Brain & Language 212 (2021) 104878

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Understanding particularized and generalized conversational implicatures: Is theory-of-mind necessary?

Wangshu Feng [a,b], Hongbo Yu [c], Xiaolin Zhou [b,d,e,f,*]

a
b
c
d
e
f

ARTICLE INFO

Conversational implicature
Theory of mind
fMRI
tDCS
dmPFC

ABSTRACT

A speaker's intended meaning can be inferred from an utterance with or without reference to its context for particularized implicature (PI) and/or generalized implicature (GI). Although previous studies have separately revealed the neural correlates of PI and GI comprehension, it remains controversial whether they share theory-of-mind (ToM) related inferential processes. Here we address this issue using functional MRI (fMRI) and transcranial direct current stimulation (tDCS). Participants listened to single-turn dialogues where the reply was indirect with either PI or GI or was direct for control conditions (i.e., PIC and GIC). Results showed that PI and GI comprehension shared the multivariate fMRI patterns of language processing; in contrast, the ToM-related pattern was only elicited by PI comprehension, either at the whole-brain level or within dorsal medial prefrontal cortex (dmPFC). Moreover, stimulating right TPJ exclusively affected PI comprehension. These findings suggest that understanding PI, but not GI, requires ToM-related inferential processes.

1. Introduction

Imagine that Pat asks the hotel's front-desk clerk about where his friend went. The clerk responds by saying: "Some of guests are already leaving". In this conversation, the listener needs not only to decode the context-invariant "sentence meaning", but also to infer the implicated meaning (conversational implicature) beyond the literal expression (Grice, 1989; Hagoort & Levinson, 2014; Noveck & Reboul, 2008), which can be further classified into particularized conversational implicature (PI) and generalized conversational implicature (GI) (Grice, 1975). Here the utterance can convey both a GI, which is normally carried by the usage of a certain linguistic expression (e.g.,) in the utterance and can be achieved without referring to the context of utterance, and a PI, which is intimately associated with the specific context of the utterance. Specifically, the clerk's use of the term " " warrants a GI: guests are already leaving, because the clerk used a weak scalar term " " on a scale, instead of a stronger one (e.g.,). Thus, GI is independent of the particulars of its context. In contrast, in the above example, the indirect reply may convey a PI, "perhaps your friend has already left". Yet, if Pat is asking for the time, the same utterance can be interpreted as "it must be late".

In linguistic pragmatics, it is an ongoing debate, with three most influential theories, about whether interpreting GI and PI involve distinct or identical cognitive processes. Default Theory focuses on the important feature that GI is carried by the usage of certain sub-sentential locutions or structures of utterances, instead of by the particulars of the context of utterance (Chierchia, 2004; Horn, 2004; Levinson, 2000). Thus, according to this account, GI is computed by an automatic and effortless system that supports default inferences, whereas PI is computed by a separate one that supports context-sensitive inferences. In contrast, Relevance Theory holds that both types of implicature are recovered in comprehension by a single pragmatic system (Carston, 2004; Sperber & Wilson, 1986). According to this theory, understanding the speaker's meaning of an utterance is a process of searching for an optimally relevant interpretation under the particular context of the utterance, with the constraint of spending as little processing effort as possible. Once the interlocutor gets an interpretation crossing the relevance threshold, he or she will take it as what the speaker wanted to

* Corresponding author at: Institute of Linguistics, Shanghai International Studies University, Shanghai 200083, China.
xz104@pku.edu.cn (X. Zhou).

https://doi.org/10.1016/j.bandl.2020.104878
Received 30 March 2020; Received in revised form 17 August 2020; Accepted 6 October 2020
Available online 20 October 2020

convey (Sperber & Wilson, 1986). That is to say, both PI and GI are derived from the same cognitive processes, which take contextual considerations into account from the beginning. Finally, Semantic Minimalism offers a more balanced view, which draws a distinction between semantic and pragmatic processing according to whether the recovery of the content can rely on computational operations alone (Borg, 2004; Cappelen & Lepore, 2005). Although both types of implicature require information beyond the strictly semantic information at hand, GI does not constitute full pragmatic content like PI, since it can be generated only by knowing the fact that the word "some" usually contains the meaning of "some but not all" in daily conversations. Thus, GI comprehension does not recruit the holistic, general pragmatic system that is recruited to generate PI, but involves a more limited system which runs on the basis of statistical facts about what speakers have communicated in past experience (Borg, 2009). In other words, PI is derived from fully context-based inferences, while GI is derived from constraint-based inferences.

Comparing the neurocognitive mechanisms underlying PI and GI comprehension would allow us to choose between these theoretical approaches. Prior neuroimaging studies have separately investigated the neural processes of comprehending PI and GI. On the one hand, studies adopting a reading or listening comprehension task showed that the neural substrates of PI comprehension can be divided into two subsystems (Hagoort & Levinson, 2014; Hagoort, 2013): a core language network responsible for filling in the semantic gap between the literal meaning of an utterance and its context (Ferstl & von Cramon, 2001; Siebörger, Ferstl, & von Cramon, 2007), and a theory-of-mind (ToM) network, which is commonly invoked by processes of inferring mental states of other individuals. The ToM network typically consists of medial prefrontal cortex (mPFC), bilateral TPJ, precuneus, and bilateral anterior superior temporal sulcus (Mar 2011; Van Overwalle & Baetens, 2009), and among these regions, dorsal mPFC (dmPFC) and right TPJ are likely to be the core regions supporting ToM processes (Schurz, Radua, Aichhorn, Richlan, & Perner, 2014). Studies on indirect reply used natural conversations as stimulus materials, in which the reply utterance was served as a direct or indirect reply to its preceding question (Bašnáková, Weber, Petersson, van Berkum, & Hagoort, 2014; Feng et al., 2017; Jang et al., 2013; Shibata, Abe, Itoh, Shimada, & Umeda, 2011; Tettamanti et al., 2017; van Ackeren, Smaragdi, & Rueschemeyer, 2016). By comparing indirect reply to direct reply, these studies identified a set of brain regions that are linked to PI comprehension, including left (and right) inferior frontal gyrus (IFG), right (and left) middle temporal gyrus (MTG), mPFC, right (and left) TPJ, and precuneus.

