NeuroImage: Clinical 23 (2019) 101865

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

# The effects of lesion and treatment-related recovery on functional network modularity in post-stroke dysgraphia

Yuan Tao[a,*], Brenda Rapp[a,b,c]

[a] Department of Cognitive Science, Johns Hopkins University, USA
[b] Department of Neuroscience, Johns Hopkins University, USA
[c] Department of Psychological and Brain Sciences, Johns Hopkins University, USA

ARTICLE INFO

*Keywords:*
Modularity
Graph-theoretic measures
Network property
Cognitive recovery
Neuroplasticity
Dysgraphia

ABSTRACT

A better understanding of the neural network properties that support cognitive recovery after a brain lesion is important for our understanding of human neuroplasticity and may have valuable clinical implications. In fifteen individuals with chronic, acquired written language deficits subsequent to left-hemisphere stroke, we used task-based functional connectivity to evaluate the relationship between the graph-theoretic measures (*modularity*, *participation coefficient* and *within-module degree z-score*) and written language production accuracy before and after behavioral treatment. A reference modular structure and local and global hubs identified from healthy controls formed the basis of the analyses. Overall, the investigation revealed that less modular networks with greater global and lower local integration were associated with greater deficit severity and lower response to treatment. Furthermore, we found treatment-induced increases in *modularity* and local integration measures. In particular, local integration within intact ventral occipital-temporal regions of the spelling network showed the greatest increase in local integration following treatment. This investigation significantly extends previous research by using task-based (rather than resting-state) functional connectivity to examine a larger set of network characteristics in the evaluation of treatment-induced recovery and by including comparisons with control participants. The findings demonstrate the relevance of network modularity for understanding the neuroplasticity supporting functional neural reorganization.

## 1. Introduction

The human brain, like most complex biological and social systems, can be analyzed and understood in terms of the functions of its basic elements and the connections between them. The development of graph-theoretic analytic approaches (e.g., Watts and Strogatz, 1998; Albert and Barabási, 2002; Bullmore and Sporns, 2009; Sporns, 2013) and widely available computational tools for their application to human neuroimaging data (Rubinov and Sporns, 2010) have played a key role in the rapidly increasing interest and research efforts directed at understanding both functional and structural aspects of brain connectivity. Empirical support for the relevance of these approaches comes from the numerous findings that properties of brain networks are associated with various aspects of cognitive performance in healthy individuals (e.g., Bassett et al., 2011; Kitzbichler et al., 2011; Gallen et al., 2015; Yue et al., 2017). Here, we specifically investigate the application of these approaches to understanding functional network properties that may support behavioral recovery in post-stroke impairments affecting written language processes.

Research on the neural changes supporting recovery of language functions has overwhelmingly considered the neurotopography of the mean BOLD signal, investigating neural differences between individuals with post-stroke language deficits and healthy controls either by comparing time-points earlier vs. later in the course of recovery (e.g., Cao et al., 1999; Heiss et al., 1999; Saur et al., 2006; Jarso et al., 2013; Sebastian et al., 2016) or more directly by comparing pre- vs. post-treatment time-points (e.g., Fridriksson, 2010; Fridriksson et al., 2012; Thompson et al., 2010; Marcotte et al., 2012; Abel et al., 2015). More recently, although there has been some work evaluating recovery-related changes in resting-state as well as task-based functional connectivity (e.g., Warren et al., 2009; Van Hees et al., 2014; Zhu et al., 2014; Sandberg et al., 2015; Siegel et al., 2016), very few studies have used graph-theoretic approaches to investigate this issue (although see Duncan and Small, 2016; Siegel et al., 2018). In this investigation, we examine the impact of intensive language intervention on functional networks identified via task-based fMRI, in individuals with acquired

* Corresponding author at: Department of Cognitive Science, Johns Hopkins University, 3400 N Charles St., Baltimore, MD 21218, USA.
*E-mail address:* yuan.tao@jhu.edu (Y. Tao).

https://doi.org/10.1016/j.nicl.2019.101865
Received 17 September 2018; Received in revised form 22 April 2019; Accepted 19 May 2019
Available online 22 May 2019

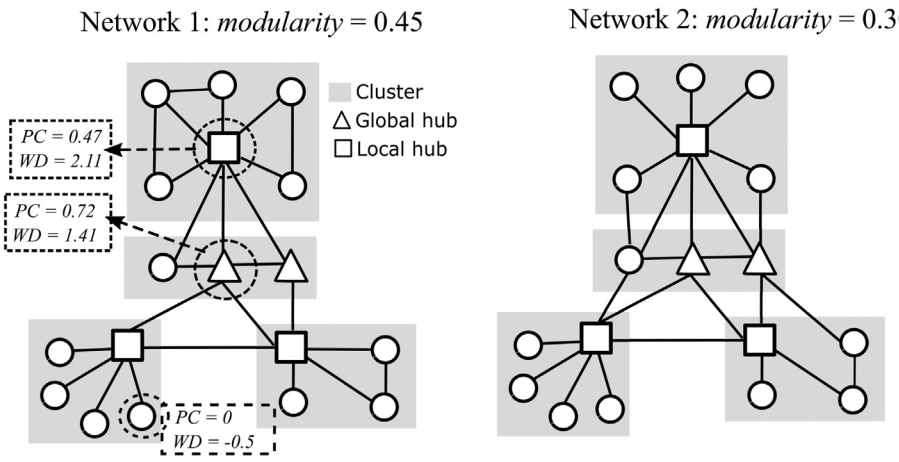


**Fig. 1.** Schematic depictions of two network structures. Left: Nodes are organized into clusters (i.e., modules) with denser within-module connections and sparser between-module connections. The *modularity* measure (Newman, 2006) quantifies overall modular segregation. Global hubs (triangles) are nodes with strong connections with multiple modules; local hubs (squares) have dense within-module connectivity. The global and local connectivities of a node are quantified by *participation coefficients* (PC) and *within-module degree z-scores* (WD), respectively (Guimera and Amaral, 2005). PC and WD values are reported for three of the nodes for illustrative purposes. Right: A network with the same total number of nodes and connections but with lower *modularity* due to fewer within-module connections and more between-module connections.

written language deficits subsequent to left hemisphere stroke. Greater knowledge of the neural changes that support recovery is important both for its clinical applications (e.g., identifying neurostimulation targets) and for its contribution to our understanding of the basic plastic capacities of the human brain.

### 1.1. Network properties

Analyses of the network properties of the brain have highlighted that, as with other complex systems, brain organization involves a balance between the forces of **segregation** and **integration**, operating at local and global scales (e.g., Hagmann et al., 2008; Bullmore and Sporns, 2009, 2012; Sporns, 2013). For example, high levels of local integration of elements (nodes) via strongly correlated activity (i.e., dense connectivity) results in segregation of nodes into sub-networks, referred to as modules, clusters or communities (Fig. 1). Segregated, modular structures of the brain are well-suited for carrying out specialized and automatized cognitive operations (Dosenbach et al., 2007; Bertolero et al., 2015; Yeo et al., 2014) and when systems with modular organization are damaged, the consequences can be relatively local and contained, without necessarily impacting the system at large (Achard et al., 2006; He et al., 2007; Honey and Sporns, 2008; Nomura et al., 2010). In addition to the local integration that supports modular organization, integration that takes place at a larger scale allows communication across modules (Fig. 1). Cross-module communication allows for, among other things, the flexible deployment and coordination of multiple processes to solve complex or novel problems (Cole et al., 2013; Braun et al., 2015a; Bertolero et al., 2015). A balance between segregation and global integration is clearly necessary for optimal use of available neural resources in support of the human brain's capacity for specialized cognitive operations as well as complex and flexible ones (Bullmore and Sporns, 2012; Sporns, 2013; Wig, 2017). A deeper understanding of the computational roles of various network properties can be achieved by investigating the relationship between network properties and different cognitive tasks and task demands (e.g., Kitzbichler et al., 2011; Braun et al., 2015a; Yue et al., 2017), aging (e.g., Meunier et al., 2009a), disruption caused by various neurological disorders (e.g., Buckner et al., 2009; Stam et al., 2006; Bassett et al., 2008; Lynall et al., 2010; Fornito et al., 2015; Siegel et al., 2018) as well as healthy learning (e.g., Bassett et al., 2011; Gallen et al., 2015) and relearning in the lesioned brain (Duncan and Small, 2016).

Graph-theoretic, network science approaches are based on the analysis of vertices and edges which, in the context of functional neuroimaging data, correspond to nodes (neural regions) and the connections (activation correlation) between pairs of nodes. These approaches have provided a large number of metrics for characterizing network properties (see Rubinov and Sporns (2010) for a review). In this investigation, we focus on the following three: *modularity*[1] (i.e., *Newman's Q* measure. Newman and Girvan, 2004; Newman, 2006), *participation coefficient* (PC) and *within-module degree z-score* (WD) (Guimera and Amaral, 2005).

*Newman's Q* is a widely used measure for quantifying, with a single value, the global (whole brain) *modularity* of a brain network (e.g., Bassett et al., 2011; Kitzbichler et al., 2011; Baggio et al., 2014; Gamboa et al., 2014; Arnemann et al., 2015; Duncan and Small, 2016; Gallen et al., 2015; Yue et al., 2017; Siegel et al., 2018). Given: (a) a matrix of pairwise connection strengths between all network nodes and (b) a pre-specified (reference) modular structure, *modularity* quantifies the extent of the pairwise connectivity *within* (reference) modules relative to what would be expected by chance given the total number of each node's connections across the entire network (referred to as a node's *degree*). Network *modularity* values approach 1 as the overall level of local integration increases relative to what would be expected by chance and they approach 0 as the numbers of connections within modules approach the value that what would be expected by chance. For example, in Fig. 1, the *modularity* value for Network 1 is 0.45 whereas for Network 2 (with the same number of nodes and connections) it is only 0.3. This is because, for each module in Network 2, there are fewer local, within-module connections as a proportion of the total degree of each node. (Note: in this example, connections are binarized as simply present/absent, but *modularity* can also be computed with graded connection strengths; Newman, 2004).

Whereas *modularity* characterizes the segregation properties of an entire network, each node can be characterized in terms of the extent to which it is locally or globally integrated. In this regard, two types of nodes are considered to play pivotal roles in determining the modular structure of a network: global and local "hubs" (Fig. 1). Global hubs are nodes with connections to multiple modules in a system and thus are also called *connectors*, or *connector hubs*. Conversely, local hubs, sometimes referred to as *provincial hubs*, or simply *hubs*, are nodes that are well-connected within a module (Note that we will refer to these two types as "global hubs" and "local hubs"). Global and local hubs can be characterized in terms *participation coefficients* (PC) and *within-module degree z-scores* (WD), respectively (Guimera and Amaral, 2005). The PC measure can be calculated for each node quantifying the extent to which the node's total connections are distributed across the modules of a network, with high PC values indicating high integration across the network. In Fig. 1, nodes with very different PC values are depicted, with high PC values indicating global hubs. The WD measure characterizes the degree of integration of a node within a module/cluster. It

[1] Note that we use the italicized term *modularity* to refer to the specific Newman's Q measure; unitalicized use of the term is used to refer to the general concept of modularity.

is a within-module z-score reflecting the number of a node's connections with the other nodes of a given cluster, expressed in standard deviation units computed from the interconnectedness of the other nodes within the same cluster. A node with a high WD value corresponds to a local hub within its module. In Fig. 1, the WD values of several nodes are indicated, with high values indicating local hubs. As illustrated, PC and WD values are independent from one another such that nodes can have high values in one, both or neither.

Clearly there is a close relationship between the number and strength of local and global hubs and global network properties. These relationships have been examined in a number of simulation studies (Achard et al., 2006; Sporns et al., 2007; Honey and Sporns, 2008; Alstott et al., 2009). This simulation work has consistently reported that lesions affecting global hubs (or similar constructs) have more widespread impact on network properties (such as *modularity*, *small-worldness*, etc.) than do lesions affecting local hubs (or similar constructs). For example, Sporns et al. (2007) specifically found that simulated lesions targeting global hubs created more segregated, less interconnected modules, while simulated lesions targeting local hubs had the opposite effect of decreasing local integration. Although lesions provide a powerful tool for gaining a deeper understanding of network organization (Albert et al., 2000; Albert and Barabási, 2002), the effects on network properties of actual lesions in the human brain and how the properties of disrupted networks change with recovery of function remain relatively unexamined topics.

