Received: 18 December 2019 | Revised: 14 August 2020 | Accepted: 19 August 2020

DOI: 10.1002/hbm.25203

RESEARCH ARTICLE

WILEY

The neural basis for mental state attribution: A voxel-based lesion mapping study

Shira Cohen-Zimerman[1,2] | Harsh Khilwani[1,3] | Gretchen N. L. Smith[1] | Frank Krueger[4,5] | Barry Gordon[6,7] | Jordan Grafman[1,2,8]

[1]Cognitive Neuroscience Laboratory, Brain Injury Research, Shirley Ryan AbilityLab, Chicago, Illinois

[2]Departments of Physical Medicine and Rehabilitation, Feinberg School of Medicine, Northwestern University, Chicago, Illinois

[3]Department of Biomedical Engineering, Northwestern University, Chicago, Illinois

[4]School of Systems Biology, George Mason University, Fairfax, Virginia

[5]Department of Psychology, University of Mannheim, Mannheim, Germany

[6]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland

[7]Department of Cognitive Science, Johns Hopkins University, Baltimore, Maryland

[8]Department of Neurology, Psychiatry, and Cognitive Neurology & Alzheimer's Disease, Feinberg School of Medicine, Department of Psychology, Northwestern University, Chicago, Illinois

Correspondence
Shira Cohen-Zimerman,
Cognitive Neuroscience Laboratory,
Brain Injury Research, Shirley Ryan AbilityLab,
Chicago, IL.
Email: shira.cohen-zimerman@northwestern.edu

Note: Questions regarding the Vietnam Head Injury Study can be directed to Dr. Jordan Grafman.
E-mail: jgrafman@northwestern.edu

Funding information
Smart Family Foundation of New York; Therapeutic Cognitive Neuroscience Fund

Abstract

The ability to infer other persons' mental states, "Theory of Mind" (ToM), is a key function of social cognition and is needed when interpreting the intention of others. ToM is associated with a network of functionally related regions, with reportedly key prominent hubs located in the dorsolateral prefrontal cortex (dlPFC) and the temporoparietal junction (TPJ). The involvement of (mainly the right) TPJ in ToM is based primarily on functional imaging studies that provide correlational evidence for brain-behavior associations. In this lesion study, we test whether certain brain areas are necessary for intact ToM performance. We investigated individuals with penetrating traumatic brain injury (n = 170) and healthy matched controls (n = 30) using voxel-based lesion-symptom mapping (VLSM) and by measuring the impact of a given lesion on white matter disconnections. ToM performance was compared between five patient groups based on lesion location: right TPJ, left TPJ, right dlPFC, left dlPFC, and other lesion, as well as healthy controls. The only group to present with lower ToM abilities was the one with lesions in the right dlPFC. Similarly, VLSM analysis revealed a main cluster in the right frontal middle gyrus and a secondary cluster in the left inferior parietal gyrus. Last, we found that disconnection of the left inferior longitudinal fasciculus and right superior longitudinal fasciculus were associated with poor ToM performance. This study highlights the importance of lesion studies in complementing functional neuroimaging findings and supports the assertion that the right dlPFC is a key region mediating mental state attribution.

KEYWORDS

connectivity, strange stories test, temporoparietal junction, theory of mind, voxel-based lesion-symptom mapping

1 | INTRODUCTION

Over the first 6 years of our lives, we typically develop an ability to attribute intentions, beliefs, and desires to other people. This important set of skills, known as Theory of mind or ToM, allows us to understand other people's actions, to predict behavioral responses and to separate our own mental state from that of others (Amodio & Frith, 2006; Apperly, 2012). Given its important role in facilitating meaningful social interactions, it is not surprising that ToM is studied extensively, with a growing focus on its neural basis (Abu-Akel & Shamay-Tsoory, 2011; Frith & Frith, 2006; Young et al., 2010).

A widely accepted theoretical distinction separates affective ToM (i.e., the ability to infer others' emotional states and feelings) from cognitive ToM (i.e., the ability to infer others' beliefs, intentions, and desires), and there is evidence to show that the two differ in their underlying neural network (Corradi-Dell'Acqua et al., 2020; Leopold et al., 2012; Shamay-Tsoory, Tibi-Elhanany, & Aharon-Peretz, 2006; Shamay-Tsoory & Aharon-Peretz, 2007). Cognitive ToM, which is the focus of this study, is associated with a number of functionally related regions as would be expected given the complexity and heterogeneity of this ability (Gallagher & Frith, 2003; Saxe, 2006). Specifically, regions in the prefrontal cortices as well as the temporoparietal junction were shown to have a specific link to this ability (Gallagher et al., 2000; Krall et al., 2015; Molenberghs, Johnson, Henry, & Mattingley, 2016; Van Overwalle, 2009; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014).

The prefrontal cortex (PFC) has long been considered to play a special role in human social behavior in general (Amodio & Frith, 2006; Forbes & Grafman, 2010; Krueger, Barbey, & Grafman, 2009) and in ToM abilities in particular (Gallagher et al., 2000). More specifically, there is data to suggest that the dorsal-lateral part of the PFC (dlPFC) is associated with the cognitive processes involved in ToM (Carrington & Bailey, 2009). For example, neuroimaging studies linked dlPFC activation to integrating social information for impression formation (Brosch, Schiller, Mojdehbakhsh, Uleman, & Phelps, 2013), and to thinking about what other people are thinking (Kobayashi, Glover, & Temple, 2007). Moreover, neuromodulation work has shown that repetitive transcranial magnetic stimulation (Kalbe et al., 2010) over the right dlPFC induced a selective effect on mental state attribution. Similarly, the dlPFC was identified as essential for mentalizing processes using direct electrical stimulation during awake brain surgery (Yordanova, Duffau, & Herbet, 2017), and was shown to be functionally coupled with other mentalizing-related sites (Yordanova, Cochereau, Duffau, & Herbet, 2019). Last, several lesion mapping studies reported impairments in understanding and predicting other peoples' thoughts and intentions following damage to the dlPFC (Corradi-Dell'Acqua et al., 2020; Shamay-Tsoory & Aharon-Peretz, 2007; Xi et al., 2011).

The second prominent hub, the temporoparietal junction (TPJ), is localized at the conjunction of the posterior superior temporal sulcus, the inferior parietal lobule and the lateral occipital cortex. This region is frequently reported to be selectively activated in imaging studies during tasks that require attribution of mental states, beliefs and intentions to others. The activation is often reported to be predominantly in the right hemisphere (rTPJ) (Kubit & Jack, 2013; Otti, Wohlschlaeger, & Noll-Hussong, 2015; Perner, Aichhorn, Kronbichler, Staffen, & Ladurner, 2006; Saxe, 2010; Saxe & Kanwisher, 2003; Scholz, Triantafyllou, Whitfield-Gabrieli, Brown, & Saxe, 2009), although some studies report activation in the left TPJ (lTPJ) as well (Perner et al., 2006; Young, Dodell-Feder, & Saxe, 2010). Moreover, one study used transcranial direct current stimulation (tDCS) to show that ToM performance decreased after inhibitory cathodal stimulation to the rTPJ (Mai et al., 2016). Based on these findings, some have argued that the right TPJ is specifically linked to the process of mental state attribution. While there are no studies reporting ToM deficits in patients with a focal lesion to the rTPJ, there are a few studies reporting such deficits in patients with a focal lesion to the lTPJ, implying that this region is not only associated with, but necessary for, ToM performance (Apperly, Samson, Chiavarino, & Humphreys, 2004; Biervoye, Dricot, Ivanoiu, & Samson, 2016; Samson, Apperly, Chiavarino, & Humphreys, 2004). However, the ToM tasks used in these latter studies included nonverbal videos and not the standard stories typically used in most ToM tasks.

Both lesion mapping and functional neuroimaging studies suggest an overall laterality effect with ToM being linked to the right hemisphere (Baldo, Kacinik, Moncrief, Beghin, & Dronkers, 2016; Happé, Brownell, & Winner, 1999; Kalbe et al., 2010; Sommer et al., 2007; Stuss, Gallup Jr, & Alexander, 2001; Weed, McGregor, Feldbæk Nielsen, Roepstorff, & Frith, 2010; Winner, Brownell, Happé, Blum, & Pincus, 1998). Yet, earlier PET (positron emission tomography) studies found activation in the left PFC during tasks that required participants to consider the thoughts and feelings of another person (Fletcher et al., 1995), to make inferential reasoning about the beliefs and intentions of others (Goel, Grafman, Sadato, & Hallett, 1995), and to respond to stories which require mental state attribution (Happé et al., 1996).

Altogether, despite the ample research on the neural underpinning of ToM, it is still unclear whether the dlPFC and the TPJ in the right or left hemisphere are necessary for cognitive ToM, or play an indirect role that can be compensated for in the case of a brain lesion. The reason for this ambiguity is that most of the data associating ToM performance to the TPJ and to frontal brain regions is based on functional brain imaging and thus cannot determine whether a particular brain region is necessary for a specific function. While a brain lesion mapping approach can reliably establish a causal link between ToM performance and a brain area, most lesion-mapping studies of ToM have focused on one patient group at a time with damage to one specific brain area and did not compare different lesion locations within the same experimental design. Only one study has attempted to assess the relative performance of different lesion groups using an established ToM task, finding no differences in cognitive ToM performance between groups (Shamay-Tsoory et al., 2006). However, the relatively small number of patients in each lesion group (range 5–14), and the grouping together of the right and left TPJ and dlPFC patients might have contributed to the null finding.