On the other hand, neuroimaging studies of GI adopted a picture-sentence verification paradigm (Shetreet, Chierchia, & Gaab, 2014; Zhan, Jiang, Politzer-Ahles, & Zhou, 2017), comparing experimental conditions of mismatched generalized implicature (e.g., , following a cartoon in which all rabbits have keys), matched generalized implicature (e.g., , following a cartoon in which two of five rabbits have keys), no-implicature control (e.g., , following a cartoon in which all rabbits have keys). Participants were presented a cartoon and a sentence, and were required to decide if the sentence matched the picture. Shetreet et al. (2014) found that the mismatched and matched GI conditions commonly activated left IFG relative to a control condition. By comparing the mismatch and match GI conditions, they further found that GI mismatch activated additionally mPFC/anterior cingulate cortex and left middle frontal gyrus (MFG). The authors speculated that GI processing is possibly associated with semantic processing (IFG) and high-order cognitive functions (mPFC), like conflict control or ToM. Using similar constructions, Zhan and colleagues (2017) found that both GI mismatch and semantic mismatch activated bilateral ventral IFG, whereas GI mismatch uniquely activated left dorsal IFG and basal ganglia, relative to semantic mismatch. The activation in basal ganglia, together with IFG, suggests that the processing of GI mismatch may involve executive functions beyond semantic unification.

Although these two lines of research on PI and GI have made remarkable achievements, we still have little direct evidence for the relationship between PI and GI processing for the following reasons. First, for the studies on GI, the picture-sentence verification paradigm provides a temporary linguistic context in which a GI recovered from the sentence is inconsistent with its paired picture (in the mismatch condition). Due to the fact that implicatures could be potentially cancelled by linguistic or extra-linguistic cues (Eckardt, 2007; Grice, 1975), it is difficult to know to what extent the neurocognitive mechanism revealed in this paradigm truly reflects the system underlying GI comprehension in normal conversations. Second, these two lines of research used different experimental paradigms and different participants, making it difficult to compare the findings across PI and GI processing. Third, previous neuroimaging research used mostly univariate data analysis approaches and showed common activations in IFG and mPFC for PI and GI processing. However, such overlapping brain activity does not necessarily imply shared neural representations and cognitive processes. Thus, to what extent comprehension of PI and GI share the same neurocognitive processes is still an open question.

Here we aim to identify both the shared and distinct neurocognitive processes underlying PI and GI comprehensions by comparing these two types of conversational implicature in the same experiment. In particular, we aim to investigate whether ToM processing is necessary for interpreting both PI and GI. To this end, we adopted a a ii ss he ss

processing share the same or similar neural representation; thus we could hardly distinguish the fMRI multivariate patterns of PI and GI processing. Semantic Minimalism predicts that PI and GI processing share similar language processing systems, but distinguish in inferential processing. Accordingly, we could identify a neural pattern of language processing that responds to both PI and GI generation, and an inference-related pattern that specifically responds to PI processing.

2. fMRI experiment

Twenty-nine university students were recruited for the fMRI experiment. One participant was excluded from data analysis on the basis of binary judgment accuracy (three SDs lower than group average), leaving 28 participants for data analysis (14 females; mean age 21.5, SD = 1.9). All participants were right-handed Chinese native speakers with normal or corrected-to normal vision. None of them suffered from neurological, psychiatric, or hearing disorders. This study was approved by the Ethics Committee of the School of Psychological and Cognitive Sciences at Peking University. Written informed consents were obtained from all the participants.

We used single-turn dialogue scenarios as stimulus materials. Each dialogue scenario was comprised of three parts - a cover story, a yes/no question, and an indirect or direct reply to the preceding question (Table 1, see for pretests). In the critical conditions (i.e., PI and GI), the reply was indirectly related to the question. For the control of PI condition, namely PIC, we used the same sentence as a direct reply to the preceding question. For the control of GI condition, namely GIC, we replaced the weak scalar term (e.g.,) in the reply utterance of GI condition with its implicated meaning (e.g.,), and thus the modified utterance served as a direct reply to the question. Various pairs of scalar items were included in GI pairs to minimalize the repetition of certain lexical items, such as, “ in Chinese vs. ”, “ (vs. ”, “ vs. ”, “ vs. ”, “ vs. ”, “ vs. ”, “ vs. ”, and “ vs. ”. For each scenario, the question was strongly expected to receive a “ ” or a “ ” answer and the reply gave a definite answer. Within each pair of scenarios, both direct and indirect replies were equivalent in giving a definite answer (“ ” or “ ”) to the preceding questions. For the PI pairs, half of the replies answered “ ” to the questions while the other half answered “ ”. However, for the GI pairs, all replies would give negative answers to the questions, rendering interpreting the scalar implicature of a weak term (i.e., the stronger term is not true) necessary for understanding the speaker's meaning of the reply. For example, in Table 1, the utterance “ ” triggered a “no” answer to the question “ ”. In this case, to understand the reply, listeners need to know that the usage of “ ” warrants a GI “ ”. But, in the case that the same utterance gives a “ ” answer to the question “ ”, it is unnecessary for listeners to notice that “ ” has the meaning of “ ”.

Apart from the scenarios in the four conditions, we created filler scenarios, which were similar to the critical scenarios in form and content. For each filler scenario, the question included a stronger term. Among these filler scenarios, 20 replies with strong terms were direct answers to the preceding questions, while the other 20 replies with weak terms were indirect. We added these fillers to balance the yes/no judgment of the scenarios, and to balance the yes/no response to replies with strong/weak terms (“ ”, “ ”, “ ”, “ ” etc.), which made the materials more diversified and prevented the participants from formulating a certain response strategy.

To simulate natural conversations in daily life, all parts of dialogue scenarios were presented aurally. Fourteen Chinese native speakers were recruited to record specified parts of materials. One female and one male speaker were responsible for recording the cover stories, while six other female and six other male speakers recorded the single turn dialogues. For a particular scenario, the dialogue always occurred between a female and a male speaker. Each auditory stimulus was recorded in a sound-proof booth with a microphone (RODE NT1-A), digitized at 11.0 kHz sampling rate in a 16-bit format, and equated for the maximum sound intensity.

For fMRI scanning, participants first performed a listening comprehension task. This task was separated into two sessions, each lasting about twenty minutes. All scenarios were divided into four experimental lists based on a Latin-square design, with each list further separated into two sessions. Each list consisted of 120 scenarios, including 20 scenarios for each experimental condition (i.e., PI, PIC, GI, and GIC) and 40 fillers. Scenarios in each list were sorted pseudorandomly, such that 1) no more than three scenarios in a certain experimental condition showed up consecutively; and 2) no more than four scenarios requiring an identical response showed up consecutively. In each trial, participants experienced the following events. First, a fixation cross was presented in the middle of the screen and remained for a

quickly as possible as to whether the latter speaker really intended to answer “yes” or “no” to the question. The judgment was indicated by a button press with the index or middle finger of the participants' right hand. Reaction time (RT) was measured as the latency of his/her response to the presentation of “yes” and “no” choices.