### 1.2. Networks and lesions

Previous research concerned with the effects of lesions/disease on network properties has most often focused on specific networks/regions, such as the frontoparietal network (e.g., Zhu et al., 2014), the cognitive control network (e.g., Nomura et al., 2010), and on general deficit types including Alzheimer's Disease (e.g., Stam et al., 2006; Buckner et al., 2009), schizophrenia (e.g., Bassett et al., 2008; Lynall et al., 2010), among others. Instead, the investigation we report on here involves analyses of the relationship between whole-brain network properties and a specific language impairment. Therefore, we situate our review and discussion within the context of similar types of studies, leaving aside the many excellent papers examining connectivity properties relating to specific networks or more general deficit types (e.g., He et al., 2007; Stam et al., 2006; Buckner et al., 2009; Carter et al., 2010; Lynall et al., 2010; Nomura et al., 2010; Carter et al., 2012; Zhu et al., 2014).

#### 1.2.1. The consequences of hub damage on network properties and behavior

In one of the few studies to have examined the consequences of brain lesions on network properties, Gratton et al. (2012) tested the predictions of prior simulation studies regarding the consequences for *modularity* of damage to global vs. local hubs. To do so, they calculated *modularity* and the extent of global and local hub damage (using PC and WD, respectively) for individuals with chronic, focal lesions (stroke, TBI, tumors) affecting either the right or left hemisphere. They found that individuals with lesions had lower *modularity* than healthy controls, the lesioned hemisphere had lower *modularity* values than the non-lesioned hemisphere but that the non-lesioned hemisphere still had lower *modularity* than the same hemisphere in healthy controls (but see Arnemann et al. (2015) and Siegel et al. (2018) for findings that individuals with chronic brain lesions did not differ in *modularity* from healthy controls). Moreover, they found that *modularity* was not correlated with lesion size or with local hub damage but was instead correlated with the extent of damage to global hubs. In sum, their finding that actual lesions that damage to global hubs had a greater impact on network *modularity* than did local hub damage was generally consistent with the simulation-based predictions described above.

Relatedly, Warren et al. (2014) examined the behavioral consequences of damage to target vs. control locations (which generally corresponded to global hubs and local hubs, respectively) in individuals with chronic, focal lesions recruited because their lesions primarily affected one type of hub or the other. They found that lesions to target locations were associated with impairment to a larger number of cognitive domains than was damage to control hubs, again consistent with the general simulation prediction of more widespread consequences of damage to global vs. local hubs.

#### 1.2.2. Modularity and recovery

Three recent studies have specifically examined the relationship between *modularity* and recovery of function with all three finding that higher *modularity* values were associated with greater behavioral gains (Arnemann et al., 2015; Duncan and Small, 2016; Siegel et al., 2018). Siegel et al. (2018) examined longitudinal changes of *modularity* values in a very large cohort of stroke cases at three time-points from the acute to chronic stage (2 weeks to 1 year post-stroke), finding that *modularity* values were below those of healthy controls at the acute time-point, and then increased to normal levels after 3 months. Furthermore, with respect to behavior, they found that better recovery was associated with larger *modularity* increases (in some, but not all, cognitive domains).

Arnemann et al. (2015) and Duncan and Small (2016) specifically considered *modularity* in the context of treatment for cognitive deficits. Arnemann et al. (2015) specifically focused on the question of whether *modularity* values prior to therapy *predict* response to treatment. They examined pre-treatment *modularity* in individuals (primarily with brain trauma) who received 5 weeks of cognitive training for attention/self-regulation and found that higher pre-treatment *modularity* values were associated with greater subsequent improvement in attention and executive functions from pre to post training. Relatedly, in a study by Gallen et al. (2015) of older healthy adults who underwent 12 weeks of cognitive control training, the authors found that baseline *modularity* values positively predicted subsequent training-related gains in performance in the training group but not an untrained control group. Duncan and Small (2016), instead, focused on the relationship between *changes* in *modularity* and language performance before and after 6 weeks of spoken language therapy in individuals with chronic post-stroke aphasia. They found that, even controlling for pre-treatment language severity, *modularity* values increased from before to after treatment and were positively and significantly correlated with the amount of language improvement.

### 1.3. The current study

The current study examines the relationship between network properties: *modularity*, *participation coefficient* (PC), *within-module degree z-score* (WD), and recovery of language function in individuals who received behavioral treatment for written language deficits (dysgraphia) acquired subsequent to a single left-hemisphere stroke. Unlike the vast majority of previous studies of network properties that have either evaluated white matter connectivity or resting-state functional connectivity, the current study evaluated network properties of task-related BOLD response in individuals performing a spelling task. We specifically analyzed what Norman-Haignere et al. (2011) and Al-Aidroos et al. (2012) refer to as "background connectivity" which is the residual of the task-based GLM (see also, Cole et al., 2013, 2014). Studies have shown that the "background connectivity" is very similar to the functional connectivity observed at rest, but with additional and significant task-related components (Fair et al., 2007; Cole et al., 2014). A number of studies have examined connectivity changes associated with recovery of function using task-related BOLD. However, these have used measures such as effective (e.g., Seghier et al., 2010, 2012; Kiran et al., 2015; Meier et al., 2016) or functional (Abel et al., 2015; Sandberg et al., 2015) connectivity to evaluate connectivity strength within specific networks rather than measures that characterize more general network properties.

Most studies involving assessment of neural changes associated with

recovery of language functions have considered deficits affecting spoken naming (e.g., Van Hees et al., 2014; Kiran et al., 2015; Sandberg et al., 2015), speech production (e.g., Duncan and Small, 2016; Marangolo et al., 2016) or sentence processing (e.g., Thompson et al., 2010). Instead, given the increasingly prominent role of written language production in communication in daily life (e-mail, texting, etc.), we examined changes associated with treatment induced recovery of spelling skills. Meta-analyses of neuroimaging studies of spelling (Purcell et al., 2011; Planton et al., 2013) have shown that, in neurologically healthy individuals, spelling recruits a highly left-lateralized network of areas that include: the left ventral occipital-temporal region known as the *visual word form area*, or VWFA (e.g., Cohen et al., 2002; McCandliss et al., 2003), that is also recruited for reading, as well as the left inferior frontal gyrus, superior frontal gyrus, supplementary motor area, intra-parietal sulcus/superior parietal lobule, supramarginal gyrus, and bilateral superior temporal gyrus/sulcus. Spelling involves a number of cognitive functions that are supported by this network of brain areas. These include the retrieval and selection of stored representations of word spellings from orthographic long-term memory as well as orthographic working memory processes that are necessary for the serial production of the constituent letters that make up a word's spelling. In addition, there are sublexical processes that map language sounds to possible/plausible spellings that rely on information developed on the basis of an individual's experiences with the systematic relationships between the sounds and letters (or other graphic symbols) of their language. These central processes of spelling serve to generate letter strings that can be expressed in a variety of spelling formats such as handwriting, typing, oral spelling, etc. by means of "peripheral" motor planning and production processes (see Rapp et al. (2016) for a review of spelling processes and the neural substrates that support them). The dysgraphia treatment protocol that was used in this investigation (although it involves written practice) targets these central spelling processes that support orthographic production across multiple production formats and modalities.

In this study we address the following questions. Regarding pre-treatment network properties, we ask: (1) Do pre-treatment *modularity*, PC or WD index deficit severity? (2) Are pre-treatment *modularity*, PC or WD related to subsequent recovery of function? Regarding the network properties associated with recovery, we ask: (3) Do *modularity*, PC or WD change in response to treatment, and, if so, what are the relationships among changes in these properties? (4) How do network changes relate to behavioral changes? (5) How are treatment-related changes distributed throughout the brain, in terms of brain regions and hemispheres? Answers to these questions provide a complementary view of neural recovery to that provided by the more traditional examinations of the distribution of local, mean BOLD changes and, in so doing, this approach allows us to deepen our understanding of the network properties involved in the neuroplastic changes that support the recovery of complex cognitive functions.

## 2. Methods and materials

### 2.1. Participants

Participants consisted of fifteen individuals (4 females, mean age 61 ± 10) with no history of reading/spelling disabilities who suffered acquired dysgraphia due to a single left hemisphere stroke (> 1 yr post stroke, mean 58 ± 34 months). Years of education ranged from 12 to 19 and all but three (including one who was ambidextrous) were (according to self-report) right-handed prior to the stroke, although 8 of the fifteen (including the three original left-handed participants) used their left hand for writing after the stroke. See Table 1 for further details. Fig. 2a depicts the distribution of the lesions, indicating greatest density of damage in the region of the left insula and the left superior longitudinal fasciculus. Participants received an average of 26 bi-weekly dysgraphia rehabilitation sessions and functional and structural MRI scanning was carried out before and after the rehabilitation period.

In addition, there were two groups of age-matched, neurologically healthy individuals with no history of reading/spelling disabilities. Control Group 1 consisted of 10 participants (8 females, age: 60.7 ± 12, education: 12–18 years, all right-handed) who were administered the same spelling task protocol during fMRI scanning as the participants in the Lesion Group. Control Group 2 consisted of 13 participants (10 females, age: 57 ± 7.39, with 12–18 years of education (three were left-handed) who performed a picture version of the spelling task (both tasks are described below). All participants provided informed consent following procedures approved by Johns Hopkins University Institutional Review Board.

### 2.2. Cognitive/language assessments and treatment

In addition to an evaluation of their spelling performance (JHU Dysgraphia Battery; PALPA 40, Kay et al., 1992), each participant was administered an extensive battery of cognitive and language tests before and after dysgraphia treatment. Test results are reported in Fig. 2b (group data) and Table A1 (individual participant data) for the following cognitive and language domains: oral reading of single words (PALPA 35; Kay et al., 1992), written single-word comprehension (PALPA 51; Kay et al., 1992); auditory single-word comprehension (Northwestern Naming Battery; Thompson et al., 2012); visual recognition memory (Doors and People Test; Baddeley et al., 1995); semantic comprehension (Pyramid and Palm Trees; Howard and Patterson, 1992); spoken picture naming (Northwestern Naming Battery; Thompson et al., 2012).

For the dysgraphia treatment and fMRI scanning, 40 Training and 30 Known Words were selected for each individual. These individualized sets were identified by administering a number of different word lists twice for spelling to dictation. Training Words were those on which letter accuracy was between 25% and 80% on the two administrations and Known Words had 100% accuracy on the two assessments. The average length of the Training Words across participants is 6.5 letter (ranged from 4.3 to 8.3), and the average frequency is 19.4 (ranged from 6 to 82. *The English Lexicon Project*. Balota et al., 2007). For the dysgraphia treatment, a spell-study-spell technique (Rapp and Kane, 2002) was administered for approximately 60–80 min sessions, typically 2×/week. Each training trial was as follows: (1) The individual heard a target word, repeated it, and attempted to write the spelling, (2) Regardless of accuracy, the individual was shown the correct spelling while the experimenter said aloud the word's letters, the individual copied the word once and was instructed to study the word. If the word had been spelled correctly at Step 1, then the training trial ended and the experimenter continued to the next item. Otherwise, the word was removed from view and Steps 1 and 2 were repeated until the word was spelled correctly, or for a maximum of 3 times before moving to the next item. Treatment sessions continued until participants achieved 90% or greater letter accuracy on the Training items on two consecutive sessions or if performance remained constant over 6 sessions.

