothesis that the herding effect reflects the speaker’s influence on the listeners through storytelling, we examined whether neural alignment among the listeners (i.e. LL coupling) corresponds to the level of engagement evoked by the story. To behaviorally assess how engaging the spoken narrative was moment by moment, we collected continuous engagement scores from a separate group of listeners. In agreement with a previous study (Song et al. 2021), we found that engagement scores correlate with LL neural similarity; higher LL similarity occurs at more engaging moments of a story. In a similar vein, higher neural alignment has been reported for more memorable (Simony et al.

a). Engagement effect: LL similarity correlates with engagement score



b). Higher engagement effect in brain regions showing stronger herding effect

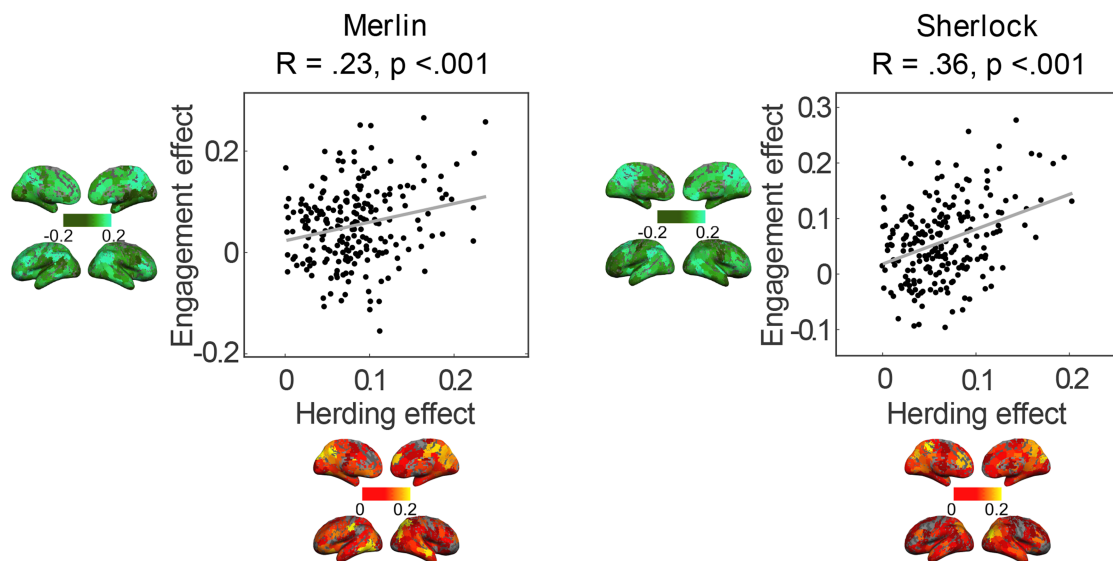


Figure 4. The engagement effect. (a) Brain regions where the behavioral engagement score significantly correlated with LL neural similarity ($P < .05$, FDR correction). In these regions, moments in the story with higher engagement ratings elicit higher LL neural similarity. The color bar indicates the magnitude of the correlation between the engagement score and LL similarity. (b) Brain regions with a stronger herding effect tend to show a stronger engagement effect. This finding provides behavioral evidence that the herding effect reflects how engaging listeners find the narrative.

2016), surprising (Brandman et al. 2021), and emotional moments during stories (Nummenmaa et al. 2014, Smirnov et al. 2019).

Among regions showing a significant herding effect, a significant engagement effect was found for both stories in the precuneus, posterior cingulate cortex, cuneus, and superior/middle occipital gyrus (Fig. 4a). More importantly, the engagement effect is more extensive in areas showing a more substantial herding effect (Fig. 4b). This finding provides behavioral evidence that the more “herded” regions are more synchronized by engaging moments of the story.

Discussion

This study examined the multibrain neural dynamics underlying storytelling. We first verified that the listeners’ neural activation patterns echoed the speaker’s neural activation patterns with a temporal lag (Fig. 2) (Stephens et al. 2010, Dikker et al. 2014, Zadbood et al. 2017, Zheng et al. 2018, Davidesco et al. 2023, Liu et al. 2022). As predicted by the herding hypothesis (Fig. 1), the

tighter the alignment between the brain activity of each listener and the speaker, the more closely the listeners clustered together (Fig. 3). We also show that LL neural similarity increases at more engaging moments of the story (Fig. 4a). This engagement effect is more substantial in the DMN (Fig. 4b), supporting the hypothesis that the “herding” effect reflects the speaker’s ability to align higher-order cognitive areas across listeners through engaging storytelling.