In the current study, we tested the causal contribution of specific brain areas (the two key brain regions discussed above in the right and left hemisphere, namely the right dlPFC, the left dlPFC, the right TPJ, and the left TPJ) to cognitive ToM performance following penetrating traumatic brain-injury (pTBI). We examined a large group of veterans from the Vietnam Head Injury Study (VHIS) who sustained focal pTBIs during combat (n = 170) and matched controls (n = 30) who also

served in combat in Vietnam but did not have a brain injury. We used a voxel-based lesion-symptom mapping (VLSM) analysis to explore the causal role of focal brain lesions on mental state attribution. Given that VLSM is not free of limitations (Mah, Husain, Rees, & Nachev, 2014), we chose to complement our analysis using a network level approach, in order to reveal specific white matter tracts which support mental state attribution. Participants were asked to read and respond to a set of stories about everyday situations. Half of the stories required mental state attribution to understand the meaning of the scenario while the other half of the stories served as control stimuli focusing on the physical characteristics of the story (strange stories test; Happé, 1994). We compared ToM performance in the four lesion groups mentioned above as well as a group of pTBI patients with lesions in areas other than the TPJ or dlPFC. We hypothesized that patients with lesions in each of the target regions (the right and left dlPFC and the right and left TPJ groups) will all be more likely to show impaired performance on a cognitive ToM task compared to patients with lesions in other brain areas or healthy controls. We also hypothesized no differences in ToM performance among the four target patient groups.

TABLE 1 Demographics and neuropsychological measures [mean (SD)] for veterans with pTBI and healthy controls

Variables/group	pTBI *n* = 170	Control *n* = 30
Demographics:		
Age (years)	57.92 (2.46)	58.57 (2.02)
Education (years)	14.61 (2.47)	14.70 (2.59)
Handedness (R:L:A)[a]	141:23:6	25:4:1
Neuropsychological:		
Preinjury IQ[b]	60.48 (25.43)	68.66 (22.09)
Theory of Mind[c]	0.24 (3.16)	−0.33 (3.97)
Working memory[d]	96.47 (14.72)	104.73 (13.54)*
Verbal comprehension[e]	106.05 (15.55)	110.33 (9.84)

Note: * denotes significant group difference $p < .05$.
Abbreviations: pTBI, penetrating traumatic brain-injury; WAIS, Wechsler Adult Intelligence Scale.
[a]Handedness (L:R:A), Left, right, and ambiguous.
[b]Percentile score of Armed Forces Qualification Test (AFQT).
[c]Strange stories task: difference score between ToM and Control condition. Lower score reflects lower ToM performance.
[d]WAIS Working Memory Index score.
[e]WAIS Verbal Comprehension Index score.

2 | METHODS

2.1 | Participants

Participants were male combat veterans who participated in Phase 3 of the VHIS (Raymont, Salazar, Krueger, & Grafman, 2011). This phase was conducted between 2003 and 2006 at the National Naval Medical Center, Bethesda, MD, and included detailed neuropsychological, neuroimaging, neurological and psychiatric evaluations for each one of the study participants. In total, we collected data from 170 patients with pTBI and 30 control participants who also served in combat in Vietnam but had no history of brain injury or other neurological disorders. The groups were matched on age, years of education, preinjury intelligence, and handedness (Table 1).

While this is the first study to investigate cognitive ToM using the VHIS database, our group has previously published data on the neural underpinnings of affective theory of mind using the reading the mind in the eyes test (Dal Monte et al., 2014) and the Faux Pas Recognition task (Leopold et al., 2012) based on this registry.

All participants understood the study procedures and gave written informed consent, as approved by the National Institutes of Health Neuroscience Institutional Review Board, Bethesda Naval Hospital and Department of Defense Institutional Review Boards. The Institutional Review Board at Northwestern University approved the current analysis of the data.

2.2 | Materials

2.2.1 | Theory of mind

Theory of mind was measured using the strange stories test (Happé, 1994). Each participant was presented with 16 stories, eight ToM stories and eight control stories, which were selected from the 24 original stories and had been used in prior imaging and neuropsychological studies (Fletcher et al., 1995; Happé et al., 1999).

Both sets of stories involved people and required attention to sentence meaning, memory, and question answering, however only the ToM questions were based on understanding the beliefs and intentions of characters in the stories, while the control story questions were based on physical inferences made about the story. Stories are of comparable difficulty in healthy young adults (White, Hill, Happé, & Frith, 2009).

The following story represents an example of a selected ToM story (story number 21): "Simon is a big liar. Simon's brother Jim knows this; he knows that Simon never tells the truth! Now, yesterday, Simon stole Jim's ping-pong bat, and Jim knows Simon has hidden it somewhere, though he can't find it. He is very cross. So he finds Simon and he says, 'Where is my ping-pong bat? You must have hidden it either in the cupboard or under your bed, because I've looked everywhere else. Where is it, in the cupboard or under your bed?' Simon tells him the bat is under his bed. Q: Why will Jim look in the cupboard for the bat?"

The following story represents an example of a selected physical (control) story (story number 12): "A burglar is about to break into a jewelers' shop. He skillfully picks the lock on the shop door. Carefully he crawls under the electronic detector beam. If he breaks this beam it will set off the alarm. Quietly he opens the door of the storeroom and sees the gems glittering. As he reaches out, however, he steps on something soft. He hears a screech and something small and furry runs out past him, towards the shop door. Immediately the alarm sounds. Q: Why did the alarm go off?"

All eight stories of each type were administered as a group, but the order of the two sets was counterbalanced among participants. For each story, participants were instructed to read the story and then answer a question on a separate page. Participants received two

points for each fully explicit correct answer, one point for partial, implicitly correct answers, and zero points for an incorrect answer or no response. Two scores were calculated for each participant: (a) ToM story score: the sum of the scores for each ToM story question, range 0–16 and (b) physical story score: the sum of the scores for each physical story question, range 0–16. The primary outcome in the current analysis was the difference between these two scores, with zero reflecting no difference in performance on the ToM and Physical stories, and a negative score reflecting a lower score on the ToM stories compared to the physical stories condition.

2.2.2 | Control measure—Space perception

To ensure that our key results were specific to ToM, we compared the groups on the dot counting test, a subtest from the Visual Object and Space Perception (VOSP) Battery measuring space perception. In this dot counting test, the patient is asked to count how many black dots there are on a white card. There are 10 cards and a point is awarded for every correct count, with the maximum score being 10.

2.2.3 | Additional neuropsychological testing

Other neuropsychological tests examined in this study included the Armed Forces Qualification Test (AFQT- 7A, 1960), a standardized test which is highly correlated with Wechsler Adult Intelligence Scale (WAIS) scores and hence used as a surrogate for IQ (Cohen-Zimerman, Salvi, Krueger, Gordon, & Grafman, 2018; Grafman et al., 1988). Preinjury AFQT scores were obtained from all participants upon enlistment in the military. During Phase 3 of the study (when ToM abilities were assessed), the AFQT was readministered, as well as the WAIS-III (Wechsler, 1997). Given that the ToM abilities were shown to covary with working memory (Gokcen, Bora, Erermis, Kesikci, & Aydin, 2009) and verbal comprehension abilities (White et al., 2009), the WAIS Working Memory Index (WMI) and Verbal Comprehension Index (VCI) were calculated for each participants and later used as covariates. The WMI includes the arithmetic and digit span subtests from the WAIS, while the VCI includes the information, similarities, and vocabulary subtests.

2.3 | Neuroimaging assessment and image preprocessing

Axial computerized tomography (CT) scans without contrast were acquired using a GE Medical Systems Light Speed Plus CT scanner at the Bethesda Naval Hospital. Magnetic resonance imaging (MRI) could not be performed with patients in this study due to the possible presence of metal fragments from shrapnel or bullet wounds, or residual metallic surgical clips or cranioplasties from surgery. Images were reconstructed with an in-plane voxel size of 0.4×0.4 mm, an overlapping slice thickness of 2.5 mm and a slice interval of 1 mm. We determined lesion location and volume from CT images using the Analysis of Brain Lesion (ABLe) software (Solomon, Raymont, Braun, Butman, & Grafman, 2007) contained in MEDx v3.44 (Medical Numerics, Germantown, MD) with enhancements to support the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). A trained neuropsychiatrist manually traced individual lesions, which were then reviewed by a researcher who was blind to the results of the Phase 3 evaluation (JG). Scans were spatially normalized to Montreal Neurological Institute MNI space (Collins, Neelin, Peters, & Evans, 1994) using the Automated Image Registration program (Woods, Mazziotta, & Cherry, 1993) using a 12-parameter affine model on de-skulled CT scans. We did not include voxels with lesions in the spatial normalization procedure in order to reduce image distortions. Lesion volume was calculated by summing the traced areas in all relevant slices of the CT image, and then and multiplying by slice thickness.

2.4 | Statistical analyses

2.4.1 | Lesion localization and grouping

In order to test our hypothesis about four specific brain areas [dlPFC (r + l) and TPJ (r + l)], we identified percent volume loss to each region as follows: The right and left dlPFC was defined using the AAL within a range of MNI coordinates (Gozzi, Raymont, Solomon, Koenigs, & Grafman, 2009). The dlPFC region of interest (ROI) included bilateral portions of the superior frontal gyrus (dorsolateral), the middle frontal gyrus (lateral), and the inferior frontal gyrus (triangular part), based on the following MNI coordinates: $x > 10$ (right), $x < -10$ (left), $z > 1$. Percentage of these AAL structures that were intersected by the lesion was determined again by analyzing the overlap of the spatially normalized lesion image with the AAL atlas image. The right and left TPJ were defined based on a 3D threshold map (Dufour et al., 2013). For each patient, we analyzed the overlap of the TPJ maps with their spatially normalized lesion image to calculate the percent of volume loss in each region separately.

Participants with damage primarily to the rTPJ ($n = 16$), left TPJ ($n = 7$), r dlPFC ($n = 30$) or l dlPFC ($n = 28$) were identified. Thirty-four veterans had suffered damage in more than one of the structures mentioned above and were excluded from the group analysis. All of the pTBI patients who had no lesion in the PFC or the TPJ bilaterally were selected as a control group ($n = 34$), which is referred to as the "Other TBI" group. A subgroup of 21 participants had frontal damage outside the dlPFC and was excluded from further analysis. A lesion overlay map was created for each of the five groups of participants with brain damage (Figure 1).