Treatment-induced improvements of the group were statistically evaluated using a generalized linear mixed-effects model (LMEM, binomial family) because participants had different sets of individually selected Training Words. The dependent variable was percentage letter accuracy for all the Training Words from each individual participant at each of the three time-points (pre, post and a 3-month follow-up). The fixed effect variables were: word length, word frequency, and time-point (simple-coded with pre-treatment as the reference level). The following random-effects were included: by-participant random intercept and slopes for word length, word frequency, and time-point, and by-words random intercept and slope for time-point. To evaluate whether treatment effects generalized beyond the trained set of items, we evaluated the pre- to post-treatment improvement for each participant on a larger pool of untrained words (each participant had a

**Table 1**
Characteristics of participants in the Lesion Group. "Spelling severity" corresponds to the error rate on the spelling-to-dictation test PALPA 40 (Kay et al., 1992); "Improvement on Training Words" corresponds to the percent change in letter accuracy on spelling-to-dictation assessments of Training Words administered before and after the training period.

| Subject ID | Sex | Age | Education (years) | Handed-ness | Month post-stroke | Lesion vol. (mm^3) | # Rehab sessions | Spelling severity (error rate) | Improvement on training words |
|---|---|---|---|---|---|---|---|---|---|
| 1 – ABS | M | 58 | 18 | R | 97 | 198,480 | 19 | 30% | 44% |
| 2 – AEF | F | 55 | 16 | R | 101 | 280,768 | 24 | 30% | 27% |
| 3 – DSK | M | 67 | 16 | R | 59 | 207,472 | 11 | 52.5% | 21% |
| 4 – DTE | F | 80 | 18 | R | 14 | 111,968 | 48 | 72.5% | 24% |
| 5 – ESG | M | 62 | 16 | L | 38 | 155,232 | 27 | 90% | 31% |
| 6 – FCE | M | 64 | 12 | R | 119 | 68,456 | 19 | 72.5% | 43% |
| 7 – JGL | F | 72 | 16 | R | 32 | 74,840 | 48 | 57.5% | 31% |
| 8 – KMN | M | 55 | 15 | R | 28 | 96,192 | 48 | 77.5% | 24% |
| 9 – KST | M | 61 | 14 | R + L | 46 | 46,400 | 29 | 35% | 33% |
| 10 – MSO | M | 45 | 18 | R | 103 | 217,440 | 30 | 57.5% | 66% |
| 11 – PQS | M | 54 | 18 | R | 17 | 143,128 | 17 | 47.5% | 27% |
| 12 – RFZ | M | 60 | 18 | R | 46 | 98,984 | 16 | 17.5% | 19% |
| 13 – RHH | M | 45 | 16 | R | 82 | 145,368 | 18 | 7.5% | 27% |
| 14 – RHN | F | 75 | 19 | L | 27 | 17,712 | 16 | 37.5% | 26% |
| 15 – TCK | M | 69 | 16 | R | 68 | 41,096 | 26 | 55% | 44% |

different number of words, ranging from 40 to 454, mean = 201). Generalization was evaluated with the same linear mixed-effects model as the one used for Training Words, except the dependent variable was the accuracy of the untrained words.

For the purposes of relating neural measures to behavior (see Section 2.5.7), the *severity* of the spelling impairment of each individual was quantified as the pre-treatment performance on the PALPA 40 spelling sub-test, consisting of 40 high/low frequency and high/low imageability words presented for spelling to dictation (Kay et al., 1992). Note that we quantified spelling severity based on this standard test rather than the Training Words because each individual in the Lesion Group had a different, individualized set of Training Words, making this an inappropriate comparative measure of severity across the group; *Pre-post improvement* in spelling was quantified as the percentage change in letter accuracy on Training Words measured at the pre- and post-treatment time-points (Fig. 2b). For this measurement, each

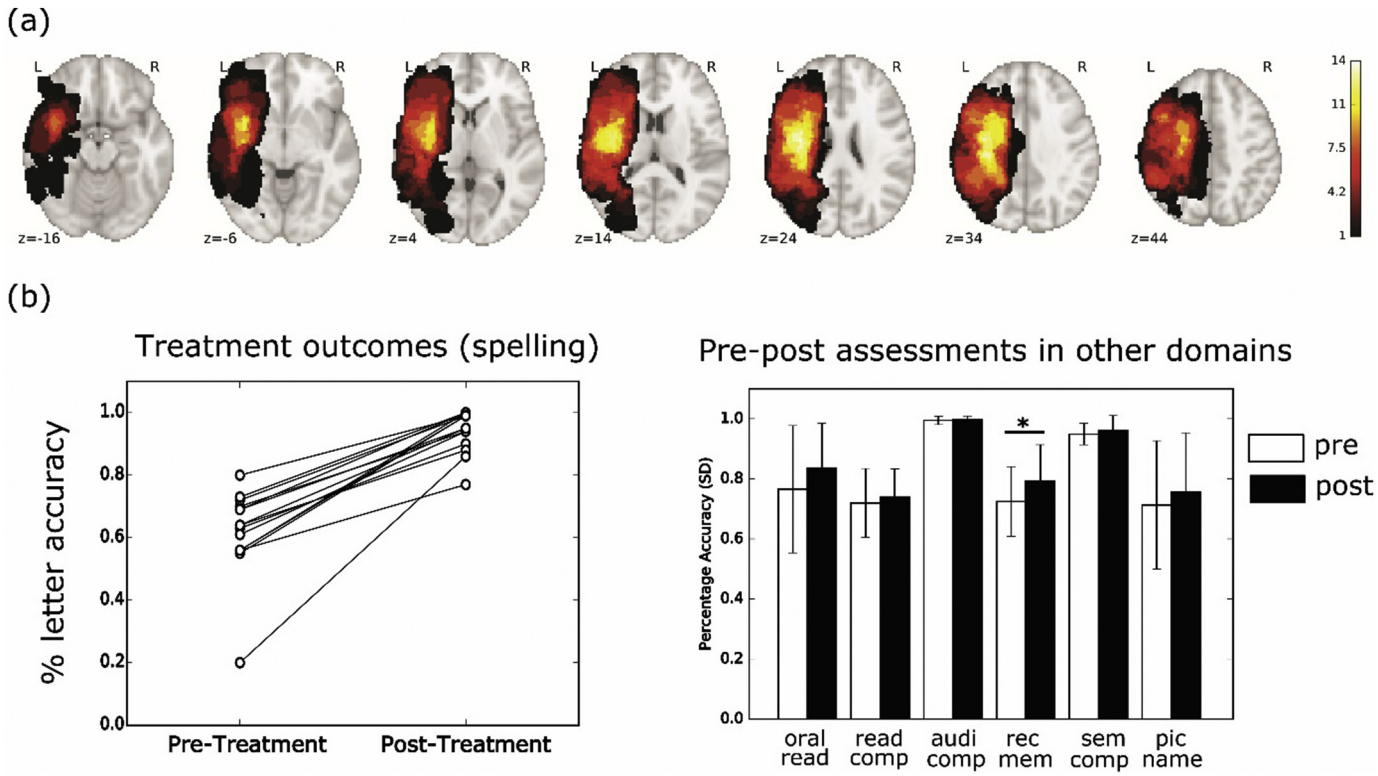


**Fig. 2.** (a) Lesion distribution of the 15 participants in the Lesion Group. The highest concentration of lesioned tissue is in the left insula and along the superior longitudinal fasciculus. (b) Results of pre- and post-treatment cognitive assessments. Left: Spelling outcomes. Pre- and post-treatment percent letter accuracies for Training Words for each individual. Improvements range from 19% to 66% (mean 32.47 ± 11.81, median 27); all are statistically significant ($p < .05$). Right: Pre- to post-treatment assessments in other language/cognitive domains. Only the visual recognition memory task (Doors and People Test) shows a significant change ($p < .05$, uncorrected). Oral-read: single-word oral reading (PALPA 35; Kay et al., 1992); Read-comp: single-word written comprehension (PALPA 51; Kay et al., 1992); Audi-comp: auditory single-word comprehension (Northwestern Naming Battery; Thompson et al., 2012); Rec-mem: recognition memory (Doors and People test; Baddeley et al., 1995); Sem-comp: semantic association (Pyramids and Palm Trees; Howard and Patterson, 1992); Pic-name: oral picture naming (Northwestern Naming Battery; Thompson et al., 2012).

participant's set of Training words was presented for spelling to dictation once or twice both before and after the treatment period and the results at each time point were averaged. The *severity* and *improvement* of each participant are reported in Table 1.

### 2.3. Neuroimaging experiment

Individuals in the Lesion Group were scanned on 2 days at each of the pre- and post-treatment time-points (an average of 6.3 days apart). Each of the two scanning sessions included two 7.7-min runs of the spelling task, generating a total of 30.8 min of data (4 runs in total) for each participant at pre- and post-treatment (Fig. S1). In addition to the spelling task, participants performed several tasks during fMRI scanning and were evaluated with multiple structural scanning protocols, including the T1-weighted imaging we report on here.

The spelling task protocol included a Spelling Probe task and a Case-Verification control task (Rapp and Lipka, 2010) with event-related designs.[2] Each spelling probe trial consisted of the following sequence of events: (1) task prompt ("Is the letter in the word?") presented both visually and auditorily for 1800 ms; (2) a 500 ms central fixation cross, (3) an auditory target word plus a variable period of silence depending on the word length, for a total duration of 1500 ms during which the fixation cross remained on the screen, (4) A response period consisting of a single visually presented probe letter presented for 1500 ms, followed by a 1700 ms fixation cross. During this time period, participants responded whether or not the letter was in the spelling of the target word with a button press using with their left index or middle finger. There was a random inter-trial interval (2–7.5 s) between trials. The Case-Verification trials were identical, except that the task prompt was "Is the letter uppercase?" and participants were instructed to ignore the auditory stimulus and judge the case of the letter with a button press to indicate if the letter was upper or lower case. Control Group 1 was administered the same experimental protocol (both Spelling Probe and Case Verification tasks) and as indicated, Control Group 2 was administered a modified protocol which was the same as the one just described except that the auditorily presented target words were replaced with black and white line-drawings and participants were instructed to respond on each trial if the visually presented letter was in the spelling of the word depicted in the drawing (or to judge the letter case for Case-Verification trials). Participants were familiarized with the names of the drawings prior to the experiment.

For the Lesion Group, each run had 45 trials consisting of the following: 15 Known Word spelling-probe trials, 15 Training Word spelling-probe trials, and 15 Case-Verification trials with words that were matched in length and frequency to those of the spelling trials. In this way, for each participant, each scanning session included presentation of 30 Known Words and 30 Training Words that were randomly selected from the individual's training list. The same stimuli were presented at both scanning sessions although their order within runs varied. In order to minimize task-switching costs, 3–6 trials of each task were presented consecutively in "mini-blocks". For the Control Groups the same procedures were followed except that the stimuli were selected to represent a range of lexical frequencies and lengths: word frequency ranged from 0.27 to 557.12 with a median = 8.3 (*English Lexicon Project*. Balota et al., 2007) and they were 4 or 7 letters in length. Participants in the Control Groups also completed four runs of the task but were scanned within the same day.

### 2.4. Imaging data acquisition

All MRI data were collected using a Phillips 3 T scanner at the F.M. Kirby Research Center for Functional Brain Imaging (Baltimore, MD). The acquisition parameters for all participant groups were as follows: TR = 1500 ms, TE = 30 ms, FOV = 216*120*240 mm (ap, fh, rl), flip angle = 65°, voxel dimension = 1.875*1.875*3 mm (1.5 mm interslice gap), data matrix = 128*128*27. One run contained 308 TRs (7.7 min). The T1-weighted structural MRI acquisition parameters were as follows: TR = 6 ms, TE = 2.91 ms, FOV = 256*256*176 mm (ap, fh, rl), flip angle = 9°, voxel dimension = 1*1*1 mm, data matrix = 256*256*176.