A significant herding effect was found in several high-order brain areas in the DMN, including the precuneus, middle/posterior cingulate cortex, lateral parietal cortex, and right antero-lateral temporal cortex (Fig. 3). The posterior medial regions, in particular, have been shown to encode paragraph-level narrative structures (Lerner et al. 2011). These regions are thought to host content-specific, supramodal event representations (Honey et al. 2012, Chen et al. 2017, Baldassano et al. 2017, Yeshurun et al. 2017, Nguyen et al. 2019, Chang et al. 2021), linking the production and comprehension of spoken narratives (Chen et al. 2017, Zadbood et al. 2017, Liu et al. 2022). The

current results add a continuous dynamic perspective to this body of work, suggesting that the speaker's neural trajectory through high-level event features may guide the listeners' upcoming event representations with varying effectiveness throughout a narrative. It is crucial to acknowledge that achieving such multibrain dynamics may not solely depend on the content of external narrative stimuli but may also depend on the internal states of individual listeners and the speaker (Yeshurun et al. 2021). For example, previous studies have demonstrated that brain-to-brain coupling may vary as a function of social closeness (Dikker et al. 2017, Bevilacqua et al. 2019) or whether the speaker and the listener share similar beliefs (e.g. similar political orientation) (Katabi et al. 2023). This dynamic convergence and divergence of idiosyncratic internal representations with the unfolding narrative would be particularly interesting for further investigation.

A significant herding effect emerges at negative SL lags on the scale of several seconds (6 s on average, ranging from 3 to 12 s across ROIs and stories). The scale of these lags is consistent with that observed in previous studies on SL coupling (Stephens et al. 2010, Dikker et al. 2014, Silbert et al. 2014, Zadbood et al. 2017, Liu et al. 2022), which have demonstrated that higher-level narrative features and event-level representations are constructed over several seconds along the cortical processing hierarchy when listening to naturalistic narratives (Chang et al. 2022). This temporally extended integrative process for narrative construction may yield SL lags on the order of seconds in higher-level cortical areas. More broadly, the current findings add to a body of work suggesting that the neural activity supporting verbal communication unfolds over surprisingly long timescales that likely reflect natural language's slow-evolving narrative and contextual structures.

We do not observe a significant herding effect with SL at lag 0 (Supplementary Fig. S3). Given that neural coupling driven by low-level auditory features would be expected to peak at lag 0, the observed lag-SL similarity seems to not be driven by auditory features (Fig. 2). The pseudo-speaker analyses (Supplementary Fig. S2)—in which we systematically substitute each listener as a “speaker” in turn—also affirm that the herding effect does not result from the speaker receiving the same auditory input as the listeners. Instead, the speaker uniquely leads the listeners through a trajectory of brain states throughout the narrative. Consequently, the listeners tend to cluster along the speaker's path, displaying varying degrees of proximity with each other—sometimes tightly, sometimes loosely—but consistently remaining a few steps behind the speaker.

The herding hypothesis illustrates a distinct form of multibrain neural dynamics, in which a neural state initially arises in the speaker's brain and then is reinstated in the listeners' brains. In our stories, most of the time, the listeners converge to trail the speaker, and when they lose track, they disperse in different directions (low LL and low lag-SL; time points in the lower left quadrant of the scatterplot in Fig. 3a). However, there are moments where LL is high despite low lag-SL (time points in the upper left quadrant of the scatterplot in Fig. 3a, also see Supplementary Fig. S1) or LL is low despite relatively high lag-SL (e.g. time points in the lower right quadrant of the scatterplot in Fig. 3a). We speculate that in the former case, the listeners might share the same misunderstanding or the speaker might not undergo the experience from the same perspective as the listeners (Sun et al. 2020), while in the latter case, the listeners might only form a loose group around the speaker due to ambiguous speech or heterogeneous apprehension (Nguyen et al. 2019).