2.4.2 | Behavioral data analysis

Behavioral data analysis was carried out on the difference score between the ToM condition and physical condition in the strange stories task. We performed statistical testing using SPSS 26.0 (IBM

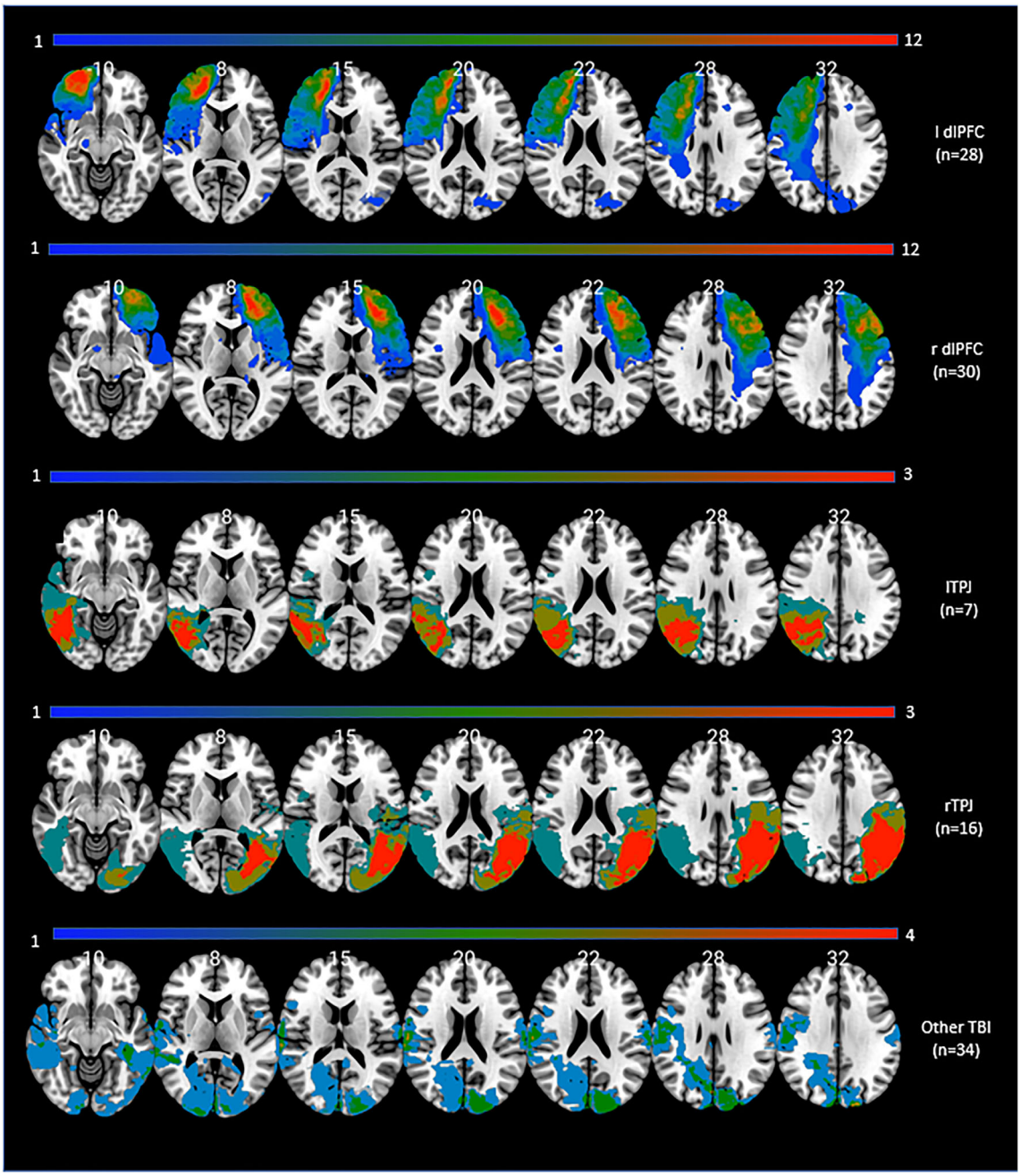

FIGURE 1 Lesion overlay maps of TBI patients (n = 115) grouped by lesion location. Numbers on the top of the brain slices indicate the z coordinates (MNI) of each axial slice. The color indicates the number of veterans in the group with damage to a given voxel. The greatest lesion overlap (red) occurred in the regions of interest. Images are in radiological space (i.e., right is left). For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article. TBI, traumatic brain-injury

Corp., Armonk, NY) and JASP 0.12.2 (JASPTeam, 2020) and significance level was set to $p < .05$ (two-tailed unless otherwise specified). We checked normality of data using the Kolmogorov–Smirnov test and homogeneity of variance using Levene's test. The ToM measure (strange stories difference score) score fulfilled normality and homogeneity of variance assumptions and therefore parametric tests were conducted (analysis of covariance variance [ANCOVA] and regression analysis). The space perception measure did not fulfill the normality assumption (Kolgomorov–Smirnov test: $p < .05$), nor met the homogeneity of variance assumption (Levene's test: $p < .05$), therefore nonparametric tests (Kruskal–Wallis and Mann–Whitney U tests) were used.

The overall difference between the five groups (rTPJ, lTPJ, right dlPFC, left dlPFC, and other TBI) was assessed applying ANCOVA, with participants' pre-injury intelligence score, years of formal education, WAIS working memory index, and WAIS verbal comprehension index as covariates. Then, post hoc tests were conducted ($p < .05$, two-tailed) with Tukey correction for multiple comparisons. Effect

sizes (eta square: η^2 = 0.01 indicates a small effect size, η^2 = 0.06 a medium effect size and η^2 = 0.14 a large effect size, and Cohen's *d*: *d* = 0.2 indicates a small effect size, *d* = 0.5 a medium effect size and *d* = 0.8 a large effect size) were calculated when appropriate.

2.4.3 | Voxel-based lesion-symptom mapping

A VLSM analysis (Bates et al., 2003) was applied in order to test the association between damaged tissue and ToM performance on the strange stories test. In VLSM analysis, the scores of patients with a lesion in a given voxel is compared to the score of patients without a lesion in this voxel using a *t* test. The primary behavioral outcome in the VLSM analysis was the ToM difference score. Additionally, participants' preinjury intelligence score, years of formal education, WAIS working memory index, WAIS verbal comprehension index, and lesion size were used as covariates in order to account for the possible influence of those variables. In order to have sufficient statistical power and to be able to test regions all over the brain, voxels that did not have at least four patients with damage were excluded from the analysis. To correct for multiple comparisons, a false discovery rate correction of 0.05 was used, and at least 100 adjacent voxels must have been statistically significant for a cluster to be reported. The analysis was carried out using the VLSM package version 2.60 (https://aphasialab.org/vlsm/) on MATLAB R2017a (Mathworks, Natick, MA) software. Identification of the brain regions associated with the significant voxels was made using the AAL atlas and Natbrainlab atlas of white matter pathways (Thiebaut de Schotten et al., 2011) in MRICronGL (https://www.nitrc.org/projects/mricrogl) on an MNI standard brain.

2.4.4 | White matter tracts disconnection analysis

To assess the degree to which specific lesions impact brain connectivity, we conducted an analysis of white matter disconnections contributing to ToM deficits. This was done by mapping the normalized lesion from each patient onto tractography reconstructions of white matter pathways obtained from a group of healthy controls (Rojkova et al., 2016) and quantifying the probability that the tract was disconnected by a given lesion (Thiebaut de Schotten et al., 2014) using Tractotron software as part of the BCBtoolkit (Foulon et al., 2018; http://www.toolkit.bcblab.com).

Our goal was to test whether disconnection within specific pathways predicted performance on the ToM tasks. We analyzed a total of 15 tracts: association (three segments of the arcuate, superior longitudinal I, II, III, inferior frontal occipital, inferior longitudinal, uncinate, cingulum), commissural (corpus callosum, anterior commissure), and projection (cortico-spinal, fornix, Optic radiations) tracts. For each individual patient, we considered a given white matter tract to be disconnected if the patient's lesion overlapped a voxel within the white matter pathway map with a probability higher than 50% (above the chance level). We then calculated the percentage of patients with the disconnection within specific white matter tracts within the left and right hemispheres (the analyses were conducted separately for patients with left and right lesions, and excluded patients with bilateral lesions). Within each group of patients, we conducted chi-square tests to compare percentage of disconnection for each tract between patients with versus patients without a deficit in ToM as assessed by the Strange stories task. Patients were classified into groups using zero as a cut-off score, with scores lower than zero reflecting a ToM deficit. This analysis was subjected to Bonferroni correction for multiple comparisons (α-level; p = .003 based on 15 tracts analyzed). For a similar method, see Chechlacz, Rotshtein, and Humphreys (2014).

3 | RESULTS

3.1 | Group analysis

Demographics and neuropsychological testing results of veterans with pTBI and healthy controls (HC) are shown in Table 1. The groups were matched with respect to age (t_{198} = −1.89, p = .17, d = −.027), total years of education (t_{198} = −0.17, p = .85, d = −.03), handedness ($X^2_{2,N=200}$ = .004, p = .99), and preinjury intelligence (t_{198} = −1.65, p = .1, d = −.32). The groups were also matched on their performance in the ToM task (t_{198} = 0.88, p = .38, d = .17), and their verbal comprehension abilities (U = 2,188, p = .21, d = −.28). However, the HC group scored higher than the pTBI group on the WAIS working memory index (t_{198} = −2.86, p = .05, d = −.56).

We next compared five patients' groups based on lesion location: right dlPFC, left dlPFC, rTPJ, lTPJ, and other TBI (see Section 2.4.1 for grouping procedure). Demographics and neuropsychological testing results of the group analysis are shown in Table 2. All groups were matched on age ($F_{4,110}$ = 0.89, p = .46, η^2 = 0.03), total years of education ($F_{4,110}$ = 0.76, p = .55, η^2 = .02), preinjury IQ scores ($F_{4,110}$ = 0.35, p = .83, η^2 = 0.01), and verbal comprehension score ($F_{4,110}$ = 1.3, p = .27, η^2 = 0.04). The groups differed on working memory scores ($F_{4,110}$ = 3.18, p = .001, η^2 = 0.10), with the left TPJ group scoring lower than the all the other groups (all p_{tukey} < .04) except for the left dlPFC group (p_{tukey} = .09).