### 2.5. fMRI data analysis

#### 2.5.1. Pre-processing

Functional data were preprocessed with FEAT in FSL 5.0.9 (Jenkinson et al., 2012) with the following pre-processing steps: removal of the first 4 volumes of each run, motion correction with MCFLIRT, slice-timing correction, non-brain tissue removal with BET, spatial smoothing using a 5 mm FWHM Gaussian kernel, grand-mean intensity normalization, high-pass temporal filtering (0.01 Hz). Registration was carried out with FSL's two-step method: functional images were first registered to the subject's T1 image with the *boundary-based registration* (BBR) method, then to standard MNI space (spatial resolution 2 mm, Jenkinson et al., 2002; Greve and Fischl, 2009). A lesion mask for each participant was manually drawn on the original T1 image with MRIcron (http://people.cas.sc.edu/rorden/mricron/index.html). The masks were then aligned to standard MNI space using the transformation calculated with the T1 image.

#### 2.5.2. GLM analysis

A general linear model (GLM) analysis was evaluated at each voxel with FEAT in FSL (Jenkinson et al., 2012). The model included the following regressors: task prompt, Known Word trials, Training Word trials, and Case-Verification trials (these trial-type regressors modeled the 3 s period including target word presentation, visual letter probe and response), and the six motion parameters as confound regressors. The fixation periods and inter-trial intervals were left unmodelled. All the task-related regressors were convolved with a double-gamma hemodynamic response function provided in FSL.

#### 2.5.3. Background functional connectivity estimation

Following Norman-Haignere et al. (2011) and Al-Aidroos et al. (2012), we quantified background functional connectivity using the residual time-series of the GLM. The analysis was performed by *Nilearn*, a Python package for neuroimaging data analysis (Abraham et al., 2014). We used the 264 locations identified by Power et al. (2011), excluding 29 locations in subcortical areas based on the segmentation template provided in FSL (*MNI152_T1_2mm_strucseg.nii*). For the remaining 235 locations, we extracted the averaged time-series from 5 mm-radius spheres centered at each location (Fig. 3a). For each participant in the Lesion Group, spheres in which > 25% of voxels were lesioned (32 voxels) were excluded from further analysis. Prior to calculating pairwise-connectivity, we additionally regressed out the first derivatives of the 6 motion parameters and the averaged time-series of the brainstem, the cerebellum, the basal ganglia, and the ventricles defined by the segmentation template in FSL. For each run (7.7 min), we calculated pairwise Pearson correlations for the 235 nodes (and fewer for each participant with lesion, see below) which were Fisher's z-transformed and converted to absolute values. For each time point and each participant, the correlation matrices of the four runs were averaged to yield one connectivity matrix. Because, for each individual, nodes with > 25% lesioned voxels were excluded, the individual connectivity matrices had fewer than 235 rows/columns, ranging from 207 to 232. The analysis procedure is depicted in Fig. S1.

#### 2.5.4. Identifying a "reference" modular structure

The graph-theoretic measures used to evaluate the structure and organization of the connectivity patterns (*modularity*, *participation*

[2] This experiment was originally designed for a "subtraction-based" GLM analysis and, for this reason, included a control task.

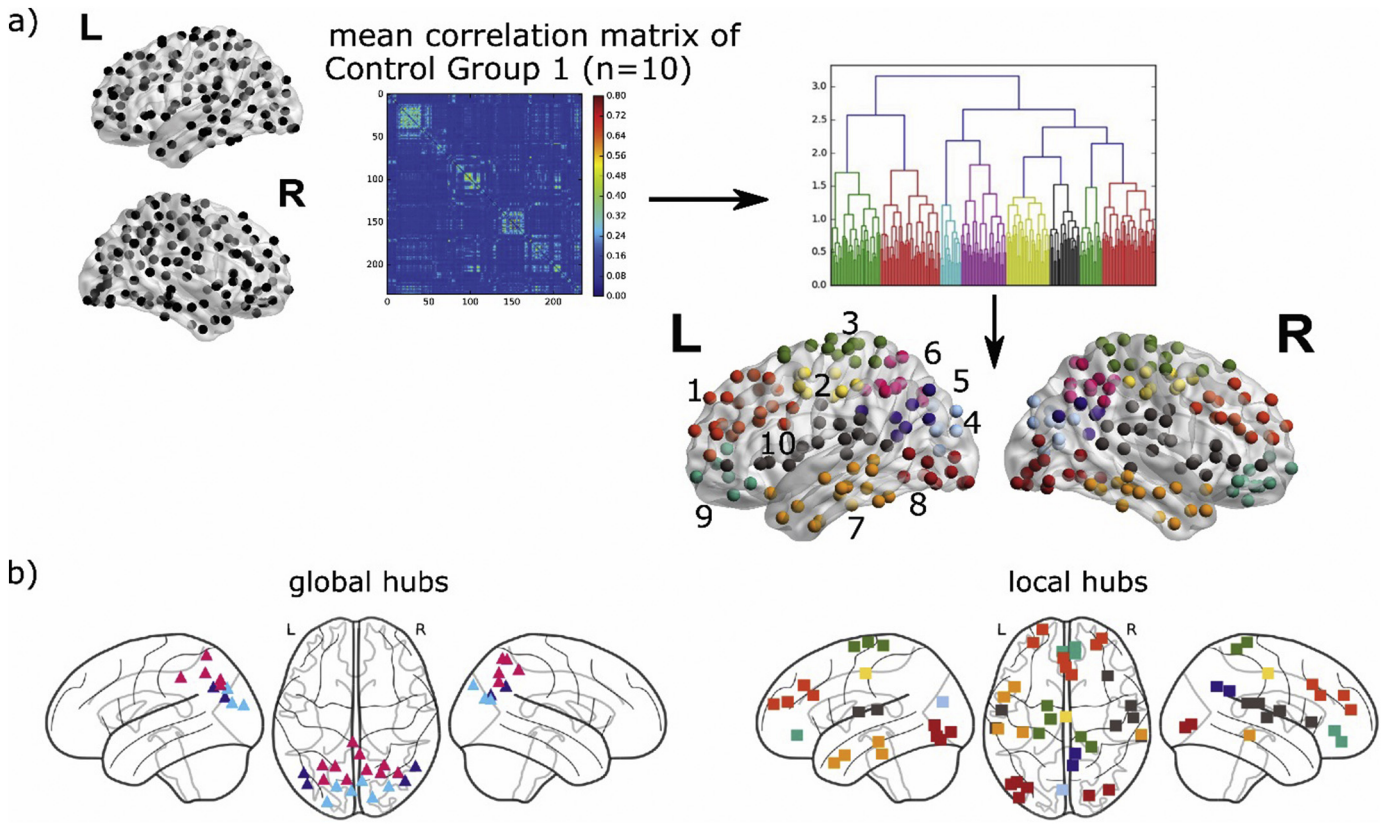


**Fig. 3.** (a) The reference modular structure was calculated with hierarchical clustering using the mean background connectivity correlation matrix of Control Group 1 ($n$ = 10). A 10-cluster partitioning scheme derived from the dendrogram corresponded to the "reference" modular structure used in subsequent analyses (see Table 2 for information about the 10 clusters). (b) Distribution of global (left) and local hubs (right) identified in Control Group 1, color-coded for their cluster membership as in (a). The global hubs ($n$ = 20; 10 in each hemisphere) were identified as the nodes with high *participation coefficient* (PC) and were located in bilateral posterior parietal and occipital lobes. The local hubs ($n$ = 37; ranging from 0 to 7 per cluster) were identified as the nodes with high *within-module degree z-scores* (WD).

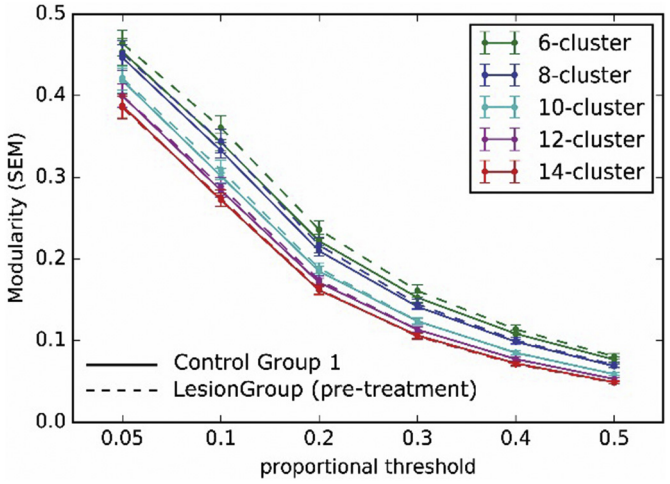


**Fig. 4.** *Modularity* values calculated based on different dendrogram partitionings (i.e., different numbers of clusters) and proportional thresholds. Cluster numbers range from 6 to 14 and proportional thresholds from 5% to 50%. Results from Control Group 1 (solid line) and the Lesion Group (dashed line) at the pre-treatment time-point are shown. For both groups, *modularity* values are highly similar and decrease as the number of clusters and the proportional threshold increase. The 10-cluster structure and 0.4 proportional threshold are used for all subsequent analyses.

*coefficient* (PC) and *within-module degree z-score* (WD)) require a "reference" modular structure (functional partitioning) relative to which the connectivity values of each participant are evaluated. Because we used a task-based connectivity measure rather than the more commonly used RS-fMRI connectivity, there was no established functional partitioning. Instead, we developed a characterization of the brain's modular organization during spelling using the background connectivity values from Control Group 1. Note that we used a single reference modular structure to calculate *modularity* values following Duncan and Small (2016) and Siegel et al. (2018) rather than computing individual modular structures for each individual participant (e.g., Gratton et al., 2012). We adopted this approach in order to have a single un-biased measure that could be used for comparison across subject groups and time-points. Further, we assumed that the participants with lesions were not an appropriate source for the reference modular structure given variability in terms of lesion size and location as well as in recovery and response to treatment.

A reference modular structure was generated based on the averaged correlation matrix of Control Group 1 and using the hierarchical clustering (agglomerative) method with Ward's criterion implemented in *SciPy* (https://scipy.org) (Fig. 3a). Hierarchical clustering is an efficient clustering approach and has been shown to be particularly suitable for performing brain parcellation (Thirion et al., 2014). One of the most valuable characteristics of hierarchical clustering is that it does not require the number of clusters to be pre-defined. Instead, the algorithm constructs a dendrogram based on the distances between the nodes. A specific clustering solution can be derived from the dendrogram by selecting a cut-off threshold in terms of dendrogram height or number of clusters. Subsequent analyses are based on the 10-cluster solution

(Fig. 3a). This solution was adopted given that our examination of *modularity* results indicated (see Fig. 4) that the number of clusters had a very similar effect on *modularity* of participants in both the Lesion and the Control Groups, such that *modularity* decreased as the number of clusters increased. The 10-cluster simply represented the middle value of those examined and, moreover, the results were most reliable across the two Control Groups with the 10-cluster structure (Fig. S2).

*2.5.5. Evaluating network properties*

The following network analyses were carried out with the Python version of the Brain Connectivity Toolbox (Rubinov and Sporns, 2010. https://pypi.org/project/bctpy). For these analyses, each correlation matrix was first thresholded to preserve the strongest connections and then converted to an undirected binary graph (in which all connection strengths correspond to 1's or 0's). To select a threshold, we examined a range of values (ranging from the top 5% to the top 50% of connection strengths) to evaluate the reliability of the results (Fig. 4 and Fig. S2). Like the number of clusters, the proportional threshold also had a similar systematic effect across participant groups such that *modularity* decreased as the cut-off threshold value increased (i.e., more connections included). All subsequent analyses are based on matrices using a top 40% proportional threshold which yielded highly reliable results across the two Control groups.

*2.5.5.1. Modularity*. For each individual (Lesion and Control Groups) we calculated a global *modularity* measure (Newman, 2006) for each (binarized) connectivity matrix relative to the reference modular structure described in Section 2.5.4 (Note that, as indicated earlier, the connections of the lesioned nodes of each participant with lesion were discarded at an earlier analysis stage).

$$Modularity\ Q = \frac{1}{L}\sum_{i,j\in N}\left(a_{ij} - \frac{k_i k_j}{L}\right)\delta \quad (1)$$

$Q$ is calculated for the set of all $N$ nodes. $L$ is the number of total connections, $a$ is the connection value between node i and node j, $k$ is the degree of a node, $\delta = 1$ if node i and j are in the same cluster, otherwise $\delta = 0$ (Newman, 2006).