In real-world settings, communicating effectively with audience can be a challenge, given the unique position of the speaker and the varying perspectives of the listeners. These differences will shape how the speaker and listeners move through a shared meaning space. Our herding framework provides a holistic way to quantify how well the speaker guides the audience's brains through a sequence of brain states that encode the meaning the speaker intends to convey. However, given our limited sample size, we could not test whether different multibrain dynamics systematically relate to different subjective experiences during communication. For instance, a well-rehearsed or well-scripted speaker may deliver their words by rote memorization and may not fully engage the brain systems supporting spontaneous speech. This manner of speaking would display lower neural coupling with the audience, even if the audience exhibits good comprehension of the speech and robust within-group neural synchronization. By incorporating a more diverse range of narratives, speakers, and audiences, we hope that future work will provide deeper insights into the occurrences of communication breakdowns and how they manifest in the brain. Moreover, while the whole-brain coverage of fMRI can provide valuable insights, its practical application in improving real-world communication is hindered by cost, constrained scanning conditions, and scanning noise. Future research endeavors should explore alternative neuroimaging techniques that can better capture the dynamics in everyday communication, ideally in real-time, face-to-face interactions (Redcay and Schilbach 2019).

In conclusion, the efficacy of communication with an audience is not static in real-life situations. The audience's level of engagement fluctuates over time with their neural alignment. In this study, we provide an intuitive multibrain framework for capturing the fluctuations in neural alignment induced by continuous verbal communication. We find that neural representations arising spontaneously in the speaker's brain subsequently re-emerge in the listeners' brains and that alignment among listeners coincides with their alignment to the speaker's preceding brain states. By incorporating moment-by-moment neural coupling among listeners (LL) and between the speaker and the audience (SL), this study measures the extent to which a speaker shapes the multibrain neural network as a story unfolds, guiding the audience's brains just as shepherds guide their flocks.

Author contributions

Claire H. C. Chang (Conceptualization), Asieh Zadbood (Data collection), Asieh Zadbood, Samuel A. Nastase (Data curation), Claire H. C. Chang (Software), Claire H. C. Chang (Data analysis), Claire H. C. Chang (Writing—original draft), Samuel A. Nastase, Uri Hasson (Writing – review & editing), and Uri Hasson (Funding acquisition).

Supplementary data

Supplementary data is available at SCAN online.

Conflict of interest

None declared.

Funding

This study was supported by the National Institute of Mental Health, USA (R01-MH112357 and DP1-HD091948), the National Science and Technology Council, Taiwan (113-2410-H-038-002), Taipei Medical University (TMU112-AE1-B21), and the Ministry of Education, Taiwan (Higher Education Sprout Project).

Data availability

This study relied on openly available spoken story datasets from the “Narratives” collection (OpenNeuro: <https://openneuro.org/datasets/ds002245>) (Nastase et al. 2021).