In addition, the total brain volume loss did not differ among the four target lesion groups ($F_{3,77}$ = 1.21, p = .3, η^2 = 0.04). The "other TBI" group had significantly less volume loss compared to all the target groups (rTPJ: U = 126, p = .002; lTPJ: U = 17, p < .001; r dlPFC: U = 267.5, p = .001; and l dlPFC: U = 232, p = .001).

3.2 | Performance in the ToM task

3.2.1 | Analysis of covariance

We began by performing an analysis of covariance (ANCOVA) to determine whether lesion location affects ToM performance. The primary outcome for the ToM task was a difference score calculated by subtracting the physical stories score from the ToM Story scores, such

TABLE 2 Demographics and neuropsychological measures [mean (SD)] for veterans grouped by lesion location

Variables/group	Right dlPFC *N* = 30	Left dlPFC *N* = 28	Right TPJ *N* = 16	Left TPJ *N* = 7	Other TBI *N* = 34
Demographics:					
Age (years)	57.33 (2.56)	57.57 (1.66)	58.25(2.04)	58.28 (2.75)	58.206 (2.39)
Education (years)	14.16 (2.15)	14.91 (2.67)	14.43 (1.97)	14.42 (2.93)	15.10 (2.62)
Handedness (R:L:A)[a]	24:4:2	24:4:0	14:1:1	7:00:00	26:6:2
Neuropsychological:					
Preinjury IQ[b]	63 (24.51)	60 (22.92)	62.75 (27.69)	69.42 (31.4)	66.52 (25.15)
Theory of Mind[c]	−1.167 (2.80)	0.82 (3.50)	0.5 (3.26)	0.28 (3.63)	1.08 (3.09)
Working memory[d]	98.03 (12.85)	95.32 (13.38)	99.18 (13.67)	80.71 (9.51)	100.38 (15.55)
Verbal comprehension[e]	105.6 (13.08)	105.39 (17.03)	111.62 (12.57)	104.14 (19.54)	111.67 (14.10)
Space perception[f]	9.75 (0.51)	9.82 (0.39)	9.31 (0.94)	10.00 (0)	9.70 (0.57)
Total Brain Volume Loss (cc^3)	33.18 (27.76)	39.30 (40.52)	26.34 (13.57)	50.50 (25.21)	15.31 (14.27)

Abbreviations: dlPFC, dorsolateral prefrontal cortex; TBI, traumatic brain-injury; TPJ, temporoparietal junction; WAIS, Wechsler Adult Intelligence Scale.
[a]Handedness (L:R:A), Left, right, and ambiguous.
[b]Percentile score of Armed Forces Qualification Test (AFQT).
[c]Strange stories task: difference score between ToM and Control condition. Lower score reflects lower ToM performance.
[d]WAIS Working Memory Index score.
[e]WAIS Verbal Comprehension Index score.
[f]Scores on the dot counting test, a subtest from the Visual Object and Space Perception battery.

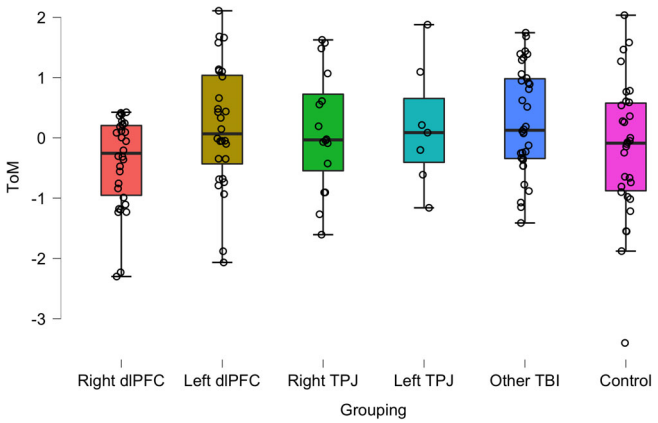

FIGURE 2 Box plots describing the performance on the Strange Stories task across the study groups. The Y axis represents the standardized residuals of the difference score between ToM and physical story conditions after controlling for individual differences in preinjury intelligence, years of education, WAIS working memory index score, and the WAIS verbal comprehension index score. The horizontal line within each box indicates the value of the group median. Figures constructed using JASP Version 0.12.2 (JASPTeam, 2020). WAIS, Wechsler Adult Intelligence Scale

that scores below zero represents poorer performance on ToM. In order to control for individual differences in preinjury intelligence, years of formal education, working memory, and verbal comprehension, these four measures were entered as covariates. ANCOVA comparing ToM performance across the five target groups and healthy controls revealed an overall significant difference in ToM performance ($F_{5,135}$ = 2.21, p = .05, η^2 = 0.07; Figure 2). Moreover this effect remained significant after repeating the analysis without the healthy control group and adding total brain volume loss as a fifth covariate ($F_{4,105}$ = 2.46, p = .05, η^2 = 0.07). A post hoc test using a Tukey correction for multiple comparisons revealed that the right dlPFC group had significantly lower scores on the ToM Story Score than the other TBI group (t = −2.86, p_{tukey} = .05). No other group differences were found.

3.2.2 | Regression analyses

Next, we conducted linear regression analyses to provide further evidence that damage to right dlPFC was specifically associated with ToM performance. The regression model included the ToM differences score as the dependent variable, and the following as covariates: pre-injury intelligence score, years of formal education, WAIS working memory index score, WAIS verbal comprehension index score, percent damage to right dlPFC, percent damage to left dlPFC, percent damage to right TPJ, and percent damage to left TPJ. Overall, the model explained a significant proportion of variance in ToM performance (R^2 = 0.13, $F_{8,106}$ = 2.02, p = .05), with more damage to the right dlPFC predicting lower ToM performance (β = −.10; t = −3.50, p < .001, one-tailed). In contrast, no other lesion volume contributed significantly to the model. Adding total volume loss as a covariate to the model resulted in a larger portion of explained variance in ToM performance (R^2 = 0.17, $F_{9,105}$ = 2.38, p = .01), with both right dlPFC β = −.05; t = −1.60, p = .04, one-tailed) and total volume loss (β = −.03; t = −2.10, p = .01, one-tailed) predicting a significant deficit in ToM. No other covariate contributed significantly to the model.

3.3 | Space perception

To ensure that our key results were specific to ToM and not some unanticipated factor, we compared the groups on a measure typically associated with parietal cortex function, namely a space perception measure. The space perception measure did not fulfill the normality or homogeneity of variance assumption and therefore nonparametric tests were used. In order to account for preinjury intelligence, education, and total brain volume loss, the linear regression model was fitted with the score of the dot counting test as the dependent variable, and the three variables mentioned above as predictors. The standardized residuals from this model indexed space perception after controlling for individual differences in preinjury intelligence, education, and total brain volume loss. This standardized score was used as the dependent measure for this analysis. First, the Kruskal–Wallis test was used to compare scores across all five groups, revealing a significant overall difference ($H[df = 4, n = 114] = 17.34$, $p = .002$). Follow-up Mann–Whitney tests showed that the rTPJ group scored lower than the right dlPFC group ($U = 139.5$, $p = .027$), the left dlPFC group ($U = 131$, $p = .023$), and the left TPJ group ($U = 16$, $p = .008$). There was a nonsignificant difference in performance between the right TPJ group and the other TBI group ($U = 250$, $p = .64$) on this measure.

3.4 | VLSM analysis

The overlay map of lesion locations for all 170 patients is presented in Figure 3. Note that the map shows brain regions with lesions present for at least four participants in each voxel consistent with the constraints of the VLSM (described above), where analyses were confined to voxels where a minimum of four patients have a lesion. The map shows a sufficient degree of overlap in order to draw conclusions for all the target brain regions, namely dlPFC and TPJ, bilaterally.

A whole-brain VLSM analysis was performed with the ToM difference score as the outcome, and the following five measures as covariates: preinjury intelligence score, years of formal education, WAIS working memory index score, WAIS verbal comprehension index score, and total brain volume loss. The VLSM analysis revealed four significant clusters with over 100 voxels each, which are listed in Table 3 (see also Figure 4a). There were two main clusters with over 1,000 voxels each: the first (volume = 2,514 voxels, Max $t = 3.15$) was located predominantly in the frontal middle gyrus in the right hemisphere. The peak MNI coordinates were (40 46 28), and the center coordinates were (24 51 7, see Figure 4a). The second (volume = 2,302 voxels, Max $t = 3.10$), cluster was located primarily within the left inferior parietal gyrus. The peak MNI coordinates were (−24 −54 −60), and the center coordinates were (−30 −48 42; see Figure 4c).