After the listening comprehension task, participants also completed a ToM task in the scanner. Stimulus materials of this task were obtained from the Saxelab website (http://saxelab.mit.edu/localizers; credit David Dodell-Feder, Nicholas Dufour, and Rebecca Saxe), containing 10 “false belief” and 10 control stories. We first translated these stories and its corresponding statements into Chinese. Then an English-Chinese bilingual, with English as his native language, translated the Chinese version back to English. This English translation and the original version were almost identical, indicating that the Chinese version was consistent with what the English version intended to convey. For each trial, a story was visually shown for 12 s, followed by a statement about the preceding story for 4 s. Each participant made a binary judgment as to whether the statement was True or False according to the story. A fixation interval of 12 s was presented between the trials.

Prior to fMRI scanning, all participants received written instructions concerning how to complete the tasks and performed a short practice for each task. After scanning, each participant completed a Chinese version of Autism Spectrum Quotient (AQ) questionnaire which is intended to measure individuals' social skills (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). The subscale scores of this questionnaire reflect the degree of autistic-like social and communication difficulties; that is, the higher the score, the poorer the social or communication skills.

Functional images were gathered on a research-dedicated 3-Tesla MRI scanner (GE MR750, General Electric, Fairfield, Connecticut), with a T2*-weighted echo-planar imaging sequence. Each volume contained 35 transversal slices, with repetition time/echo time/flip angle = 2000 ms/30 ms/90°, slice thickness/inter-slice gap = 4 mm/0.75 mm, field of view = 192 × 192 mm^2, resolution within slice = 64 × 64, and voxel size = 3.0 × 3.0 × 4.0 mm^3. Slices of each voS egya sli scann

comparisons at cluster-level (whole-brain or within the dmPFC ROI using small-volume-correction; Chen, Jimura, White, Maddox, & Poldrack, 2015).

To identify the distributed neural representations of PI and GI processing, we used linear support vector machines (SVMs) to train multivariate fMRI pattern classifiers for PI and GI, respectively. We implemented the SVMs using Spider toolbox (https://people.kyb.tuebingen.mpg.de/spider). We trained three classifiers on individual contrast maps to discriminate PI from PIC, GI from GIC, and PI from GI. For illustration purposes, we carried out bootstrap tests to assess the significance of voxel classifier weights. We performed SVMs on 10,000 bootstrap samples (with replacement). In each voxel, two-tailed, uncorrected -value was computed according to the distribution of classifier weights. For the whole-brain analysis, the weight maps were thresholded at < 0.001 uncorrected (cluster size >10) to illustrate clusters that contributed most reliably to the classification (c.f., Wager et al., 2013). In classification and further similarity analysis, we used all the voxels in the training data. We performed a force-choice test with a leave-one-participant-out cross-validation method (cf., Chang, Gianaros, Manuck, Krishnan, & Wager, 2015; Woo et al., 2014) to calculate the classification accuracies of the SVM classifiers for PI vs. PIC and GI vs. GIC. The classifier trained to discriminate between PI and PIC (i.e., PI classifier) and the classifier trained to discriminate between GI and GIC (i.e., GI classifier) represented the neural patterns that were modifiable by PI and GI (Woo et al., 2014; Woo, Chang, Lindquist, & Wager, 2017). On the one hand, if PI and GI comprehension shared neural representations, then the PI classifier should accurately discriminate GI from GIC, and the GI classifier should accurately discriminate PI from PIC. On the other hand, if the cross-validated accuracy for classifier trained to discriminate between PI and GI was significant, there might be distinct cognitive processes between PI and GI comprehension.

Next, we used the Neurosynth Image Decoder (http://neurosynth.org/decode; Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011) to quantify the neural representation similarity between our pattern classifiers and reverse-inference maps obtained from previous studies (i.e., thousands of published neuroimaging studies included in the Neurosynth database at Jan 2017). The Pearson correlation coefficients (r) between the unthresholded weight map of PI/GI classifier and the reverse inference z-map of each of the 2911 terms in the Neurosynth database was calculated to indicate pattern similarity. Here, we focused on pattern correlations between PI/GI comprehension and 15 core concepts in psychology and psycholinguistics: , , , , , , , , , , , , , and .

Furthermore, we investigated to what extent language and ToM processing could be involved in PI and GI comprehension. Independently defined language and ToM prototypical brain patterns by the term "language" and "theory mind" in the Neurosynth database, were used to discriminate PI and GI from their respective controls. With a leave-one-participant-out cross-validation scheme, we computed the classification accuracies of the language and ToM pattern classifier for PI vs. PIC, GI vs. GIC, and PI vs. GI.

With the same procedure above, we also trained a ToM classifier to discriminate the false belief and control conditions in the ToM task both on the whole-brain and within dmPFC ROI. For the ROI analysis, predefined voxels (the number of voxels = 1038) from the conjunction analysis illustrated above were selected as training and testing data. We calculated the classification accuracies of the ToM classifiers both on the whole-brain and within dmPFC ROI for PI vs. PIC, GI vs. GIC, and PI vs. GI.

For fMRI scanning, a 2 (scenario pair: PI pair vs. GI pair) × 2 (implicature: critical vs. control) repeated-measures ANOVA for participants' task accuracy revealed a significant interaction, $(1,27) = 8.20$, $= 0.008$, $^2 = 0.23$ (see Table 2). Tests for simple effects indicated that for the GI pairs, accuracies were lower in the GI condition than in its corresponding control condition, < 0.001; this effect was not significant for the PI pairs, $= 0.08$. Trials with incorrect response or no response within the time limit (3 s) were excluded from the following behavioral and fMRI analyses.