References

- Baldassano C, Chen J, Zadbood A et al. Discovering event structure in continuous narrative perception and memory. *Neuron* 2017;**95**:709–21.e5. <https://doi.org/10.1016/j.neuron.2017.06.041>
- Baldassano C, Hasson U, Norman KA. Representation of real-world event schemas during narrative perception. *J Neurosci* 2018;**38**:9689–99. <https://doi.org/10.1523/JNEUROSCI.0251-18.2018>
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B (Methodological)* 1995;**57**:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bevilacqua D, Davidesco I, Wan L et al. Brain-to-brain synchrony and learning outcomes vary by student–teacher dynamics: evidence from a real-world classroom electroencephalography study. *J Cogn Neurosci* 2019;**31**:401–11. https://doi.org/10.1162/jocn_a_01274
- Boccignone G, Conte D, Cuculo V, Lanzarotti R. AMHUSE: a multimodal dataset for humour sensing. In: *Proceedings of the 19th ACM International Conference on Multimodal Interaction. ICMI' 17*, Association for Computing Machinery, pp. 438–45. New York, NY, 2017.
- Brandman T, Malach R, Simony E. The surprising role of the default mode network in naturalistic perception. *Commun Biol* 2021;**4**:1–18. <https://doi.org/10.1038/s42003-020-01602-z>
- Bruner J. *Actual Minds, Possible Worlds*. Cambridge, MA and London, England: Harvard University Press, 1986. <https://doi.org/10.4159/9780674029019>
- Chang CHC, Lazaridi C, Yeshurun Y et al. Relating the past with the present: information integration and segregation during ongoing narrative processing. *J Cogn Neurosci* 2021;**33**:1106–28. https://doi.org/10.1162/jocn_a_01707
- Chang CHC, Nastase SA, Hasson U. Information flow across the cortical timescale hierarchy during narrative construction. *Proc Natl Acad Sci USA* 2022;**119**:e2209307119. <https://doi.org/10.1073/pnas.2209307119>
- Chen J, Leong YC, Honey CJ et al. Shared memories reveal shared structure in neural activity across individuals. *Nat Neurosci* 2017;**20**:115–25. <https://doi.org/10.1038/nn.4450>
- Chen J, Qian P, Gao X et al. Inter-brain coupling reflects disciplinary differences in real-world classroom learning. *NPJ Sci Learn* 2023;**8**:11. <https://doi.org/10.1038/s41539-023-00162-1>
- Cohen SS, Madsen J, Touchan G et al. Neural engagement with online educational videos predicts learning performance for individual students. *Neurobiol Learn Mem* 2018;**155**:60–64. <https://doi.org/10.1016/j.nlm.2018.06.011>
- Davidesco I, Laurent E, Valk H et al. The temporal dynamics of brain-to-brain synchrony between students and teachers predict learning outcomes. *Psychol Sci* 2023;**34**:633–43. <https://doi.org/10.1177/09567976231163872>
- Dikker S, Silbert LJ, Hasson U et al. On the same wavelength: predictable language enhances speaker-listener brain-to-brain synchrony in posterior superior temporal gyrus. *J Neurosci* 2014;**34**:6267–72. <https://doi.org/10.1523/JNEUROSCI.3796-13.2014>
- Dikker S, Wan L, Davidesco I et al. Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Curr Biol* 2017;**27**:1375–80. <https://doi.org/10.1016/j.cub.2017.04.002>
- Finn ES, Glerean E, Khojandi AY et al. Idiosynchrony: from shared responses to individual differences during naturalistic neuroimaging. *NeuroImage* 2020;**215**:116828. <https://doi.org/10.1016/j.neuroimage.2020.116828>
- Garrido L, Vaziri-Pashkam M, Nakayama K et al. The consequences of subtracting the mean pattern in fMRI multivariate correlation analyses. *Front Neurosci* 2013;**7**:174. <https://doi.org/10.3389/fnins.2013.00174>
- Hasson U, Chen J, Honey CJ. Hierarchical process memory: memory as an integral component of information processing. *Trends in Cogn Sci* 2015;**19**:304–13. <https://doi.org/10.1016/j.tics.2015.04.006>
- Hasson U, Ghazanfar AA, Galantucci B et al. Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn Sci* 2012;**16**:114–21. <https://doi.org/10.1016/j.tics.2011.12.007>
- Honey CJ, Thesen T, Donner TH et al. Slow cortical dynamics and the accumulation of information over long timescales. *Neuron* 2012;**76**:423–34. <https://doi.org/10.1016/j.neuron.2012.08.011>
- Katabi N, Simon H, Yakim S et al. Deeper than you think: partisanship-dependent brain responses in early sensory and motor brain regions. *J Neurosci* 2023;**43**:1027–37. <https://doi.org/10.1523/JNEUROSCI.0895-22.2022>
- Lerner Y, Honey CJ, Silbert LJ et al. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J Neurosci* 2011;**31**:2906–15. <https://doi.org/10.1523/JNEUROSCI.3684-10.2011>
- Liu J, Zhang R, Geng B et al. Interplay between prior knowledge and communication mode on teaching effectiveness: interpersonal neural synchronization as a neural marker. *NeuroImage* 2019;**193**:93–102. <https://doi.org/10.1016/j.neuroimage.2019.03.004>
- Liu L, Li H, Ren Z et al. The “two-brain” approach reveals the active role of task-deactivated default mode network in speech comprehension. *Cereb Cortex* 2022;**32**:4869–84. <https://doi.org/10.1093/cercor/bhab521>
- Meshulam M, Hasenfratz L, Hillman H et al. Neural alignment predicts learning outcomes in students taking an introduction to computer science course. *Nat Commun* 2021;**12**:1922. <https://doi.org/10.1038/s41467-021-22202-3>
- Murphy K, Birn RM, Handwerker DA et al. The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? *NeuroImage* 2008;**44**:893–905. <https://doi.org/10.1016/j.neuroimage.2008.09.036>
- Nastase SA, Gazzola V, Hasson U et al. Measuring shared responses across subjects using intersubject correlation. *Soc Cogn Affect Neurosci* 2019;**14**:667–85.
- Nastase SA, Liu Y-F, Hillman H et al. The “Narratives” fMRI dataset for evaluating models of naturalistic language comprehension. *Sci Data* 2021;**8**:250.
- Nguyen M, Chang A, Micciche E et al. Teacher–student neural coupling during teaching and learning. *Soc Cogn Affect Neurosci* 2022;**17**:367–76. <https://doi.org/10.1093/scan/nsab103>
- Nguyen M, Vanderwal T, Hasson U. Shared understanding of narratives is correlated with shared neural responses. *NeuroImage* 2019;**184**:161–70. <https://doi.org/10.1016/j.neuroimage.2018.09.010>
- Nummenmaa L, Saarimäki H, Glerean E et al. Emotional speech synchronizes brains across listeners and engages large-scale dynamic brain networks. *NeuroImage* 2014;**102**:498–509. <https://doi.org/10.1016/j.neuroimage.2014.07.063>