3.5 | White matter tracts disconnection

Patients' lesions were compared to an atlas of white matter connections in order to identify the probability of tract disconnections (Foulon et al., 2018; Rojkova et al., 2016). The percentage of patients with disconnected tracts was calculated separately for patients with

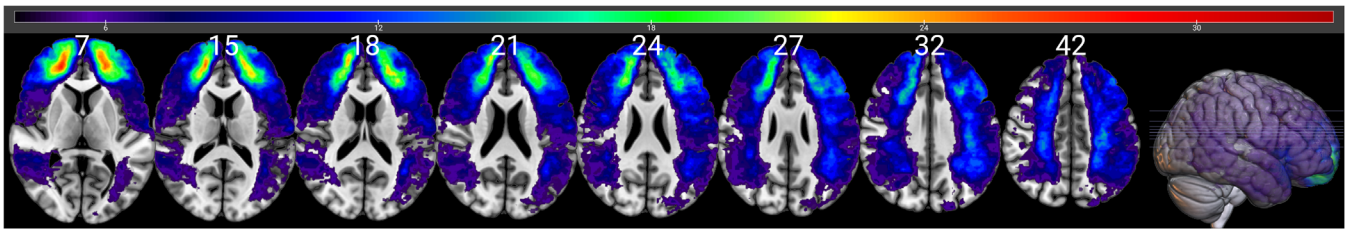

FIGURE 3 Lesion overlay demonstrating overlap in lesions across 170 participants included in the VLSM analysis, with a minimum of four participants' lesions in each voxel and a maximum of 33. Values in white indicate the z coordinates (MNI) of each axial slice. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article. VLSM, voxel-based lesion-symptom mapping

TABLE 3 Results from voxel-based lesion-symptom analyses showing regions of damage associated with lower Theory of Mind and Physical Story Scores

Structure	Voxels	Peak MNI coordinates			Max *t*-value
		x	*y*	*z*	
Frontal middle gyrus (right)	2,514	24	51	7	3.15
Inferior and superior parietal gyrus (left)	2,302	−30	−48	42	3.10
Frontal middle gyrus (right)	222	45	2	56	2.83
Superior frontal gyrus, medial Corpus callosum, anterior cingulate (left)	118	−9	48	34	2.72

Note: Regions defined using Automated Anatomical Labeling (AAL) atlas and the Natbrainlab atlas of white matter pathways. We report clusters with 100 or more voxels.

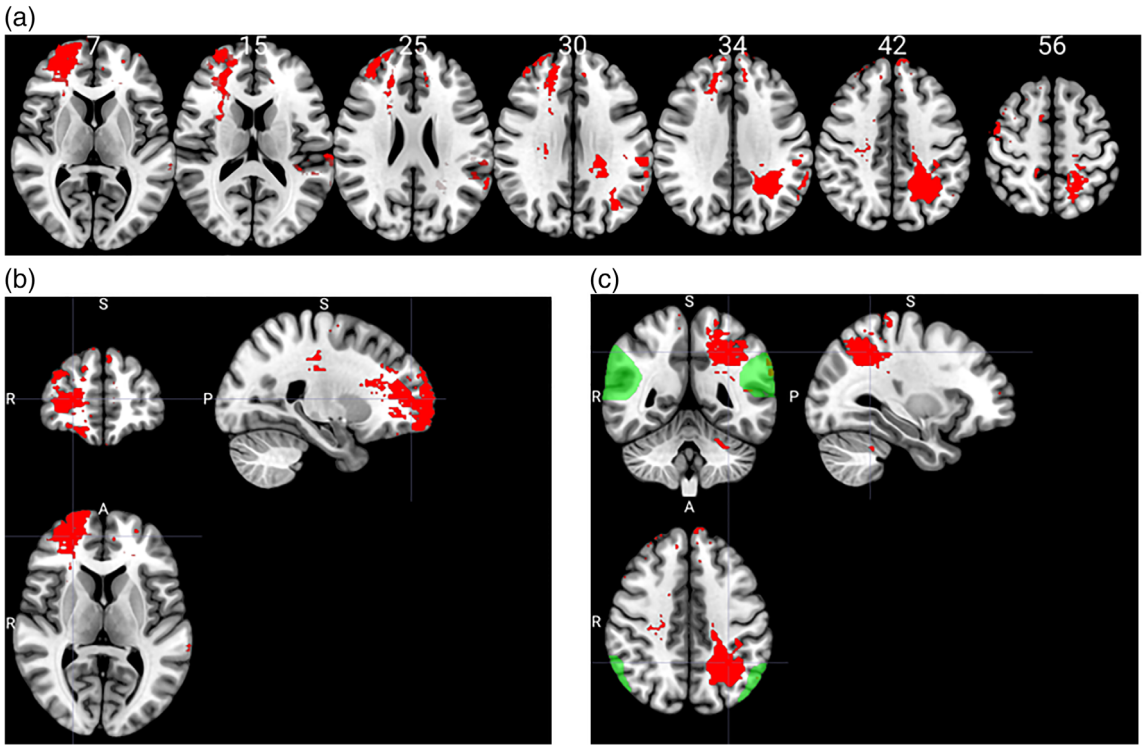

FIGURE 4 (a) Voxel-based lesion-symptom mapping analysis results. In red are areas of damage that were associated with a deficit in ToM. Values in white indicate the z coordinates (MNI) of each axial slice. (b) Center MNI coordinates for the main cluster in the right hemisphere. (c) Center MNI coordinates for main cluster in the left hemisphere. In green are the right and left temporoparietal junction (maps downloaded from https://saxelab.mit.edu/use-our-theory-mind-groupmaps). Images are in radiological space (i.e., right is left)

and without deficits in ToM (determined based on zero as a cut-off score, difference score equal or higher than 0 reflects no deficit, score < 0 reflects deficit), and separately for patients with left (n = 62) or right hemisphere lesions (n = 80). We compared the groups of patients with and without presumed damage for each white matter tract separately, using a chi-square test. This analysis revealed that disconnections of the left Inferior longitudinal fasciculus (X^2 [1, N = 62] = 4.08, p = .04), and the right superior longitudinal fasciculus 1 (X^2 [1, N = 80] = 4.21, p = .04) were associated with a poorer performance in the ToM task (Figure 5). However, these comparisons did not survive Bonferroni correction for multiple comparisons.

4 | DISCUSSION

In this study, we used different methods to examine the causal association between focal brain lesions and mental state attribution. We found that lesions in the right frontal middle gyrus (overlapping with the right dlPFC) and in the inferior and left superior parietal gyrus can lead to deficits in mental state attribution. We also found that brain lesions that disconnect the left Inferior longitudinal fasciculus and right superior longitudinal fasciculus 1 are more likely to result in ToM deficits.

Patients with damage to the *right dlPFC* were impaired on the ToM task compared to the other lesion groups, suggesting that this region is necessary for mental state attribution. This is supported by the results of the VLSM analysis that revealed a cluster in the right dlPFC which was linked to a lower performance in the ToM task. Second, patients with damage to the *left dlPFC* performed no different than any other patient group. Yet, the VLSM analysis revealed a small lesion cluster in the left dorsomedial prefrontal cortex (dmPFC) which was linked with poorer performance on the ToM task. This finding supports other previous studies claiming involvement of dmPFC in cognitive theory of mind (Corradi-Dell'Acqua et al., 2020). Third, patients with damage to the *left TPJ* did not differ from the other groups on ToM performance. However, the VLSM analysis also revealed a cluster in the left TPJ which was linked to lower ToM performance. Given that the group of left TPJ patients was the smallest patient group in this study (n = 7), these findings suggest a role for the left TPJ in mental state attribution.

Last, patients with damage to the *right TPJ* performed similarly to the other patient groups. Moreover, for each patient with rTPJ damage and impaired performance on the ToM task, there is an example of another patient with a similar lesion in the rTPJ and intact ToM performance. In addition, VLSM analysis did not reveal any cluster overlapping with the rTPJ. In sum, these results suggest that an intact rTPJ is not necessary for mental state attribution. Given that these findings are not in line with previous imaging studies, it might be argued that this specific group of rTPJ patients had exceptionally well-preserved cognition in general. This was not the case, however,

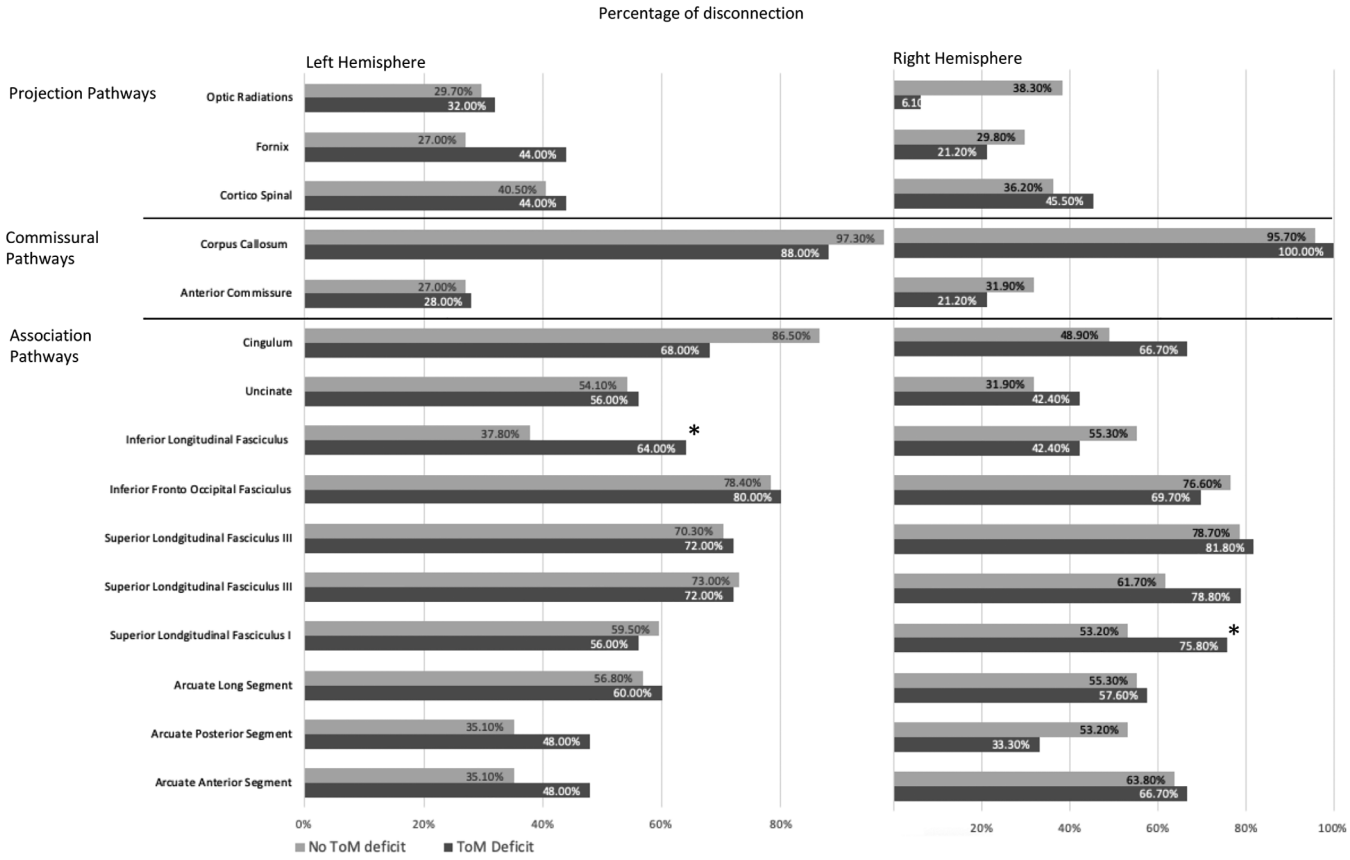

FIGURE 5 Percentage of patients with disconnection in projection, commissural and association white matter pathways within the left and right hemisphere, calculated for groups with and without a deficit in ToM performance. * denotes a significant group difference, $p < .05$

since the same group scored significantly lower on a space perception task.