A 2 × 2 repeated-measures ANOVA for participants' RTs revealed a significant interaction, $(1,27) = 23.69$, < 0.001, $^2 = 0.47$. Tests for simple effects indicated that for the PI pairs, RTs were longer in the PI condition than in its corresponding control condition, < 0.001; this effect was smaller for the GI pairs, $= 0.004$. To deal with the possible speed-accuracy tradeoff in PI and GI conditions, we calculated the inverse efficiency score in each condition, which consisted of the average RT of correct trials divided by accuracy (Townsend & Ashby, 1978, see Table 2). An ANOVA for inverse efficiency scores showed also a significant interaction, $F(1,27) = 8.95$, $p = 0.006$, $^2 = 0.25$: the inverse efficiency scores were larger in the PI condition than in its control; this effect was smaller for the GI pair. These findings indicated that understanding utterances with conversational implicature involves more complex pragmatic inferential processes, relative to utterances without conversational implicature, and that understanding PI seemed to be more difficult than understanding GI.

In addition, after the experiment, all fMRI participants read each scenario again and rated how indirectly the reply was related to the preceding question on a 7-point visual analog scale, ranging from "the most direct" to "the most indirect". For this after-experiment indirectness rating, a 2 × 2 repeated-measure ANOVA for rating scores showed a significant interaction, $(1,27) = 52.41$, < 0.001, $^2 = 0.66$. Tests for simple effects showed that for the PI pairs, the replies were more indirect in the PI condition than in its corresponding control condition, < 0.001; this effect was smaller for the GI pairs, < 0.001. These results suggested that the replies with conversational implicatures were considered to be more indirect than ones without such implicatures.

To identify neural correlates of PI and GI comprehension, we examined, respectively, the contrasts PI > PIC and GI > GIC at the whole-brain level. The contrast PI > PIC (Fig. 1A and Table S1 in) revealed activations in bilateral IFG, MTG, TPJ, mPFC (extending posteriorly to pre-SMA), precuneus (extending to post cingulum cortex), and bilateral MFG. The contrast GI > GIC (Fig. 1B and Table S1) revealed activations in bilateral IFG, left MTG, and mPFC/pre-SMA. Note that, after masking out the activations of the contrast GI > GIC at a voxel-level threshold < 0.01 uncorrected, the contrast PI > PIC showed activations in bilateral anterior temporal lobe, bilateral TPJ, middle mPFC, and precuneus (Fig. 1D, in blue); after masking out the activations of the contrast PI > PIC, the contrast GI > GIC showed activation in pre-SMA (Fig. 1D, in orange).

A whole-brain conjunction analysis of the contrasts PI > PIC and GI > GIC revealed clusters of activation in bilateral IFG, left MTG, and dmPFC (extending to pre-SMA), as shown in Fig. 1C and Table S1. These results indicated that the comprehension of PI and GI may involve both overlapping and distinct neural correlates.

Table 2
Mean accuracy, RT, inverse efficiency score, and degree of indirectness, and standard deviation (in parenthesis) for each condition.

Measurement	PI	PIC	GI	GIC
Accuracy (%)	93.8 (5.2)	95.9 (4.9)	89.1 (7.5)	97.1 (4.0)
RT (ms)	852 (275)	586 (235)	669 (251)	577 (249)
Inverse Efficiency (RT/Acc)	917 (320)	616 (256)	756 (299)	595 (260)
Indirectness	4.82 (0.94)	2.10 (0.53)	3.40 (1.04)	2.04 (0.77)

To identify brain regions activated by ToM processing, we examined the false belief > control contrast at the whole-brain level. This contrast evoked clusters of activation in bilateral TPJ extending inferiorly to anterior temporal gyrus, mPFC, precuneus extending to post cingulum cortex, bilateral IFG and MFG. These results are highly consistent with the ToM network identified in previous studies (Dodell-Feder et al., 2011; Lee & McCarthy, 2016). As shown in Fig. 1E, PI-specific activations (in blue) were almost completely embedded in ToM processing network identified in this study (in red).

imd)

To test the hypothesis that PI and GI processing have shared neural representations, we first trained and tested multivariate patterns at the whole-brain level. Multivariate fMRI pattern classifier trained to dissociate PI vs. PIC could discriminate PI from its control with 96% accuracy (95% confident interval (CI): 90–100%, < 0.001). When this classifier was applied to discriminate GI and its control, an accuracy approaching 100% (95% CI: 100–100%, < 0.001) was obtained. Similarly, the classifier trained to dissociate GI vs. GIC could discriminate GI condition from its control with 96% accuracy (95% CI: 90–100%, < 0.001), and could be generalized to discriminate PI vs. PIC with an accuracy of 96% (95% CI: 90–100%, < 0.001). These findings provided evidence for the existence of functionally shared neural representations for PI and GI. In addition,