- Pan Y, Dikker S, Goldstein P et al. Instructor-learner brain coupling discriminates between instructional approaches and predicts learning. *NeuroImage* 2020;**211**:116657. <https://doi.org/10.1016/j.neuroimage.2020.116657>
- Pickering MJ, Garrod S. Toward a mechanistic psychology of dialogue. *Behav Brain Sci* 2004;**27**:169–90. <https://doi.org/10.1017/S0140525X04000056>
- Redcay E, Schilbach L. Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nat Rev Neurosci* 2019;**20**:495–505. <https://doi.org/10.1038/s41583-019-0179-4>
- Shen X, Tokoglu F, Papademetris X et al. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage* 2013;**82**:403–15. <https://doi.org/10.1016/j.neuroimage.2013.05.081>
- Silbert LJ, Honey CJ, Simony E et al. Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proc Natl Acad Sci USA* 2014;**111**:E4687–96.
- Simony E, Honey CJ, Chen J et al. Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat Commun* 2016;**7**:12141. <https://doi.org/10.1038/ncomms12141>
- Smirnov D, Saarimäki H, Glerean E et al. Emotions amplify speaker–listener neural alignment. *Human Brain Mapp* 2019;**40**:4777–88. <https://doi.org/10.1002/hbm.24736>
- Song H, Finn ES, Rosenberg MD. Neural signatures of attentional engagement during narratives and its consequences for event memory. *Proc Natl Acad Sci USA* 2021;**118**:e2021905118. <https://doi.org/10.1073/pnas.2021905118>
- Stephens GJ, Silbert LJ, Hasson U. Speaker–listener neural coupling underlies successful communication. *Proc Natl Acad Sci USA* 2010;**107**:14425–30. <https://doi.org/10.1073/pnas.1008662107>
- Sun B, Xiao W, Feng X et al. Behavioral and brain synchronization differences between expert and novice teachers when collaborating with students. *Brain Cogn* 2020;**139**:105513. <https://doi.org/10.1016/j.bandc.2019.105513>
- Willems RM, Nastase SA, Milivojevic B. Narratives for Neuroscience. *Trends Neurosci* 2020;**43**:271–73. <https://doi.org/10.1016/j.tins.2020.03.003>
- Yeshurun Y, Nguyen M, Hasson U. Amplification of local changes along the timescale processing hierarchy. *Proc Natl Acad Sci USA* 2017;**114**:9475–80. <https://doi.org/10.1073/pnas.1701652114>
- Yeshurun Y, Nguyen M, Hasson U. The default mode network: where the idiosyncratic self meets the shared social world. *Nat Rev Neurosci* 2021;**22**:181–92. <https://doi.org/10.1038/s41583-020-00420-w>
- Zadbood A, Chen J, Leong YC et al. How we transmit memories to other brains: constructing shared neural representations via communication. *Cereb Cortex* 2017;**27**:4988–5000. <https://doi.org/10.1093/cercor/bhx202>
- Zhang L, Xu X, Li Z et al. Interpersonal neural synchronization predicting learning outcomes from teaching–learning interaction: a meta-analysis. *Front Psychol* 2022;**13**:1–11.
- Zheng L, Chen C, Liu W et al. Enhancement of teaching outcome through neural prediction of the students' knowledge state. *Human Brain Mapp* 2018;**39**:3046–57. <https://doi.org/10.1002/hbm.24059>
- Zhu Y, Leong V, Hou Y et al. Instructor–learner neural synchronization during elaborated feedback predicts learning transfer. *J Educ Psychol* 2022;**114**:1427–41. <https://doi.org/10.1037/edu000707>