4.1 | The role of the TPJ in mental state attribution

Given the amount of data linking brain activation in the rTPJ and ToM performance (Saxe & Kanwisher, 2003; Saxe & Powell, 2006; Saxe & Wexler, 2005), it is likely that this region contributes to making mental state attributions, yet in our study we did not find supporting evidence to suggest that this area is *necessary* for this ability. Historically, lesions to rTPJ have been often associated most with attention deficits in general, and unilateral spatial neglect in particular (Vallar & Calzolari, 2018). The important role of the rTPJ in attentional processes—specifically in attentional reorienting—has been established in numerous functional imaging and lesion studies (Corbetta, Patel, & Shulman, 2008; Dugué, Merriam, Heeger, & Carrasco, 2018; Geng & Vossel, 2013; Jakobs et al., 2012; Mitchell, 2008; Rinne et al., 2013). It was also shown to play a causal role in spatial representation and visuospatial memory (Chechlacz et al., 2014). Given that different studies linked the rTPJ to both attentional reorientation and ToM tasks, some researchers have suggested that the TPJ has a role in shifting orientation from self to other, which can be utilized during ToM tasks (Arzy, Thut, Mohr, Michel, & Blanke, 2006; Krall et al., 2016; Martin, Huang, Hunold, & Meinzer, 2019). Since we did not directly test attentional reorientation in the VHIS, we cannot rule out or support this claim directly.

The current VLSM analysis revealed a large cluster in the left Inferior and Superior parietal gyrus, which partially overlapped with the left TPJ, and was linked to lower performance on the ToM task. This finding is consistent with previous findings indicating a causal role for the left TPJ in mental state attribution (Biervoye et al., 2016; Samson et al., 2004). However, the group of patients with damage to the left TPJ did not differ from the other groups on their average ToM score. Interestingly, this group also scored lower on the Wechsler WM scale, which implies that their brain damage resulted in an impaired ability to maintain information (including that required for ToM) for further processing rather than a specific impairment in mental state attribution.

4.2 | The role of the dlPFC in mental state attribution

Our finding regarding the crucial role of the dlPFC in mental state attribution supports and extends previous (Corradi-Dell'Acqua et al., 2020; Kalbe et al., 2010; Rowe, Bullock, Polkey, & Morris, 2001; Shamay-Tsoory & Aharon-Peretz, 2007; Stuss et al., 2001; Xi et al., 2011). Yet, they may seem in contrast to findings reported by Herbet, Lafargue, Bonnetblanc, Moritz-Gasser, and Duffau (2013) who studied 10 patients

with a diffuse low-grade glioma (DLGG) in right frontal areas. Patients in that study were assessed on a ToM task just before, immediately after, and 3 months after brain surgery. Unlike the current study, Herbert et al. did not find long-term effects of right frontal lesions on ToM performance. However, a closer look reveals that the two studies are difficult to compare due to several reasons: first, Herbet and his colleagues focused on patients with dorsomedial frontal lesions and not dorsolateral frontal lesions as we did. Second, patients with dmPFC tumors often have lower scores preoperatively, therefore stable scores postoperatively may still reflect overall impairment. Third, one patient in that study did show significant long-term deficits similar to those we are reporting in this study but it may be that the extent of the resection was larger for this patient perhaps affecting more lateral prefrontal cortex. Last, given that DLGG is characterized by slow-growing progression which can lead to dramatic brain plasticity (Duffau, 2005), studies focusing on patients with DLGG may teach us as much about the brain's ability to compensate for neuronal loss as about brain-behavior associations in healthy individuals.

The specific role of the right dlPFC in ToM processes is still an unknown. Given the inhibitory role this area plays in other cognitive domains (Floden & Stuss, 2006; Oldrati, Patricelli, Colombo, & Antonietti, 2016), it has been argued that the right dlPFC is crucial for inhibiting one's own point of view and considering the others'. This type of inhibition is required in order to be successful on the strange stories test; however, we did not test this hypothesis in the current study and therefore cannot directly support this claim.

4.3 | The role of the white matter tracts in mental state attribution

The results of a white matter disconnection analysis indicated that poor performance measured in patients with brain damage can be associated in part with white matter tract disconnections, specifically in the right superior longitudinal fasciculus and the left inferior longitudinal fasciculus (ILF). The superior longitudinal fasciculus (SLF) is a frontoparietal white matter tract which is thought to play a role in attention and visuospatial processing, while the ILF is a white matter pathway that primarily connects the anterior temporal lobe with the occipital lobe (Herbet, Zemmoura, & Duffau, 2018). The disconnection in the right superior longitudinal fasciculus is consistent with results from the VLSM analysis indicating clusters in the right frontal middle gyrus. Moreover, the definition of the ILF in the tractotron encompasses temporoparietal fibers more commonly assigned to the middle longitudinal fasciculus, therefore this finding is compatible with the VLSM cluster in the left inferior parietal gyrus.

Taken together, these findings support the view that white-matter connectivity is essential for mental state attribution. Our data fit well with previous studies suggesting an important role for the arcuate fasciculus (AF)/superior longitudinal fasciculus in theory of mind. For example, two studies on patients with glioma used a lesion mapping approach to reveal that disconnections in the SLF/AF resulted in poorer ability to identify one's affective state by looking at one's eyes (Herbet et al., 2014; Nakajima, Yordanova, Duffau, & Herbet, 2018). A different study on healthy children (ages 3–4) used diffusion-weighted MRI and showed that the arcuate fascicle and the inferior fronto-occipital fascicle (IFOF) both play an important role in the development of ToM abilities (Grosse Wiesmann, Schreiber, Singer, Steinbeis, & Friederici, 2017).

Moreover, studies focusing on individuals with autism spectrum disorder (ASD), a population with profound difficulty in mental state attribution, suggest that both the SLF and ILF are affected in the population. Lower white matter integrity in the right ILF was shown to characterize children with ASD, compared to typically developing children (Koldewyn et al., 2014). Moreover, recent studies showed that left ILF may be involved in the presentation of autistic traits among non-diagnosed individuals (Bradstreet, Hecht, King, Turner, & Robins, 2017), and in language development in individuals with ASD (Naigles et al., 2017). Several studies also found alterations in the SLF among individuals with ASD (Fitzgerald et al., 2018; Poustka et al., 2012; Weinstein et al., 2011). Moreover, a recent study showed that 6-week-old infants at risk for autism showed higher fractional anisotropy in the right SLF compared to low risk infants (Liu et al., 2019).

4.4 | Current and previous ToM related findings from the VHIS

Findings from this study can be viewed together with two previously published papers which analyzed VHIS data focusing on ToM. One study (Dal Monte et al., 2014) focused on mental state recognition from facial expressions using the Reading the Mind in the Eyes Task (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). While this task is often presented as a ToM task, recent findings support the claim that it actually measures emotion recognition and not ToM abilities (Oakley, Brewer, Bird, & Catmur, 2016). Another study (Leopold et al., 2012) examined performance on an affective ToM task in which participants are required to judge whether something that was said in a specific social context was appropriate or not (the Faux Pas Recognition task; Stone, Baron-Cohen, & Knight, 1998). This study found that damage to the left ventromedial prefrontal cortex (vmPFC) was associated with performance on the Faux Pas task but not on the strange stories task.

Taken together with the findings of the current analysis, these two studies support the distinction between neural networks which are involved in processing affective and cognitive mental states: the vmPFC being involved specifically in affective ToM tasks, while a larger frontal network including the dlPFC is involved in cognitive ToM tasks (Abu-Akel & Shamay-Tsoory, 2011; Corradi-Dell'Acqua et al., 2020; Kalbe et al., 2010; Shamay-Tsoory & Aharon-Peretz, 2007).

4.5 | Limitations

Despite the unique opportunity to sample a large, relatively homogenous set of patients with focal brain lesions, we acknowledge specific

limitations to the present study. First, our sample was composed entirely of older adult male combat veterans. Therefore, it may be difficult to infer how our results might generalize to a more diverse population, particularly given the possibility that ToM is associated with different brain structures in men versus women (Adenzato et al., 2017). Furthermore, given that our study was conducted 30 years postinjury, it could be argued that plasticity would make it difficult to attribute the roles of different brain regions in ToM. Nevertheless, we still observed consistent ToM impairments with right dlPFC damage over 30 years postinjury providing further support for the argument that this brain area is necessary for this function.

5 | CONCLUSIONS

To conclude, the current study provides direct causal neuropsychological evidence for the role of the right dlPFC in mental attribution. It also demonstrates that intact ToM performance is evident despite damage to the rTPJ, challenging conventional wisdom about its *essential* role in Theory of Mind processing (Samson et al., 2004; Saxe & Kanwisher, 2003; Saxe & Powell, 2006; Saxe & Wexler, 2005). Our results are consistent with the idea that Theory of Mind requires a network of modular processes which rely on several cognitive domains (Leslie, Friedman, & German, 2004; Mitchell, 2008) with the dlPFC gaining processing priority depending on the stimuli used and the task design.