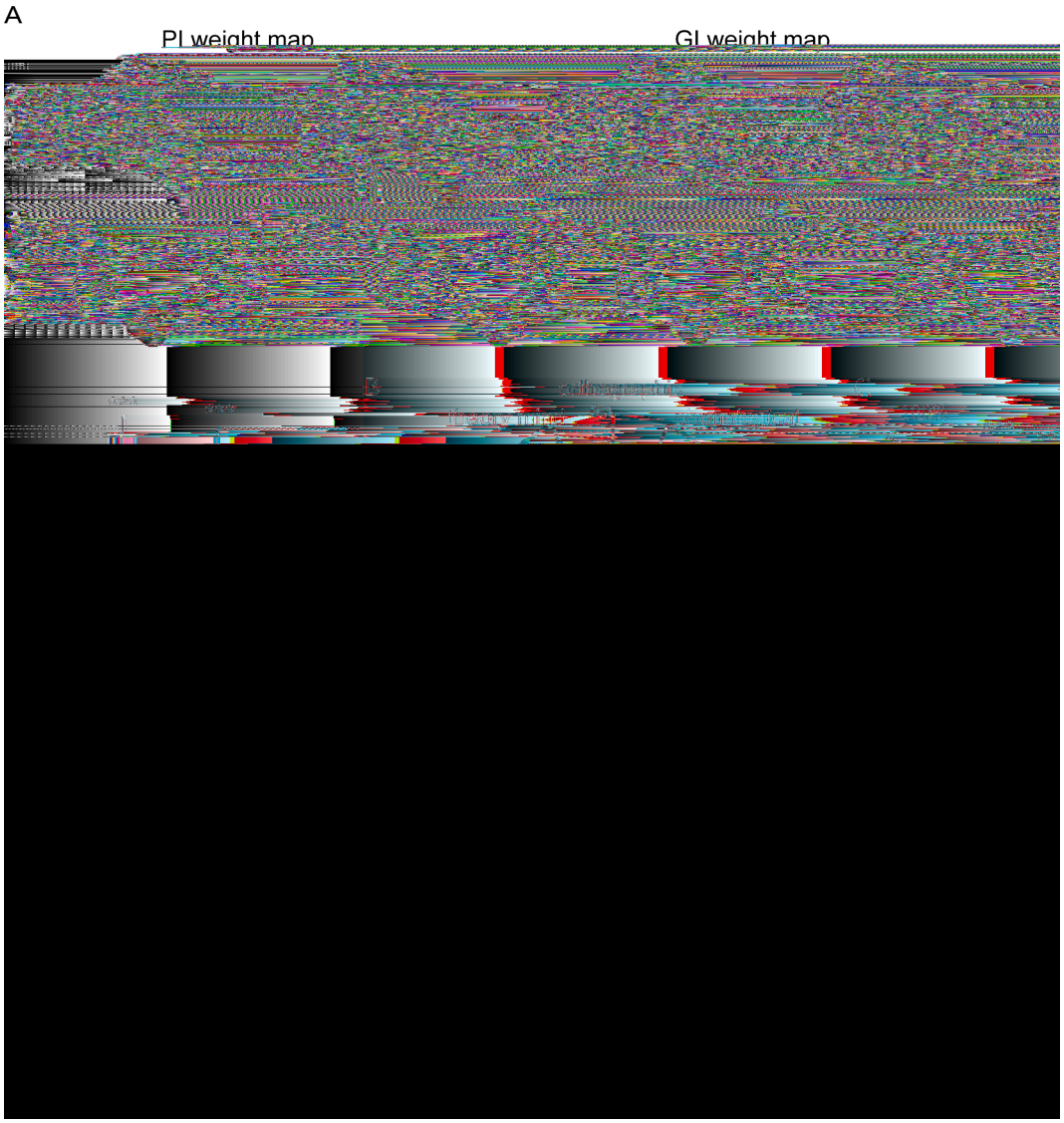

Fig. 2. Results of the whole-brain MVPA. (A) The whole-brain weight maps show voxels whose activity reliably classify PI vs. PIC conditions (i.e., PI weight map) or GI vs. GIC condition (i.e., GI weight map). Positive (warm color) and negative (cool color) weights indicate that more PI/GI processing was predicted by increased and reduced activity, respectively. (B) shows the results of neural similarity analysis using Neurosynth Image Decoder. (C) shows the accuracy of the "Language" map (left three bars) and the ToM map (right three bars) classifying PI vs. PIC, GI vs. GIC, and PI vs. PIC. Error bars represent SEs. ** < 0.01, *** < 0.001, n.s. not significant. (D) shows the prototypic language and ToM maps derived from Neurosynth database.

activated dmPFC, 58.6% voxels were also significantly activated by ToM task (see Fig. 3C). Given these seemingly contradictory findings, we further investigated whether dmPFC played an identical role in PI and GI processing.

We first hypothesized that if a "ToM" neural classifier within dmPFC could discriminate PI vs. PIC, but not GI vs. GIC, then it is reasonable for us to believe that PI and GI employed distinct neural representations in dmPFC. To test this hypothesis, we trained a "ToM" multivariate pattern within dmPFC ROI to discriminate the false belief condition and its control in the ToM task. This dmPFC ROI was obtained from the univariate conjunction analysis of the contrast PI > PIC and GI > GIC. The cross-validation test showed that this "ToM" classifier could discriminate the false belief condition from its control with 100% accuracy (95% CI: 100–100%, < 0.001). When applied to discriminate the four experimental conditions (Fig. 3A), this "ToM" classifier performed significantly above chance in discriminating both PI vs. PIC (89%, 95% CI: 79–97%, < 0.001) and PI vs. GI (86%, 95% CI: 73–96%, < 0.001). However, this classifier performed at chance level in discriminating GI vs. GIC (61%, 95% CI: 45–76%, = 0.34), consistent with the whole-brain MVPA classification. These findings provided support to the hypothesis that interpreting PI and GI has distinct neural representations within dmPFC. Specifically, the representation of PI, but not GI, may involve a ToM-related inferential component.

Secondly, we carried out univariate parametric analyses for activation in dmPFC ROI. We added the participants' social skills (as measured by AQ questionnaire; see for details) as group-level covariates for the PI > PIC and GI > GIC contrasts in two separate models. As shown in Fig. 3B, the magnitude of activation in dmPFC (peak coordinates: [9, 32, 49]; cluster size = 12; $_{FWE}$ = 0.041, small-volume corrected) negatively correlated with the social skills scores during PI processing (= −0.60, = 0.001), but not during GI processing (= 0.10, = 0.61). A direct comparison confirmed that the two correlation coefficients differed significantly, = −3.22, = 0.001, with 95% CI being [−1.05, −0.29]. These findings indicated that individuals' social skills modulated dmPFC activation during PI processing, but had no effect on GI processing.

Finally, we conducted a PPI analysis by using the dmPFC (peak coordinates: [−9, 38, 43]) as seed region. We found that dmPFC

showed significantly stronger functional interplay with several brain regions, including precentral gyrus, left inferior parietal lobule (IPL), right IFG pars opercularis and pars orbitalis (extending to right anterior insula), and pre-SMA during PI processing, relatively to GI processing (Fig. 3C and Table S2).

interpreting GI engages only weakly ToM-like inferential processing at best. Second, activation in dmPFC strongly correlated with individuals' social skills during PI processing, but not during GI processing. Third, dmPFC showed significantly stronger functional connectivity with SMA, premotor cortex, right IFG and left IPL during PI processing, relatively to GI processing. The latter pattern of frontal and parietal activity is associated with domain-general cognitive/executive control (Duncan, 2010; Ye & Zhou, 2009a, 2009b). Given that PI comprehension is generally more difficult than GI comprehension, it is reasonable to predict that PI may require additional cognitive processing to monitor and resolve the conflicts between sentential representations in discourse. Thus, the increased functional connectivity may reflect how the cognitive control system was involved in pragmatic inference during PI comprehension. Thus, a related idea is that this region is engaged in strategic inferential processing to establish the relation between utterances in discourse (Ferstl, Neumann, Bogler, & von Cramon, 2008; Ferstl & von Cramon, 2002; Kuperberg,

sham) × 2 (inference type: belief vs. control) repeated measures ANOVAs on participants' task accuracy. For the anodal experiment (Fig. 4A left panel), a marginally significant interaction between the two factors was revealed, (1,65) = 3.48, = 0.067, ^{2} = 0.05. Simple effect analysis revealed that for the sham group, the accuracy rate was lower in false belief condition (70.6 ± 2.7%) than in the control condition (81.2 ± 2.2%; < 0.001, ^{2} = 0.21); for the anodal group, there was no significant difference in accuracy between false belief condition (80.0 ± 2.6%) and control condition (83.8 ± 2.1%; = 0.14, ^{2} = 0.03). For the cathodal experiment (Fig. 4A right panel), the analysis also showed a marginally significant interaction, (1,86) = 3.81, = 0.054, ^{2} = 0.04. Simple effect analysis revealed that for the sham group, the accuracy rate was lower in false belief condition (71.7 ± 2.7%) than in control condition (81.7 ± 1.9%; = 0.001, ^{2} = 0.12). This effect was larger for the cathodal group (false belief, 65.4 ± 2.6% vs. control, 83.3 ± 1.8%; < 0.001, ^{2} = 0.33). These findings confirmed that enhancing or disrupting right TPJ functions through tDCS facilitates or hinders ToM-related inferential processes.