*2.5.5.2. Global and local hubs*. We first identified "reference" global and local hubs with Control Group 1 data using the two node-level graph-theoretic measures, *participation coefficient* (PC) and *within-module degree z-score* (WD) respectively (Guimera and Amaral, 2005). For Control Group 1, using the modular structure defined in Section 2.5.4, we calculated the average across-participant PC and the WD score for each node, and then identified the reference global and local hubs as those nodes with PC and WD values greater than one standard deviation above the mean across the 235 nodes (Fig. 3b). Then, for use in subsequent analyses, we computed the average PC and WD values of these global and local hub nodes for each individual (in all groups) at each time-point.

$$PC_i = 1 - \sum_{c\in C}\left(\frac{k_i(c)}{k_i}\right)^2 \quad (2)$$

*Participation coefficient* (PC) is calculated for each node, *k(c)* is the number of connections within a cluster, *C* is the set of all clusters of the network (Guimera and Amaral, 2005).

$$WD_i = \frac{k_i(s) - \overline{k}_s}{\sigma_s} \quad (3)$$

*Within-module degree z-score* (WD) is calculated for each node, *k(s)* is the number of connections of node i within its own cluster *s* (within-cluster degree), $\overline{k}$ and $\sigma$ are the mean and standard deviation of the within-cluster degree distribution (Guimera and Amaral, 2005).

We also examined *modularity*, average PC and WD for the left (ipsilesional) and the right (contralesional) hemispheres separately. Specifically, for *modularity*, we calculated the per hemisphere values using only the within-hemisphere connections and the same reference modular structure but using only the left or the right hemisphere nodes of each cluster. For average PC and WD, we simply divided the identified hubs (Fig. 3b) into left and right sets, and calculated the mean *participation coefficient* or the *within-module degree z-score* values for either hemisphere.[3]

*2.5.6. Evaluating the network properties across participant groups and treatment time-points*

The three network properties (*modularity*, average PC and WD) were calculated for each participant and each time-point. First, as the reference modular structure was derived from Control Group 1, we compared the two Control Groups to validate the measures in the Control Group 2 (unpaired *t*-test). Note that because analyses of the three network measures were theoretically motivated and planned, they are not subject to multiple comparison correction.

Second, to assess the effects of lesion, we compared the three network properties of the participants with lesion before treatment to the healthy controls using unpaired *t*-tests. Finally, we evaluated the pre- to post-treatment differences of the participants in the Lesion Group by means of paired t-tests. In addition, the post-treatment values were also compared to the healthy controls with unpaired t-tests. All the comparisons are two-tailed. For the hemisphere analyses, changes in each neural measure and interactions between hemisphere and time-points for the participants in the Lesion Group were evaluated using repeated-measures two-way ANOVAs.

*2.5.7. Evaluating the relationship between network measures and spelling performance*

For evaluating the relationships between network measures and behavior, we developed two sets of linear regression models, one for the pre-treatment time-point and one for the pre- to post-treatment changes (with separate models for each of the graph-theoretic measures, for a total of 6 models). First, for the pre-treatment time-point, we constructed a regression model that predicted each neural measure (*modularity*, average PC and average WD) using the following predictors: *spelling severity* (see Table 1), age, years of education, lesion volume, time after stroke, and motion during fMRI scanning (RMS mean displacement). Second, for the pre- to post-treatment changes, we constructed a regression model to predict changes for each of the three neural measures, including the same predictors as in the first set of analyses, plus also: *improvement* on the Training Words from pre- to post-treatment (log transformed), the accuracy of the Training Words at pre-treatment, and the number of treatment sessions (Table 1). All regression analyses were carried out with the *lm* function in R (R Core Team, 2017) and the effects were visualized with the R-package **effects** (Fox, 2003). Note that the inclusion of pre-treatment Training Word accuracy as well as the log-transformation of the Training word improvement scores alleviate the challenges treatment studies face related to the fact that, due to different cross-participant starting accuracies, there are different cross-participant possibilities for improvement.

[3] One caveat is that the node-level graph-theoretic measures, *participation coefficient* and *within-module degree z-score*, and the resulting reference hubs were all calculated on the basis of whole-brain connectivity. Those measures would differ if calculated with within-hemisphere connections only, resulting in the possibility that, in that case, different nodes could be identified as hubs. However, defining a different set of hubs based on within-hemisphere connectivity has fundamental conceptual implications that would require more systematic investigation. Thus, in the current study we focused on the same set of hubs defined on the basis of the whole-brain connectivity structure. Essentially, the hemisphere analyses evaluated whether the whole-brain effects seen in the Lesion Group were driven by one hemisphere.

## 3. Results

### 3.1. Analysis 1: behavioral results

#### 3.1.1. Spelling and language/cognitive assessments

From pre- to post-treatment, spelling accuracy for Training Words significantly improved for all participants (Fig. 2b, Table 1) with improvements in percentage letter accuracy ranging from 19% to 66% (median = 27%). These improvements – as well as retention at the 3-month follow-up time point – were statistically evaluated using a generalized linear mixed-effects model. The results showed that there was a reliable treatment effect from pre- to post-treatment for the Training Words (beta = 3.643, $p < 2e\text{-}16$). Furthermore, performance remained significantly higher at the follow-up time-point than at pre-treatment (beta = 2.019, $p < 2e\text{-}16$) indicating that the treatment gains were long-lasting. The spelling severity measure, PALPA 40, was also administered after treatment and the participants showed significant improvement ($t(14) = 5.88$, $p = .00004$). As the PALPA 40 test only contained 40 words, to examine the effect of treatment generalization, we also evaluated the pre- to post-treatment changes using a larger pool of untrained words for each participant (each participant had different number of words, ranging from 40 to 454, mean = 201). The results showed that the spelling performance on the larger set of untrained words also improved significantly from pre- to post-treatment (beta = 0.428, $p = .001$).

In contrast, performance on other language and cognitive domains (see Table A1 and Fig. 2b) remained unchanged except for the visual recognition memory test (Doors and People Test; Baddeley et al., 1995) ($t(14) = 2.25$, $p = .04$). Given the novel nature of this task (old/new recognition memory for pictures of doors), improvement could have been due to test-retest increase in familiarity with the task. Overall, the specificity of the behavioral improvements (i.e., they were limited to the spelling task) increase confidence that any neural changes we observed from pre- to post-treatment were specific to the recovery of spelling functions, rather than to generalized cognitive changes.

#### 3.1.2. In-scanner performance

Participants performed with comparable (and high) accuracy across pre- and post-treatment time-points for both Case-Verification trials and also for Spelling Probe trials with the Known Words (Case-Verification: pre = 0.95, post = 0.95, t-test $t(14) = 0$; Spelling Probe Known Words: pre = 0.79, post = 0.82, $t(14) = 0.78$). The relatively high accuracy indicates that participants were actively engaged during the experiment and the consistency across time-points indicates that neither the spelling treatment nor the repeated scanning produced generalized improvements in scanner performance. Furthermore, consistent with spelling performance outside the scanner, the Training Words Spelling Probe condition showed both lower accuracy than the Known Words condition and accuracy improved significantly from pre-to post-treatment (pre = 0.68, post = 0.76, $t(14) = 2.5$, two-tailed $p = .0255$).

### 3.2. Analysis 2: identifying the functional modular structure of spelling and global and local hubs

Hierarchical clustering analysis of the connectivity matrices of Control Group 1 identified 10 bilateral, generally symmetrical clusters of 13–39 nodes each (Fig. 3a) that served as the reference modular structure for subsequent analyses. In neuroanatomical terms, the 10 clusters can be labeled as follows (see Table 2 for the specific anatomical areas associated with each): (1) prefrontal, (2) ventral frontoparietal, (3) dorsal frontoparietal, (4) medial occipital, (5) occipitoparietal, (6) posterior parietal, (7) temporal, (8) ventral occipitotemporal (VOT), (9) ventromedial prefrontal, and (10) perisylvian. The subset of these clusters that most closely corresponds to the spelling network as identified in neuroimaging meta-analyses with neurologically healthy individuals (Purcell et al., 2011; Planton et al.,

**Table 2**
Information regarding clusters derived from hierarchical clustering analysis, and used as the "reference" modular structure.

| Cluster # | Cluster name | *N* of nodes | Mean % nodes damaged in LH (mean (SD)) | Region labels (HO atlas) |
|---|---|---|---|---|
| 1 | Prefrontal | 36 | 13 (11) | Frontal pole; IFG-opercularis; Middle frontal gyrus; Superior frontal gyrus; ACC/Paracingulate gyrus; Precentral gyrus. |
| 2 | Ventral Frontoparietal | 17 | 17 (22) | SMA; Middle frontal gyrus; Precentral gyrus; Postcentral gyrus; ACC. |
| 3 | Dorsal Frontoparietal | 26 | 7 (14) | Middle frontal gyrus; Superior frontal gyrus; Precentral gyrus; Postcentral gyrus; Superior Parietal Lobule. |
| 4 | Medial Occipital | 15 | 1 (4) | Cuneal cortex; Intracalcarine cortex, SupLOC; InfLOC; Occipital pole; Precuneus. |
| 5 | Occipitoparietal | 13 | 4 (10) | PCC; SupLOC; Precuneus; AG. |
| 6 | Posterior Parietal | 20 | 6 (9) | PCC; SupLOC; Precuneus; Superior parietal lobule; Posterior Supramarginal gyrus; AG. |
| 7 | Temporal | 31 | 10 (12) | ITG; MTG; Temporal fusiform cortex; Lingual gyrus; Parahippocampus; STG-anterior; Temporal pole. |
| 8 | VOT | 22 | 3 (5) | ITG-temporooccipital; MTG-temporooccipital; InfLOC; supLOC; Lingual gyrus; Occipital fusiform gyrus; occipital pole. |
| 9 | Ventromedial Prefrontal | 16 | 1 (4) | Frontal orbital cortex; Frontal pole; Paracingulate gyrus; Frontal medial cortex. |
| 10 | Perisylvian | 39 | 27 (25) | Central opercular cortex; Frontal pole; IFG-triangularis; Insula; Frontal operculum; Parietal operculum; Planum temporale; Postcentral gyrus; Precentral gyrus; MTG. |

HO atlas refers to the Harvard-Oxford Atlas (Desikan et al., 2006.). VOT – ventral occipitotemporal cortex; IFG – inferior frontal gyrus; ACC – anterior cingulate cortex; SMA – supplementary motor cortex; sup/inf LOC – superior/inferior lateral occipital cortices; AG – angular gyrus; PCC – posterior cingulate cortex; STG – superior temporal gyrus; MTG – middle temporal gyrus.

2013) are the left prefrontal, posterior parietal, VOT, and perisylvian clusters.

Table 2 reports, for each cluster, the mean percentage of damaged nodes across participants. In line with the lesion distribution seen in MCA strokes, the perisylvian cluster was the most affected with mean percentage damage of 27% (SD = 25%). The total percentage of damaged nodes (those with > 25% lesioned voxels) per participant ranged from 1% to 12% (mean = 5%, SD = 3%). There was also a high correspondence between lesion volume and the percentage of damaged nodes (Pearson $r = 0.91$).

With regard to the identification of hubs based on Control Group 1 data, 20 global hubs (i.e., inter-module connector nodes) were identified with a bilateral distribution, although they were located only in the occipital and parietal clusters (medial occipital, occipitoparietal, posterior parietal). Local hubs (within-module connectors) were found in all clusters except for the posterior parietal cluster, for a total of 37; but they were primarily distributed in the prefrontal (cluster #1, 7/37) and the perisylvian (cluster #10, 7/37) clusters, followed by the VOT (cluster #8, 6/37), the dorsal frontoparietal (cluster #3, 5/37) and the temporal (cluster #7, 5/37) clusters (Fig. 3b).