ACKNOWLEDGMENTS

This research was supported by the Therapeutic Cognitive Neuroscience Fund (Barry Gordon) and the Smart Family Foundation of New York (Jordan Grafman). The funders played no role in the design of this study or the interpretation of its results. We would like to thank Arya Shariat for calculating the TPJ volume loss for each of the study participants. We would also like to thank all the Vietnam veterans who participated in this study. Without their long-term commitment to improving the health care of veterans, this study could not have been completed.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request. This study was not preregistered.

ORCID

Shira Cohen-Zimerman https://orcid.org/0000-0001-6098-2550

REFERENCES

Abu-Akel, A., & Shamay-Tsoory, S. (2011). Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia*, *49*, 2971–2984. https://doi.org/10.1016/j.neuropsychologia.2011.07.012

Adenzato, M., Brambilla, M., Manenti, R., De Lucia, L., Trojano, L., Garofalo, S., … Cotelli, M. (2017). Gender differences in cognitive theory of mind revealed by transcranial direct current stimulation on medial prefrontal cortex. *Scientific Reports*, *7*, 41219. https://doi.org/10.1038/srep41219

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews. Neuroscience*, *7*, 268–277. https://doi.org/10.1038/nrn1884

Apperly, I. A. (2012). What is "theory of mind"? Concepts, cognitive processes and individual differences. *The Quarterly Journal of Experimental Psychology*, *65*, 825–839. https://doi.org/10.1080/17470218.2012.676055

Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: Neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience*, *16*, 1773–1784.

Arzy, S., Thut, G., Mohr, C., Michel, C. M., & Blanke, O. (2006). Neural basis of embodiment: Distinct contributions of temporoparietal junction and extrastriate body area. *The Journal of Neuroscience*, *26*, 8074–8081.

Baldo, J. V., Kacinik, N. A., Moncrief, A., Beghin, F., & Dronkers, N. F. (2016). You may now kiss the bride: Interpretation of social situations by individuals with right or left hemisphere injury. *Neuropsychologia*, *80*, 133–141.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the mind in the eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*, 241–251.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion–symptom mapping. *Nature Neuroscience*, *6*, 448–450. https://doi.org/10.1038/nn1050

Biervoye, A., Dricot, L., Ivanoiu, A., & Samson, D. (2016). Impaired spontaneous belief inference following acquired damage to the left posterior temporoparietal junction. *Social Cognitive and Affective Neuroscience*, *11*, 1513–1520.

Bradstreet, L. E., Hecht, E. E., King, T. Z., Turner, J. L., & Robins, D. L. (2017). Associations between autistic traits and fractional anisotropy values in white matter tracts in a nonclinical sample of young adults. *Experimental Brain Research*, *235*, 259–267.

Brosch, T., Schiller, D., Mojdehbakhsh, R., Uleman, J. S., & Phelps, E. A. (2013). Neural mechanisms underlying the integration of situational information into attribution outcomes. *Social Cognitive and Affective Neuroscience*, *8*, 640–646. https://doi.org/10.1093/scan/nst019

Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, *30*, 2313–2335.

Chechlacz, M., Rotshtein, P., & Humphreys, G. W. (2014). Neuronal substrates of Corsi Block span: Lesion symptom mapping analyses in relation to attentional competition and spatial bias. *Neuropsychologia*, *64*, 240–251.

Cohen-Zimerman, S., Salvi, C., Krueger, F., Gordon, B., & Grafman, J. (2018). Intelligence across the seventh decade in patients with brain injuries acquired in young adulthood. *Trends in Neuroscience and Education*, *13*, 1–7. https://doi.org/10.1016/j.tine.2018.08.001

Collins, D. L., Neelin, P., Peters, T. M., & Evans, A. C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography*, *18*, 192–205.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron*, *58*, 306–324.

Corradi-Dell'Acqua, C., Ronchi, R., Thomasson, M., Bernati, T., Saj, A., & Vuilleumier, P. (2020). Deficits in cognitive and affective theory of mind relate to dissociated lesion patterns in prefrontal and insular

cortex. *Cortex*, *128*, 218–233. https://doi.org/10.1016/j.cortex.2020.03.019

Dal Monte, O., Schintu, S., Pardini, M., Berti, A., Wassermann, E. M., Grafman, J., & Krueger, F. (2014). The left inferior frontal gyrus is crucial for reading the mind in the eyes: Brain lesion evidence. *Cortex*, *58*, 9–17.

Duffau, H. (2005). Lessons from brain mapping in surgery for low-grade glioma: Insights into associations between tumour and brain plasticity. *Lancet Neurology*, *4*, 476–486.

Dufour, N., Redcay, E., Young, L., Mavros, P. L., Moran, J. M., Triantafyllou, C., … Saxe, R. (2013). Similar brain activation during false belief tasks in a large sample of adults with and without autism. *PLoS One*, *8*, e75468. https://doi.org/10.1371/journal.pone.0075468

Dugué, L., Merriam, E. P., Heeger, D. J., & Carrasco, M. (2018). Specific visual subregions of TPJ mediate reorienting of spatial attention. *Cerebral Cortex*, *28*, 2375–2390.

Fitzgerald, J., Leemans, A., Kehoe, E., O'Hanlon, E., Gallagher, L., & McGrath, J. (2018). Abnormal fronto-parietal white matter organisation in the superior longitudinal fasciculus branches in autism spectrum disorders. *The European Journal of Neuroscience*, *47*, 652–661.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. J., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition*, *57*, 109–128.

Floden, D., & Stuss, D. T. (2006). Inhibitory control is slowed in patients with right superior medial frontal damage. *Journal of Cognitive Neuroscience*, *18*, 1843–1849.

Forbes, C. E., & Grafman, J. (2010). The role of the human prefrontal cortex in social cognition and moral judgment. *Annual Review of Neuroscience*, *33*, 299–324.

Foulon, C., Cerliani, L., Kinkingnehun, S., Levy, R., Rosso, C., Urbanski, M., … Thiebaut de Schotten, M. (2018). Advanced lesion symptom mapping analyses and implementation as BCBtoolkit. *Gigascience*, *7*, giy004.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*, 531–534.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences*, *7*, 77–83.

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia*, *38*, 11–21.

Geng, J. J., & Vossel, S. (2013). Neuroscience and biobehavioral reviews re-evaluating the role of TPJ in attentional control: Contextual updating? *Neuroscience and Biobehavioral Reviews*, *37*, 2608–2620. https://doi.org/10.1016/j.neubiorev.2013.08.010

Goel, V., Grafman, J., Sadato, N., & Hallett, M. (1995). Modeling other minds. *Neuroreport*, *6*, 1741–1746.

Gokcen, S., Bora, E., Erermis, S., Kesikci, H., & Aydin, C. (2009). Theory of mind and verbal working memory deficits in parents of autistic children. *Psychiatry Research*, *166*, 46–53.

Gozzi, M., Raymont, V., Solomon, J., Koenigs, M., & Grafman, J. (2009). Dissociable effects of prefrontal and anterior temporal cortical lesions on stereotypical gender attitudes. *Neuropsychologia*, *47*, 2125–2132.

Grafman, J., Jonas, B. S., Martin, A., Salazar, A. M., Weingartner, H., Ludlow, C., … Vance, S. C. (1988). Intellectual function following penetrating head-injury in Vietnam veterans. *Brain*, *111*, 169–184.

Grosse Wiesmann, C., Schreiber, J., Singer, T., Steinbeis, N., & Friederici, A. D. (2017). White matter maturation is associated with the emergence of theory of mind in early childhood. *Nature Communications*, *8*, 14692. https://doi.org/10.1038/ncomms14692

Happé, F., Brownell, H., & Winner, E. (1999). Acquired "theory of mind" impairments following stroke. *Cognition*, *70*, 211–240.

Happé, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., … Frith, C. (1996). 'Theory of mind' in the brain. Evidence from a PET Scan Study of Asperger syndrome. *Neuroreport*, *8*, 197–201.

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*, 129–154. https://doi.org/10.1007/BF02172093

Herbet, G., Lafargue, G., Bonnetblanc, F., Moritz-Gasser, S., & Duffau, H. (2013). Is the right frontal cortex really crucial in the mentalizing network? A longitudinal study in patients with a slow-growing lesion. *Cortex*, *49*, 2711–2727.

Herbet, G., Lafargue, G., Bonnetblanc, F., Moritz-Gasser, S., Menjot De Champfleur, N., & Duffau, H. (2014). Inferring a dual-stream model of mentalizing from associative white matter fibres disconnection. *Brain*, *137*, 944–959.

Herbet, G., Zemmoura, I., & Duffau, H. (2018). Functional anatomy of the inferior longitudinal fasciculus: From historical reports to current hypotheses. *Frontiers in Neuroanatomy*, *12*. https://www.frontiersin.org/article/10.3389/fnana.2018.00077

Jakobs, O., Langner, R., Caspers, S., Roski, C., Cieslik, E. C., Zilles, K., … Eickhoff, S. B. (2012). Across-study and within-subject functional connectivity of a right temporo-parietal junction subregion involved in stimulus–context integration. *NeuroImage*, *60*, 2389–2398.

JASPTeam. (2020). *JASP (version 0.12.2)*.

Kalbe, E., Schlegel, M., Sack, A. T., Nowak, D. A., Dafotakis, M., Bangard, C., … Kessler, J. (2010). Dissociating cognitive from affective theory of mind: A TMS study. *Cortex*, *46*, 769–780. https://doi.org/10.1016/j.cortex.2009.07.010

Kobayashi, C., Glover, G. H., & Temple, E. (2007). Children's and adults' neural bases of verbal and nonverbal 'theory of mind'. *Neuropsychologia*, *45*, 1522–1532.

Koldewyn, K., Yendiki, A., Weigelt, S., Gweon, H., Julian, J., Richardson, H., … Kanwisher, N. (2014). Differences in the right inferior longitudinal fasciculus but no general disruption of white matter tracts in children with autism spectrum disorder. *Proceedings of the National Academy of Sciences*, *111*, 1981–1986.