We then analyzed behavioral data in the listening comprehension task. For each experimental condition, participants correctly responded to more than 95% of all trials. For the anodal experiment (Fig. 4B left panel), a 2 (tDCS type: anodal vs. sham) × 2 (scenario pair: PI pair vs. GI pair) × 2 (implicature: critical condition vs. control condition) repeated measures ANOVA on participants' RTs revealed a significant three-way interaction between tDCS type, scenario pair and implicature, (1, 65) = 4.30, = 0.042, ^{2} = 0.06. Separate ANOVAs on the tDCS effect were carried out for the PI and GI scenario pairs, respectively. For the PI pair, there was a significant interaction between tDCS type and implicature, (1, 65) = 4.12, = 0.046, ^{2} = 0.06. Tests for simple effects showed that for the sham group, the RTs were longer in the PI condition (765 ± 49 ms) than in the PIC condition (583 ± 40 ms; < 0.001, ^{2} = 0.41), while this effect was much larger for the anodal group (PI, 827 ± 48 ms vs. PIC, 566 ± 40 ms; < 0.001, ^{2} = 0.59), suggesting that the anodal stimulation over right TPJ causally slowed down responses to the indirect replies with PI. For the GI pair, there was neither a main effect of tDCS type, nor an interaction between tDCS type and implicature (s < 1), indicating that the anodal brain stimulation over right TPJ could not affect GI comprehension.

The same pattern of results was obtained in the cathodal experiment (Fig. 4B right panel). The ANOVA on RT showed a significant three-way interaction, (1, 86) = 4.28, = 0.042, ^{2} = 0.05. Separate ANOVAs on the tDCS effect were carried out for the PI and GI scenario pairs. For the PI pair, there was a significant interaction between tDCS type and implicature, (1, 86) = 4.97, = 0.028, ^{2} = 0.06. Tests for simple effects showed that for the sham group, the RT was longer in the PI condition (690 ± 34 ms) than in the PIC condition (514 ± 27 ms; < 0.001, ^{2} = 0.33), and this effect was much larger for the cathodal group (PI, 793 ± 33 ms vs. PIC, 534 ± 26 ms; < 0.001, ^{2} = 0.54), indicating that the cathodal stimulation over right TPJ causally showed down responses to the indirect replies with PI. For the GI pair, there was neither a main effect of tDCS type, nor an interaction between tDCS type and implicature (s < 1.5), indicating that the cathodal brain stimulation over right TPJ could not affect GI comprehension.

To further explore the relationship between brain stimulation over right TPJ and behavioral performance on PI, we examined the indirect pathway from tDCS stimulation via ToM ability (the accuracy difference between false belief and control conditions) to PI comprehension. Results showed that the association between brain stimulation over right TPJ and PI comprehension could be mediated by ToM ability, for both anodal (the indirect effect estimate ± SE = 22.97 ± 15.77, 95% CI = [0.59, 65.25]) and cathodal (16.84 ± 13.19, 95% CI = [0.41, 57.03]) experiments (Fig. 4C). Similar analyses could not be conducted for GI comprehension, as the brain stimulation over right TPJ exhibited no effect on it.

Previous studies have consistently showed that the brain stimulation over right TPJ could causally affect ToM processing (Leloup et al., 2016; Santiesteban et al., 2012; Sowden et al., 2015; Young et al., 2010). Here, to further clarify the functions of ToM network in PI and GI comprehension by distinguishing its causal roles, we selected right TPJ region to

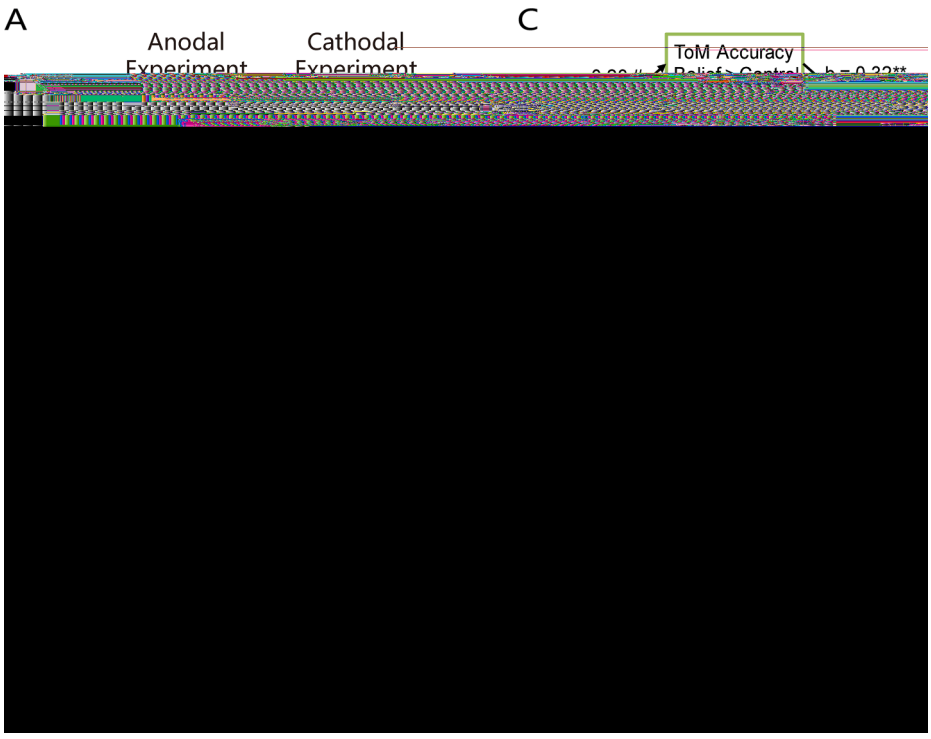

Fig. 4. tDCS results for the ToM task (A) and the listening comprehension task (B). (C) The indirect pathway from the brain stimulation over right TPJ, via ToM ability, to PI comprehension. Error bars represent between-subject SEs. # < 0.07, * < 0.05, ** < 0.01, *** < 0.001, n.s. not significant.

deliver tDCS.