### 3.3. Analysis 3: pre-treatment network properties

#### 3.3.1. Do pre-treatment modularity, average PC and/or average WD index deficit severity?

First, since the reference modular structure (Fig. 3a) that forms the basis of all the subsequent analyses was derived from Control Group 1 data, a comparison of Control Group 1 with Control Group 2's *modularity*, average PC and WD values allows for evaluating the reliability of these network properties with an independent dataset. For each group, average PC and WD were calculated as the mean *participation coefficient* and *within-module degree z-score* values for the global and local hubs respectively (identified as such based on Control Group 1 data) as shown in the glass brains in Fig. 3b. There were no significant differences between the two control groups for either *modularity* or average PC (*modularity*: $t(21) = 0.60$, $p = .55$; PC: $t(21) = 0.91$, $p = .37$. Fig. 5a and b). While the two groups were significantly different in terms of average WD ($t(21) = 2.77$, $p = .01$. Fig. 5c), each Control Group's values were very significantly larger than those of the Lesion Group (reported in the next paragraph).

Second, we evaluated whether the Lesion Group's pre-treatment *modularity*, average PC and WD values differed from those of the Control Groups (Fig. 5, comparison indicated with purple). There was no significant difference for *modularity* (two Control Groups combined: $t(36) = -0.21$, $p = .83$; Control Group 1: $t(23) = 0.10$, $p = .92$; Control Group 2: $t(26) = -0.40$, $p = .69$). This is generally consistent with the finding of Siegel et al. (2018) that *modularity* values generally normalize within a few months post-stroke. For the average PC values of global hubs, we also found no difference with the Control Groups (two groups combined: $t(36) = 0.21$, $p = .84$; Control Group 1: $t(23) = 0.59$, $p = .56$; Control Group 2: $t(26) = -0.19$, $p = .85$). In contrast, the Lesion Group's local hubs exhibited average WD scores that were lower when than those of the Control Groups (two groups combined: $t(36) = 5.37$, $p = 5e\text{-}06$; Control Group 1: $t(23) = 5.70$, $p = 8e\text{-}06$; Control Group 2: $t(26) = 3.5$, $p = .0017$).

As there is a close relationship between *modularity* and average PC and WD, we correlated *modularity* with each of the hub-based measures for the individuals in the Lesion Group. As shown in Fig. 6a, there was a significant negative correlation between *modularity* and average PC (Pearson $r = -0.63$, $p = .003$ by 5 K permutation), indicating that lower global *modularity* was related to higher degrees of global integration. On the other hand, no significant correlation was found between *modularity* and average WD ($r = 0.06$, $p = .39$). This finding is consistent with observations from both empirical and simulation studies that damage to global hubs has more widespread consequences than does damage to local hubs (Albert et al., 2000; Albert and Barabási, 2002; Sporns et al., 2007; Honey and Sporns, 2008; Alstott et al., 2009; Gratton et al., 2012; Warren et al., 2014).

Third, with regard to the relationship between pre-treatment *modularity*, average PC and WD values and deficit severity, regression analysis results (Fig. 6b, left to right) indicate that pre-treatment average PC values were significantly positively correlated with deficit severity ($t = 2.97$, $p = .0178$) such that greater spelling severity was associated with higher average PC values (stronger global connectivity). Note that, given that lesion volume was also included as a predictor variable in the regression models (Fig. S3), the relationship between severity and PC was not driven by lesion volume. As *modularity* and average PC were negatively correlated, *modularity* exhibited the reverse relationship with severity (i.e., greater spelling severity was associated with lower modularity) but this did not reach significance ($t = -1.45$, $p = .1856$). Lastly, average WD values did not correlate with severity ($t = 0.031$, $p = .971$).

With regard to the other variables included in the pre-treatment time-point regression analyses: In terms of *modularity,* higher *modularity* values were correlated with greater age ($t = 2.61$, $p = .0309$) and longer time after stroke ($t = 2.59$, $p = .0321$). For average PC, higher values were associated with lower age ($t = -3.26$, $p = .0115$), more years of education ($t = 4.55$, $p = .0019$), and smaller lesions ($t = -3.02$, $p = .0166$). Average WD values were not significantly correlated with any of these variables at pre-treatment. The amount of in-scanner motion was not correlated with *modularity* or average WD ($t = 0.59$ and $-1.56$ respectively) and only marginally correlated with average PC ($t = 2.03$, $p = .07$). (All effects of the three models are shown in Fig. S3).

#### 3.3.2. Are pre-treatment modularity, average PC, and/or WD related to future response to treatment?

To examine whether future response to treatment was related to these network properties at pre-treatment, we calculated the correlation between each of the three graph-theoretic measures at the pre-treatment time-point and the pre-to post-treatment behavioral improvement on the Training Words. As shown in Fig. 7a, we found that higher pre-treatment average WD was significantly associated with larger future treatment gains (Pearson r = 0.51, $p = .02$ by 5 K permutation) while there was no significant association for either *modularity* ($r = 0.22$) or average PC ($r = -0.15$).

### 3.4. Analysis 4: the relationship between pre to post-treatment changes in behavioral and network properties

#### 3.4.1. Do modularity, average PC and/or WD change in response to treatment? If so, what is the relationship among these variables?

The three graph-theoretic measures at the pre- and post-treatment time-points for the individuals in the Lesion Group and the comparisons with the Control Groups are depicted in Fig. 5 (purple and green respectively). The comparison of *modularity* values (Fig. 5a, orange) from pre- to post-treatment for the Lesion Group reveals a significant increase ($t(14) = 3.03$, $p = .0089$) with post-treatment *modularity* exceeding that of the Control Groups (two groups combined: $t(36) = -2.49$, $p = .02$).

As can be seen in Fig. 5, average PC values did not change in response to treatment ($t(14) = -1.05$, $p = .31$). In contrast, average WD values increased significantly from pre-to post-treatment ($t(14) = 2.2$, $p = .04$) but still remained significantly lower than those of the Control Groups at the post-treatment time-point ($t(36) = 4.30$, $p = .0001$).

Given the mathematical relationships between average PC/WD and *modularity*, in order to evaluate whether changes in either one or both contributed to the significant *modularity* changes reported in the previous analysis, we examined the correlation between these measures. Results showed that changes in average PC (global integration) changes were not related to *modularity* changes ($r = -0.17$, $p = .26$ by 5 K permutation test), although there was a significant correlation between

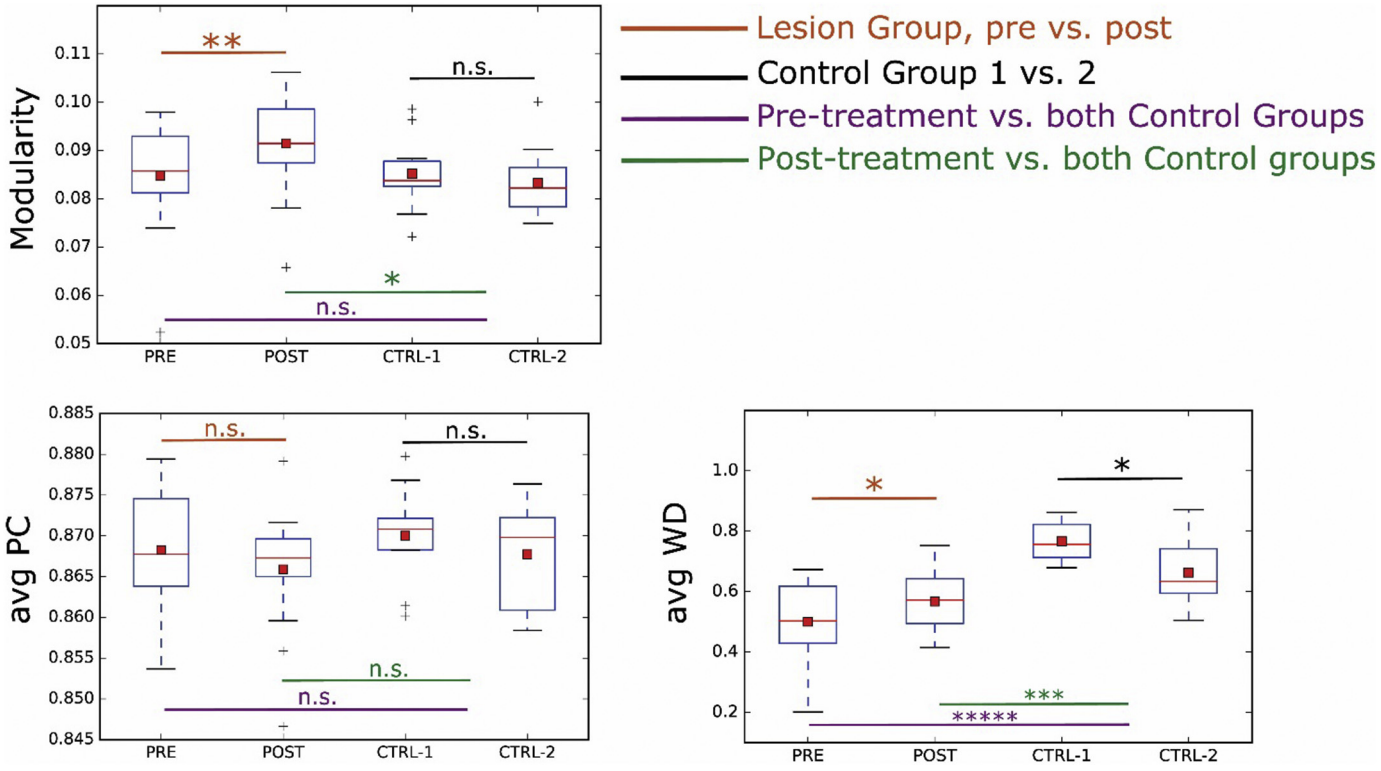


**Fig. 5.** Comparison of network properties across time-points and participant groups. Four comparisons are evaluated for each network property: Control Group 1 vs. 2 (black), pre-treatment Lesion Group vs. all Controls (purple), post-treatment Lesion Group vs. all Controls (green), and pre- vs. post-treatment for the Lesion Group (orange). The pre vs. post comparisons are calculated with paired *t*-tests and the other three with unpaired t-tests, all *p*-values are two-tailed. (a) *Modularity:* values for individuals in the Lesion Group at pre-treatment do not differ from the Control Groups, but increase significantly from pre- to post-treatment and become significantly higher than the Control Groups at post-treatment. (b) average PC: No significant difference between Lesion and Control Groups at either time-point, and within the Lesion Group the values do not differ from pre- to post-treatment. (c) average WD: Significantly lower in individuals in the Lesion Group than for those in the Control Groups at both time-points, and increase significantly from pre- to post-treatment. There is also a significant difference in average WD between the two Control Groups. (~: $p < .1$, *:$p < .05$, **: $p < .01$, ***: $p < .001$, ****: $p < .0001$, *****: $p < .00001$, n.s.: $p > .1$).

average PC and *modularity* at pre-treatment as we reported in Section 3.3.1 (Fig. 6a). In contrast, we find that average WD (local integration) changes were significantly correlated with *modularity* changes (Pearson $r = 0.62$, $p = .01$ by 5 K permutation test. Fig. 8a). These results indicate that, although the degree of global integration plays a key role in determining *modularity* (a global network measure) the changes in *modularity* that we observed to be associated with treatment were, at least in large part, driven by increases in local connectivity.

Given that head motion can have an impact on patterns of functional connectivity (e.g., Power et al., 2012; Satterthwaite et al., 2013), we compared the amount of motion across groups and time-points (mean RMS calculated by MCFLIRT, Jenkinson et al., 2002). Results revealed no significant difference between pre- and post-treatment time-points ($t(14) = 0.82$, $p = .44$), no difference between the two Control Groups ($t(21) = 0.51$, $p = .62$), nor between the Lesion Group at the pre-treatment time-point and the Control Groups (two groups combined: $t(36) = 1.61$, $p = .12$). Only at the post time-point was the Lesion Group's motion greater than that of the Controls Groups' (two groups combined: $t(36) = 2.25$, $p = .03$). Importantly, however, the fact that there was no difference between pre- and post-treatment time-points for the Lesion Group indicates that the observed changes in *modularity* cannot be attributed to motion changes. In addition, we also evaluated the relationship of inter-node distance and the correlation between motion and connectivity strength following Satterthwaite et al. (2013), as it has been suggested that motion tends to increase short-distance connectivity values and decrease long-distance connectivity values. However, we found no such relationship (Pearson $r = -0.11$ and $-0.13$ for pre- and post-time-points respectively), indicating that the observed changes in modularity cannot be explained by a distance-dependent effect of motion.