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. *Brain Structure & Function*, *220*, 587–604.

Krall, S. C., Volz, L. J., Oberwelland, E., Grefkes, C., Fink, G. R., & Konrad, K. (2016). The right temporoparietal junction in attention and social interaction: A transcranial magnetic stimulation study. *Human Brain Mapping*, *37*, 796–807.

Krueger, F., Barbey, A. K., & Grafman, J. (2009). The medial prefrontal cortex mediates social event knowledge. *Trends in Cognitive Sciences*, *13*, 103–109.

Kubit, B., & Jack, A. I. (2013). Rethinking the role of the rTPJ in attention and social cognition in light of the opposing domains hypothesis: Findings from an ALE-based meta-analysis and resting-state functional connectivity. *Frontiers in Human Neuroscience*, *7*, 323.

Leopold, A., Krueger, F., Dal Monte, O., Pardini, M., Pulaski, S. J., Solomon, J., & Grafman, J. (2012). Damage to the left ventromedial prefrontal cortex impacts affective theory of mind. *Social Cognitive and Affective Neuroscience*, *7*, 871–880.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences*, *8*, 528–533.

Liu, J., Tsang, T., Jackson, L., Ponting, C., Jeste, S. S., Bookheimer, S. Y., & Dapretto, M. (2019). Altered lateralization of dorsal language tracts in 6-week-old infants at risk for autism. *Developmental Science*, *22*, e12768.

Mah, Y.-H., Husain, M., Rees, G., & Nachev, P. (2014). Human brain lesion-deficit inference remapped. *Brain*, *137*, 2522–2531.

Mai, X., Zhang, W., Hu, X., Zhen, Z., Xu, Z., & Zhang, J. (2016). Using tDCS to explore the role of the right temporo-parietal junction in theory of mind and cognitive empathy. *Frontiers in Psychology*, *7*, 380.

Martin, A. K., Huang, J., Hunold, A., & Meinzer, M. (2019). Dissociable roles within the social brain for self–other processing: A HD-tDCS study.

Cerebral Cortex, *29*(8), 3642–3654. https://doi.org/10.1093/cercor/bhy238

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex*, *18*, 262–271. https://doi.org/10.1093/cercor/bhm051

Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience and Biobehavioral Reviews*, *65*, 276–291.

Naigles, L. R., Johnson, R., Mastergeorge, A., Ozonoff, S., Rogers, S. J., Amaral, D. G., & Nordahl, C. W. (2017). Neural correlates of language variability in preschool-aged boys with autism spectrum disorder. *Autism Research*, *10*, 1107–1119.

Nakajima, R., Yordanova, Y. N., Duffau, H., & Herbet, G. (2018). Neuropsychological evidence for the crucial role of the right arcuate fasciculus in the face-based mentalizing network: A disconnection analysis. *Neuropsychologia*, *115*, 179–187.

Oakley, B. F. M., Brewer, R., Bird, G., & Catmur, C. (2016). Theory of mind is not theory of emotion: A cautionary note on the reading the mind in the eyes test. *Journal of Abnormal Psychology*, *125*(6), 818–823. https://doi.org/10.1037/abn0000182

Oldrati, V., Patricelli, J., Colombo, B., & Antonietti, A. (2016). The role of dorsolateral prefrontal cortex in inhibition mechanism: A study on cognitive reflection test and similar tasks through neuromodulation. *Neuropsychologia*, *91*, 499–508.

Otti, A., Wohlschlaeger, A. M., & Noll-Hussong, M. (2015). Is the medial prefrontal cortex necessary for theory of mind? *PLoS One*, *10*, e0135912. https://doi.org/10.1371/journal.pone.0135912

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neuroscience*, *1*, 245–258. https://doi.org/10.1080/17470910600989896

Poustka, L., Jennen-Steinmetz, C., Henze, R., Vomstein, K., Haffner, J., & Sieltjes, B. (2012). Fronto-temporal disconnectivity and symptom severity in children with autism spectrum disorder. *The World Journal of Biological Psychiatry*, *13*, 269–280.

Raymont, V., Salazar, A. M., Krueger, F., & Grafman, J. (2011). "Studying injured minds" - The Vietnam head injury study and 40 years of brain injury research. *Frontiers Neurology*, *2*, 15.

Rinne, P., Hassan, M., Goniotakis, D., Chohan, K., Sharma, P., Langdon, D., … Bentley, P. (2013). Triple dissociation of attention networks in stroke according to lesion location. *Neurology*, *81*, 812–820.

Rojkova, K., Volle, E., Urbanski, M., Humbert, F., Dell'Acqua, F., & De Schotten, M. T. (2016). Atlasing the frontal lobe connections and their variability due to age and education: A spherical deconvolution tractography study. *Brain Structure & Function*, *221*, 1751–1766.

Rowe, A. D., Bullock, P. R., Polkey, C. E., & Morris, R. G. (2001). 'Theory of mind' impairments and their relationship to executive functioning following frontal lobe excisions. *Brain*, *124*, 600–616. https://doi.org/10.1093/brain/124.3.600

Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, *7*, 499–500.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*, 235–239.

Saxe, R. (2010). The right temporo-parietal junction: A specific brain region for thinking about thoughts. In A. Leslie & T. German (Eds.) *Handbook of theory of mind* (pp. 1–35). Brighton, England: Taylor & Francis.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *NeuroImage*, *19*, 1835–1842.

Saxe, R., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science*, *17*, 692–699. https://doi.org/10.1111/j.1467-9280.2006.01768.x

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia*, *43*, 1391–1399.

Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E. N., & Saxe, R. (2009). Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS One*, *4*, e4869. https://doi.org/10.1371/journal.pone.0004869

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, *42*, 9–34.

Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia*, *45*, 3054–3067.

Shamay-Tsoory, S. G., Tibi-Elhanany, Y., & Aharon-Peretz, J. (2006). The ventromedial prefrontal cortex is involved in understanding affective but not cognitive theory of mind stories. *Social Neuroscience*, *1*, 149–166. https://doi.org/10.1080/17470910600985589

Solomon, J., Raymont, V., Braun, A., Butman, J. A., & Grafman, J. (2007). User-friendly software for the analysis of brain lesions (ABLe). *Computer Methods and Programs in Biomedicine*, *86*, 245–254.

Sommer, M., Döhnel, K., Sodian, B., Meinhardt, J., Thoermer, C., & Hajak, G. (2007). Neural correlates of true and false belief reasoning. *NeuroImage*, *35*, 1378–1384.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, *10*, 640–656.

Stuss, D. T., Gallup, G. G., Jr., & Alexander, M. P. (2001). The frontal lobes are necessary for 'theory of mind'. *Brain*, *124*, 279–286. https://doi.org/10.1093/brain/124.2.279

Thiebaut de Schotten, M., Bizzi, A., Dell'Acqua, F., Allin, M., Walshe, M., Murray, R., … Catani, M. (2011). Atlasing location, asymmetry and inter-subject variability of white matter tracts in the human brain with MR diffusion tractography. *NeuroImage*, *54*, 49–59.

Thiebaut de Schotten, M., Tomaiuolo, F., Aiello, M., Merola, S., Silvetti, M., Lecce, F., … Doricchi, F. (2014). Damage to white matter pathways in subacute and chronic spatial neglect: A group study and 2 single-case studies with complete virtual "in vivo" tractography dissection. *Cerebral Cortex*, *24*, 691–706.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*, 273–289.

Vallar, G., & Calzolari, E. (2018). Chapter 14 - Unilateral spatial neglect after posterior parietal damage. In G. Vallar & H. B. Coslett (Eds.), *The parietal lobe* (Vol. 151, pp. 287–312). Amsterdam, The Netherlands: Elsevier. http://www.sciencedirect.com/science/article/pii/B9780444636225000140

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, *30*, 829–858.

Wechsler, D. (1997). *WAIS-iii*. San Antonio, TX: Psychological Corporation.

Weed, E., McGregor, W., Feldbæk Nielsen, J., Roepstorff, A., & Frith, U. (2010). Theory of mind in adults with right hemisphere damage: What's the story? *Brain and Language*, *113*, 65–72.

Weinstein, M., Ben-Sira, L., Levy, Y., Zachor, D. A., Ben, I. E., Artzi, M., … Ben, B. D. (2011). Abnormal white matter integrity in young children with autism. *Human Brain Mapping*, *32*, 534–543.

White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the strange stories: Revealing mentalizing impairments in autism. *Child Development*, *80*, 1097–1117.

Winner, E., Brownell, H., Happé, F., Blum, A., & Pincus, D. (1998). Distinguishing lies from jokes: Theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain and Language*, *62*, 89–106.

Woods, R. P., Mazziotta, J. C., & Cherry, S. R. (1993). MRI-PET registration with automated algorithm. *Journal of Computer Assisted Tomography*, *17*(4), 536–546. https://doi.org/10.1097/00004728-199307000-00004

Xi, C., Zhu, Y., Niu, C., Zhu, C., Lee, T. M. C., Tian, Y., & Wang, K. (2011). Contributions of subregions of the prefrontal cortex to the theory of mind and decision making. *Behavioural Brain Research*, *221*, 587–593.

Yordanova, Y. N., Cochereau, J., Duffau, H., & Herbet, G. (2019). Combining resting state functional MRI with intraoperative cortical stimulation to map the mentalizing network. *NeuroImage*, *186*, 628–636.

Yordanova, Y. N., Duffau, H., & Herbet, G. (2017). Neural pathways subserving face-based mentalizing. *Brain Structure & Function*, *222*, 3087–3105.

Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., & Damasio, A. (2010). Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron*, *65*, 845–851.

Young, L., Dodell-Feder, D., & Saxe, R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia*, *48*, 2658–2664.

How to cite this article: Cohen-Zimerman S, Khilwani H, Smith GNL, Krueger F, Gordon B, Grafman J. The neural basis for mental state attribution: A voxel-based lesion mapping study. *Hum Brain Mapp*. 2021;42:65–79. https://doi.org/10.1002/hbm.25203