First of all, results of the ToM task verified the validity of tDCS manipulation by showing that enhancing or disrupting right TPJ functions through tDCS did facilitate or hinder ToM-related inferential processes. More importantly, both anodal and cathodal stimulation causally engendered slower responses to the indirect replies with PI, and the individual's ToM

not in GI comprehension, whether at the whole-brain level or within the co-activated dmPFC region. Secondly, the tDCS experiments revealed that the brain stimulation over right TPJ could causally affect PI comprehension through its impacts upon the ToM ability, but it does not affect GI comprehension. These findings consistently indicated that the cognitive processes underlying PI and GI generation are distinct, supporting the intuitive distinction between PI and GI by Grice (1975). Thus, these findings are compatible with the accounts of either Default Theory or Semantic Minimalism. Overall, the evidence from this study suggests that compared to Default Theory and Relevance Theory, Semantic Minimalism provides more felicitous theoretical description of the cognitive processes underlying PI and GI generation and the relationship between these two types of implicatures.

Considering that we used the verbal false belief task to investigate the neural representation associated with ToM processing in the fMRI experiment and to measure the individuals' ToM ability in the tDCS experiments, one thing is noteworthy. In this ToM task, the false belief condition contains short discourses describing false beliefs, while the control condition contains discourses describing outdated photographs and maps (Dodell-Feder et al., 2011). Although this design roughly matched the domain-general inferences about outdated representations, the stimuli used in the target and the control conditions were not strictly matched in terms of linguistic variables. A recent study that matched some basic linguistic variables (such as length, sentence number, word frequency, and number of strokes per word) found that bilateral anterior superior temporal sulci and TPJ showed stronger activation in the false belief condition than the control condition from the beginning sentences of the stories, whereas the false-belief reasoning are supposed to occur only at the ending sentence of the false belief story (Lin et al., 2018). This finding indicates that these ToM-related brain activations may also reflect neurocognitive processes other than inferential processing, such as social concept retrieval. For the current study, analyses using both ToM map and language map from Neurosynth are free from the potential confounding between ToM and language processing in the ToM task. Moreover, we had reasons to believe that PI comprehension recruited ToM-like inference beyond social concept retrieval. First, in the current study, the contrast PI > PIC essentially revealed the full set of ToM network, instead of only regions linked to social concept retrieval. Second, dmPFC, which is unrelated to social concept retrieval, reflected a ToM neural pattern in understanding PI.

Finally, the role of the cognitive/executive control system in implicature comprehension is a concern. The cognitive control system, typically consisting of dmPFC, IFG, premotor cortex, and IPL, is considered to support adaptive behaviors, allowing individuals to deal with change and challenge. Previous studies indicated that the pragmatic difficulties following brain damage are due to domain-general cognitive/attentional control deficits (see Martin & McDonald, 2003). Thus, it is reasonable to predict that the cognitive control system plays a role in implicature comprehension. However, the current study did not provide strong evidence that the cognitive control system is directly engaged during either PI or GI generation. More specifically, the pattern similarity analysis with Neurosynth database did not reveal any significant correlation between PI/GI weight map and the prototypical brain patterns associated with “cognitive control” and “attention” (as shown in Fig. 2B). Nevertheless, we found that the ToM-related area (dmPFC) showed significantly stronger functional connectivity with the domain-general cognitive control network during PI comprehension, relative to GI comprehension. These findings suggest that the cognitive control system may be involved in implicature comprehension indirectly by regulating the dmPFC activity.

5. Conclusion

In this study, we identified both shared and distinct neurocognitive processes underlying PI and GI comprehension. By conducting both univariate analysis and MVPA of fMRI data, we demonstrate that PI and GI processing engage a shared language processing component, whereas the PI but not GI comprehension requires neurocognitive processes associated with ToM and intention inference. Moreover, the ROI-based fMRI MVPA and functional connectivity results suggest that the computational processes in dmPFC may rely more on knowledge of situational or social information during PI processing, relatively to GI processing. Furthermore, tDCS results provide causal evidence showing that both anodal and cathodal tDCS to right TPJ results in slower PI comprehension, but neither of them impacts GI comprehension. Our findings not only provide a deeper insight into the neurocognitive mechanisms of understanding conversational implicature, but also have broader implications for reviewing linguistic distinctions between PI and GI

Eckardt, R. (2007). Licensing 'or'. In U. Sauerland, & P. Stateva (Eds.), (pp. 34–70). Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.

Feng, W., Wu, Y., Jan, C., Yu, H., Jiang, X., & Zhou, X. (2017). Effects of contextual relevance on pragmatic inference during conversation: An fMRI study. , 52–61.

Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. (5), 581–593.

Ferstl, E. C., & von Cramon, D. Y. (2001). The role of coherence and cohesion in text comprehension: An event-related fMRI study. (3), 325–340.

Ferstl, E. C., & von Cramon, D. Y. (2002). What does the frontomedian cortex contribute to language processing: Coherence or theory of mind? (3), 1599.

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. (3), 218–229.

Friston, K. J., Holmes, A. P., Price, C. J., Büchel, C., & Worsley, K. J. (1999). Multi-subject fMRI studies and conjunction analyses. (4), 385–396.

Gavilán, J. M., & García-Albea, J. E. (2011). Theory of mind and language comprehension in schizophrenia: Poor mindreading affects figurative language comprehension beyond intelligence deficits. (1), 54–69.

Grice, H. P. (1975). Logic and Conversation. In P. Cole, & J. Morgan (Eds.), (pp. 41–58). New York: Academic Press.

Grice, H. P. (1989). . Harvard: Harvard University Press.

Hagoort, P. (2013). MUC (memory, unification, control) and beyond. (416), 1–13.

Hagoort, P., & Levinson, S. C. (2014). Neuropragmatics. In M. S. Gazzaniga (Ed.), (pp. 667–674). Cambridge, MA: MIT Press.

Horn, L. R. (2004). Implicature. In L. R. Horn, & G. Ward (Eds.), (pp. 2–28). Oxford: Blackwell.