### 3.4.2. How are pre- to post-treatment changes in network properties related to behavioral changes?

Regression analyses revealed that *modularity* increases were significantly and negatively correlated with behavioral treatment gains such that participants with smaller *modularity* increases (or decrease) had larger behavioral improvements ($t = -2.77$, $p = .0396$, Fig. 8b). Thus, despite the fact that overall there was a significant pre- to post-treatment increase in *modularity* such that *modularity* was actually significantly higher than normal at post-treatment (Fig. 6a), larger increase in *modularity* were generally associated with smaller behavioral improvements. Why might this be the case? One obvious possibility is that the negative relationship between modularity increases and treatment gains occurs because the higher *modularity* observed after treatment was maladaptive. That interpretation predicts that higher post-treatment *modularity* should have been associated with lower post-treatment spelling accuracy. However, we found no such relationship (Pearson $r = -0.17$), making such an explanation very unlikely. The alternative interpretation we propose is that individuals that started out with a healthier, more intact spelling system required less neural modification to achieve a larger benefit from treatment. This not only explains the seemingly paradoxical observation that smaller increases in *modularity* were associated with greater behavioral treatment gains, but is consistent with two additional observations. Namely, that those individuals who experienced smaller *modularity* increases from pre- to post-treatment: (a) had less severe deficits at pre-treatment ($t = 4.98$, $p = .0042$, Fig. S4) and (b) also had higher pre-treatment *modularity* (Pearson $r = -0.44$, $p < .05$ by 5 K permutation test). Thus, these

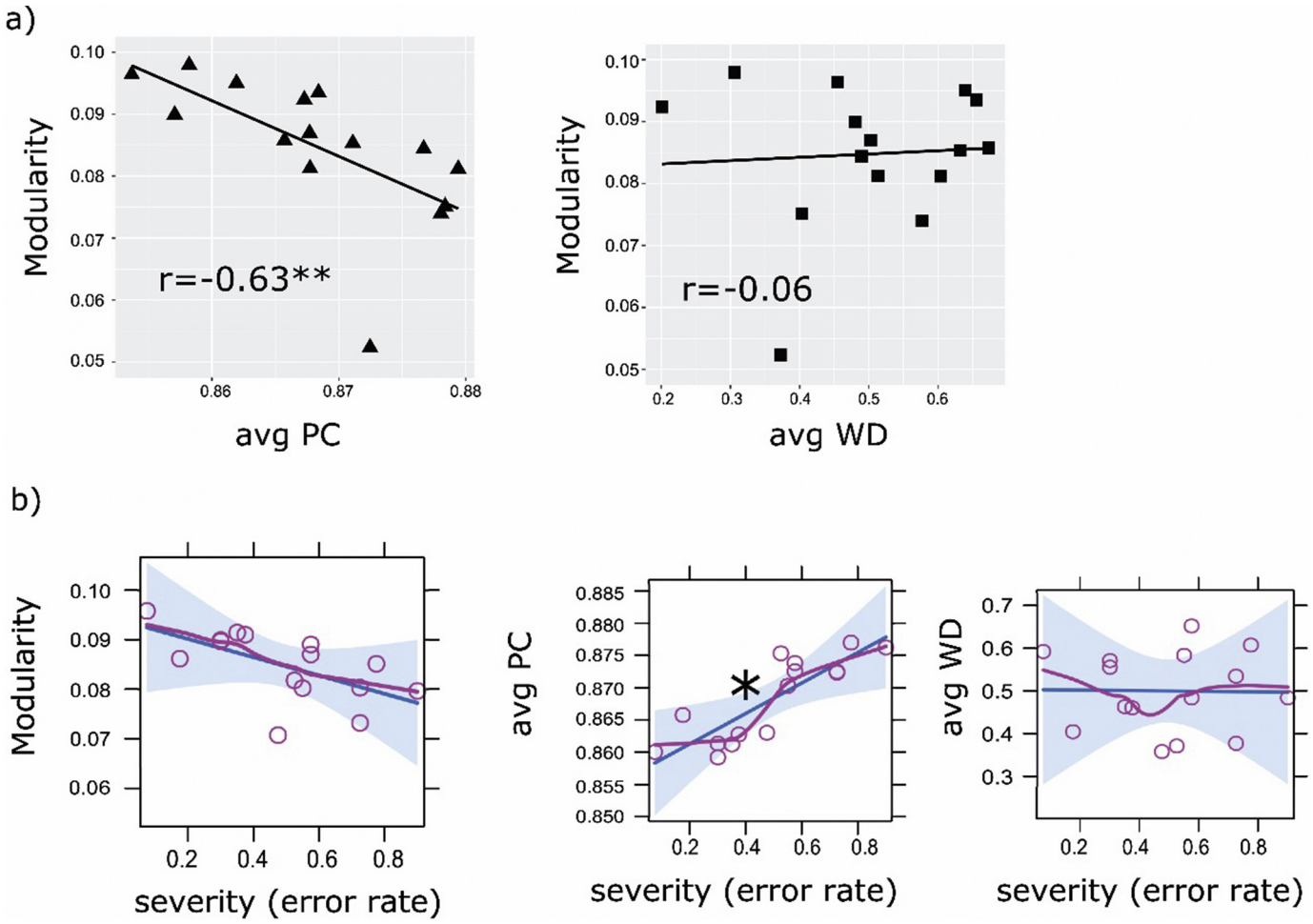


**Fig. 6.** (a) Relationship between *modularity* and average PC and WD in individuals with lesion at pre-treatment time-point. *Modularity* is significantly correlated with average PC such that lower *modularity* is associated with higher average PC values of the global hubs, i.e. higher global integration. No significant correlation is observed between *modularity* and average WD. (b) Relationships between pre-treatment network properties (from left to right: *modularity*, average PC, average WD) and deficit severity as estimated from multiple regression models (described in Section 2.5.7). The plots visualize, for each model, the estimated slopes (and confidence intervals) combined with partial residual plots of the predictor *severity* (R package **effects**, Fox, 2003). The results indicate that greater spelling severity is significantly associated with higher PC before treatment. Effects of all the other predictors can be found in Fig. S3. (∼: $p < .1$, *: $p < .05$, **: $p < .01$, ***: $p < .001$, ****: $p < .0001$, *****: $p < .00001$, n.s.: $p > .1$).

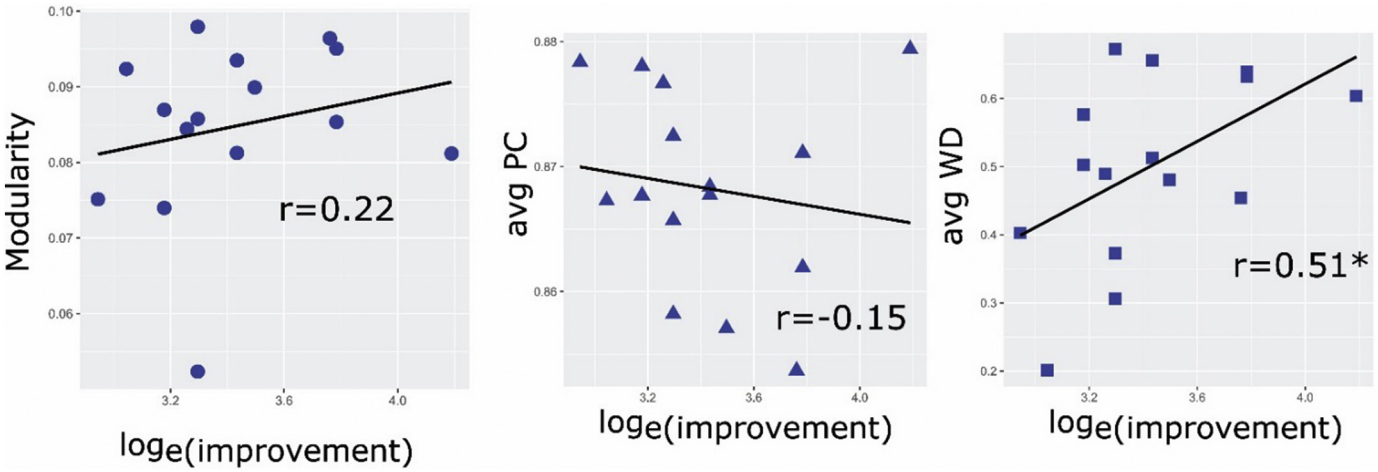


**Fig. 7.** Relationship between pre-treatment network properties and behavior changes from pre- to post-treatment, left to right are: *modularity,* average PC and average WD. Higher average WD values before treatment is correlated with the amount of improvement in spelling (Pearson $r = 0.51$, $p = .02$ by 5 K permutation test).