Jang, G., Yoon, S. A., Lee, S. E., Park, H., Kim, J., & Ko, J. H. (2013). Everyday conversation requires cognitive inference: Neural bases of comprehending implicated meanings in conversations. (6), 61–72.

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: A neural prediction problem. (5), 836–848.

Krall, S. C., Rottschy, C., Oberwelland, E., Bzdok, D., Fox, P. T., Eickhoff, S. B., … Konrad, K. (2015). The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. (2), 587–604.

Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., & Holcomb, P. J. (2006). Making sense of discourse: An fMRI study of causal inferencing across sentences. (1), 343–361.

Lee, S. M., & McCarthy, G. (2016). Functional heterogeneity and convergence in the right temporoparietal junction. (3), 1108–1116.

Leloup, L., Miletich, D. D., Andriet, G., Vandermeeren, Y., & Samson, D. (2016). Cathodal transcranial direct current stimulation on the right temporo-parietal junction modulates the use of mitigating circumstances during moral judgments. , 355.

Levinson, S. C. (2000). . Cambridge, MA: MIT Press.

Lin, N., Yang, X., Li, J., Wang, S., Hua, H., Ma, Y., & Li, X. (2018). Neural correlates of three cognitive processes involved in theory of mind and discourse comprehension. & (2), 273–283.

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. (1), 103–134.

Martin, I., & McDonald, S. (2003). Weak coherence, no theory of mind, or executive dysfunction? Solving the puzzle of pragmatic language disorders. (3), 451–466.

Martin, I., & McDonald, S. (2004). An exploration of causes of non-literal language problems in individuals with Asperger syndrome. (3), 311–328.

Mason, R. A., & Just, M. A. (2011). Differentiable cortical networks for inferences concerning people's intentions versus physical causality. (2), 313–329.

Menenti, L., Petersson, K. M., Scheeringa, R., & Hagoort, P. (2009). When elephants fly: Differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. (12), 2358–2368.

Meyer, M. L., Spunt, R. P., Berkman, E. T., Taylor, S. E., & Lieberman, M. D. (2012). Evidence for social working memory from a parametric functional MRI study. (6), 1883–1888.

Monetta, L., Grindrod, C. M., & Pell, M. D. (2009). Irony comprehension and theory of mind deficits in patients with Parkinson's disease. (8), 972–981.

Muller, F., Simion, A., Reviriego, E., Galera, C., Mazaux, J.-M., Barat, M., & Joseph, P.-A. (2010). Exploring theory of mind after severe traumatic brain injury. (9), 1088–1099.

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. (3), 653–660.

Noveck, I., & Reboul, A. (2008). Experimental pragmatics: A Gricean turn in the study of language. (11), 425–431.

Peelen, M. V., & Downing, P. E. (2007). Using multi-voxel pattern analysis of fMRI data to interpret overlapping functional activations. (1), 4–5.

Piovan, C., Gava, L., & Campeol, M. (2016). Theory of Mind and social functioning in schizophrenia: Correlation with figurative language abnormalities, clinical symptoms and general intelligence. (1), 20–29.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? (2), 59–63.

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. (5), 692–697.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. (3), 879–891.

Price, A. R., Peelle, J. E., Bonner, M. F., Grossman, M., & Hamilton, R. H. (2016). Causal evidence for a mechanism of semantic integration in the angular gyrus as revealed by high-definition transcranial direct current stimulation. (13), 3829–3838.

Rapp, A. M., Mutschler, D. E., & Erb, M. (2012). Where in the brain is nonliteral language? a coordinate-based meta-analysis of functional magnetic resonance imaging studies. (1), 600–610.

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2012). Enhancing social ability by stimulating right temporoparietal junction. (23), 2274–2277.

Saxe, R., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. (8), 692–699.

Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. (2), 65–72.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. & , 9–34.

Shetreet, E., Chierchia, G., & Gaab, N. (2014). When some is not every: Dissociating scalar implicature generation and mismatch. (4), 1503–1514.

Shibata, M., Abe, J. I., Itoh, H., Shimada, K., & Umeda, S. (2011). Neural processing associated with comprehension of an indirect reply during a scenario reading task. (13), 3542–3550.

Siebörger, F. T., Ferstl, E. C., & von Cramon, D. Y. (2007). Making sense of nonsense: An fmri study of task induced inference processes during discourse comprehension. (1), 77–91.

Sowden, S., Wright, G. R. T., Banissy, M. J., Catmur, C., & Bird, G. (2015). Transcranial current stimulation of the temporoparietal junction improves lie detection. (18), 2447–2451.

Sperber, D., & Wilson, D. (1986). (Vol. 1). Harvard: Harvard University Press.

Talbert, B. (2017). Overthinking and other minds: the analysis paralysis. (6), 545–556.

Tettamanti, M., Vaghi, M. M., Bara, B. G., Cappa, S. F., Enrici, I., & Adenzato, M. (2017). Effective connectivity gateways to the Theory of Mind network in processing communicative intention. , 169–176.

Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In J. Castellan, & F. Restle (Eds.), (Vol. 3, pp. 200–239). Hillsdale, N.J.: Erlbaum.

van Ackeren, M. J., Smaragdi, A., & Rueschemeyer, S. A. (2016). Neuronal interactions between mentalising and action systems during indirect request processing. (9), 1402–1410.

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. (3), 829–858.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. (3), 564–584.

Wager, T. D., Atlas, L. Y., Lindquist, M. A., Roy, M., Woo, C. W., & Kross, E. (2013). An fMRI-based neurologic signature of physical pain. (15), 1388.

Woo, C. W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: Brain models in translational neuroimaging. (3), 365.

Woo, C. W., Koban, L., Kross, E., Lindquist, M. A., Banich, M. T., & Ruzic, L. (2014). Separate neural representations for physical pain and social rejection. (5380), 1–12.

Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: Emergent features of word, sentence, and narrative comprehension. (3), 1002–1015.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. (8), 665–670.

Ye, Z., & Zhou, X. (2009a). Conflict control during sentence comprehension: fMRI evidence. , 280–290.

Ye, Z., & Zhou, X. (2009b). Executive control in language processing. & (8), 1168–1177.

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. (15), 6753–6758.

Zhan, J., Jiang, X., Politzer-Ahles, S., & Zhou, X. (2017). Neural correlates of fine-grained meaning distinctions: An fMRI investigation of scalar quantifiers. (8), 3848–3864.

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. (4), 399.