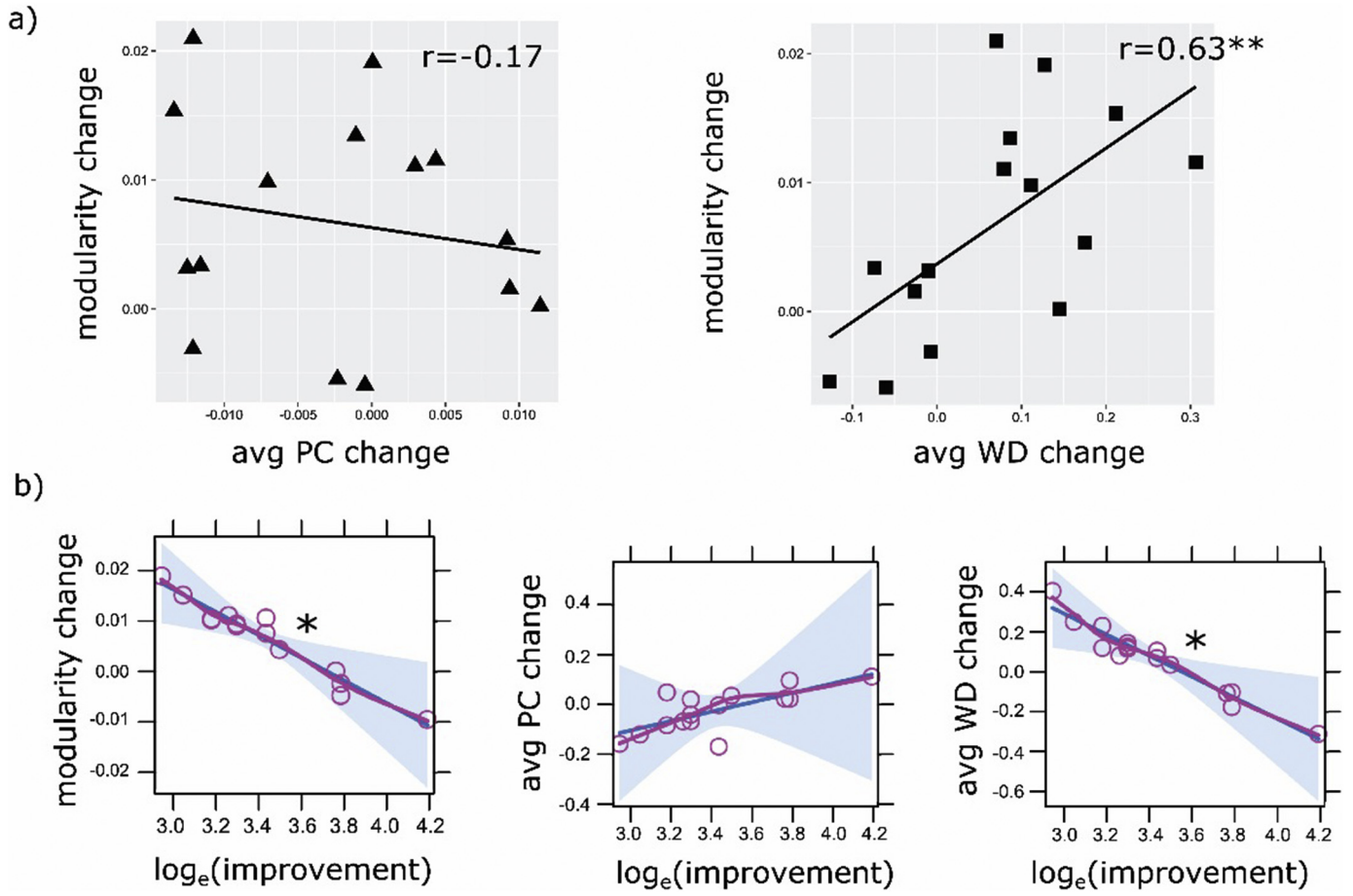


**Fig. 8.** (a) Relationship between changes in modularity and average PC and WD in individuals with lesion. No significant correlation is found between modularity and average PC changes, but changes in modularity are significantly correlated with changes in average WD such that greater modularity increases are associated with greater average WD increases. (b) Relationships between network property changes (left to right: modularity, average PC and WD) and behavioral changes from pre- to post-treatment as estimated from multiple regression models (described in Section 2.5.7). The plots visualize, for each model, the estimated slopes (and confidence intervals) combined with partial residual plots of the predictor improvement after log transform (R package **effects**, Fox, 2003). The amount of pre- to post-treatment improvement in spelling is significantly associated with pre- to post-treatment change in modularity and marginally with change in average WD, while for average PC the correlation is not significant. The correlations are negative, indicating that greater behavioral improvement is associated with smaller increases (or decreases) in modularity and average WD (see text for further discussion). Effects of all the other predictors can be found in Fig. S4. (~: $p < .1$, *:$p < .05$, **: $p < .01$, ***: $p < .001$, ****: $p < .0001$, *****: $p < .00001$, n.s.: $p > .1$).

observations all support the interpretation that less severe deficits prior to treatment, presumably due to a healthier and more modular spelling system, are characterized by higher pre-treatment *modularity* which requires less modification to achieve greater behavioral benefits.

There was no statistically significant relationship between changes in average PC and behavioral changes ($t = 0.88$, $p = .42$, Fig. 7b), but we did find a significant correlation between changes in average WD and behavioral changes such that, as we observed for *modularity,* smaller average WD increases (or decreases) were associated with greater treatment gains ($t = -3.37$, $p = .0199$, Fig. 7b). Like *modularity*, WD changes showed trend of positive correlation with severity such that smaller WD increases tended to have less severe deficits at pre-treatment ($t = 1.82$, $p = .1279$, Fig. S4), and smaller WD increases also correlated with higher WD values at pre-treatment (Pearson $r = -0.74$, $p < .001$ by 5 K permutation test). In other words, changes in average WD and *modularity* show the same relationship with behavioral gains, while average PC changes seem unrelated.

With regard to the other variables included in the pre-post treatment regression analyses, only age was significantly correlated with magnitude of change in *modularity* ($t = -3.44$, $p = .0184$), such that greater age was associated smaller *modularity* increases (or decrease). No other significant effects were observed. Results of all the independent variables of the regression models are shown in Fig. S4.

### *3.4.3. How are treatment-related network changes distributed throughout the brain?*

We examined if the significant average WD changes reported in Section 3.4.2 were evenly distributed across clusters or, instead, were concentrated in a subset of them. To do so, for the 9 clusters with local hubs (cluster #6 lacked local hubs), we evaluated the WD changes of each cluster from pre- to post-treatment with two-tailed paired t-tests. As shown in Fig. 9, the increases in average WD were not equally distributed across clusters but instead were concentrated in a few clusters: the VOT cluster (#8) had the largest increase (Fig. 8, dark red, $t(14) = 3$, $p = .0096$); the dorsal frontoparietal (#3) and the ventromedial prefrontal clusters (#9) also had marginal increases ($t(14) = 1.82$, $p = .0898$ & $1.96$, $p = .0706$ respectively). Only the VOT cluster remained marginally significant after multiple comparisons correction (Bonferroni corrected $p = .0864$). We also evaluated the relationship between behavioral recovery and WD changes for each cluster. However, none of the individual clusters (including the VOT cluster) showed statistically significant correlations between the WD changes and behavioral improvement ($p$-values ranged from 0.18 to 1). Nonetheless, all of the clusters that showed numerical increases in WD from pre to post treatment (i.e., clusters 3, 4, 6, 8, 9, 10) exhibited the same pattern (that was found to be statistically significant across the clusters) such that smaller WD increases or decreases were associated

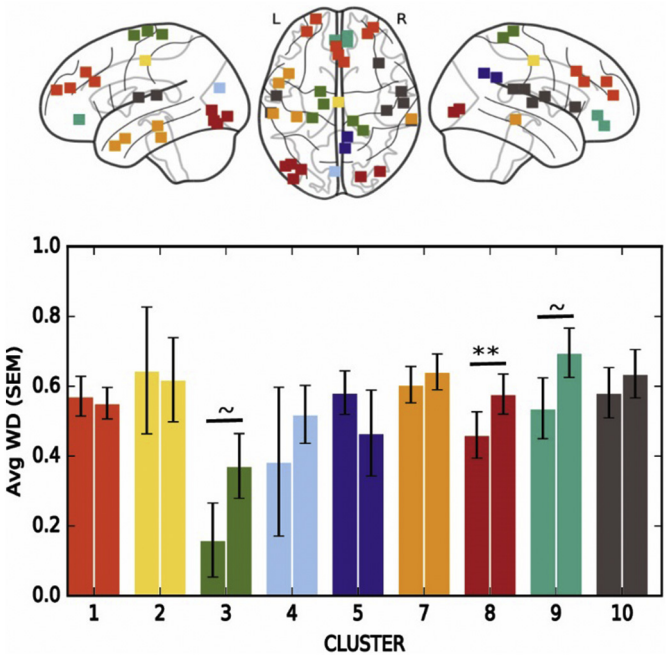


**Fig. 9.** Distribution of pre- to post-treatment changes of average WD across the 9 clusters. The average WD values of the local hubs of each cluster at pre- and post-treatment are depicted (on the x-axis are the cluster numbers as reported in Table 2). The uncorrected *p*-values of two-tailed paired *t*-tests are indicated. Among the 9 clusters with local hubs (the posterior parietal cluster #6 lacked local hubs), only the VOT cluster (dark red) shows a statistically significant pre- to post-treatment increase, with marginally significant changes in the dorsal frontoparietal cluster (green) and the ventromedial cluster (turquoise). $^{**}$:$p < .01$; $\sim p < .1$.

with larger behavioral gains.

As the VOT cluster showed the largest change and it also included the *visual word form area* (VWFA), we visualized the connectivity patterns and the WD values of the local hubs for each group and time-point in Fig. 10a to provide a concrete picture of the pre to post connectivity differences within this region. As clearly shown in Fig. 10a, the VOT connectivity pattern of the Lesion Group normalized towards the Control Groups from pre-to post-treatment. Furthermore, we compared the average WD scores of both time-points to the healthy controls (Fig. 10b). Although this cluster was well preserved in the individuals in the Lesion Group (Table 2), the average WD values at pre-treatment were lower than for the Control Group (t(36) = 3.38, $p = .0018$) and, although they significantly increased, they remained (marginally) lower than Controls' at post-treatment (t(36) = 1.92, $p = .0634$).

### 3.5. Analysis 5: hemispheric effects

The analyses and results reported thus far are based on bilateral clusters of nodes. One natural question is whether results observed from these bilateral clusters could have been driven by connectivity in a single hemisphere. To evaluate this possibility, we examined the effects reported in Sections 3.3. and 3.4 separately for the nodes in the left (ipsilesional) and the right (contralesional) hemispheres.

#### 3.5.1. Hemisphere-specific modularity, average PC and WD at pre-treatment

For *modularity*, we found that at pre-treatment, individuals in the Lesion Group showed higher LH *modularity* than the Control Group (t(36) = −2.13, $p = .0405$), but did not differ from the controls in the RH (t(36) = −0.10, $p = .3253$), with the RH results mirroring those reported in the whole-brain analysis. Regarding average PC, as with the whole-brain analysis, neither LH nor RH differed from the controls (LH: t(36) = −0.37, $p = .7119$, RH: t(36) = 0.81, $p = .4241$). Analysis of hemisphere-specific average WD showed effects similar to those found in the whole-brain analysis such that individuals in the Lesion Group had significantly lower average WD values than the healthy controls in both hemispheres (LH: t(36) = 3.94, $p = .0004$, RH: t(36) = 3.51, $p = .0012$). Those comparisons are shown in Fig. S5 (in purple). In addition, in direct comparisons of the hemispheres, there were no significant differences between the two hemispheres for any group at any time-point. In sum, at the pre-treatment time-point, the hemisphere-specific effects were the same as the whole-brain results except that *modularity* in the ipsilesional LH was higher for the Lesion Group than for the Control Group. In addition, we found that LH *modularity* was positively correlated with lesion volume (Pearson $r = 0.46$, $p = .04$) but that this was not the case for RH *modularity* ($r = 0.10$, n.s.).

#### 3.5.2. Hemisphere-specific pre- to post-treatment changes in modularity, average PC and WD

With regard to *modularity,* each hemisphere showed a similar pattern of results as was reported for the whole-brain, with *modularity* values significantly increasing in each hemisphere from pre-to post-treatment (LH t(14) = 2.48, $p = .0267$; RH: t(14) = 2.77, $p = .0150$) and no interaction between hemisphere and time-point (F(1,1) = 0.0109, $p = .9182$, Fig. S5). Moreover, as with the whole-brain, both LH and RH showed higher *modularity* than the controls after treatment (LH: t(36) = −4.19, $p = .0002$; RH: t(36) = −3.27, $p = .0024$). For average PC, also consistent with the whole-brain results, neither hemisphere showed a difference from pre- to post-treatment (LH t(14) = −1.34, $p = .2024$; RH: t(14) = −0.56, $p = .5857$), nor did the post-treatment values differ from those of the controls (LH: t(36) = 0.89, $p = .38$; RH: t(36) = 1.54, $p = .13$). In terms of average WD, we observed a marginally significant increase from pre- to post-treatment in the RH (t(14) = 1.96, $p = .0698$), while the increase in the LH was not significant (t(14) = 1.46, $p = .1671$), and there was no interaction between hemisphere and time-point (F(1,1) = 0.19, $p = .6712$). The post-treatment WD values in both hemispheres were still lower than the controls, as was seen for the whole-brain analysis (LH: t(36) = 3.67, $p = .0008$, RH: t(36) = 2.05, $p = .0480$).

Within-hemisphere regression analyses also revealed similar relations between *modularity* changes and behavioral changes as were seen for the whole brain such that smaller *modularity* increases (or decrease) in each hemisphere were associated with larger pre- to post-treatment behavioral improvements (LH $t = -2.17$, $p = .0818$; RH $t = -2.05$, $p = .0955$). Likewise, there was also a positive relationship between *modularity* increases and pre-treatment spelling accuracy (severity) in each hemisphere (LH: $t = 2.50$, $p = .0548$; RH: $t = 2.54$, $p = .0520$) such that participants with larger *modularity* increases tended to have more severe deficits, and had smaller treatment gains. As in the whole-brain analysis, only age was significantly correlated with magnitude of change in *modularity* in LH ($t = -3.05$, $p = .0285$), such that greater age was associated smaller *modularity* increases (or decrease). No other significant effects were observed.

To summarize, the pre- to post-treatment effects found for each hemisphere were by and large very similar to the whole-brain results. Together with the absence of any statistically significant interactions between hemispheres, we conclude that the significant effects observed with whole-brain connectivity were not clearly driven by a single hemisphere.

#### 3.5.3. Hemisphere-specific effects in the VOT cluster

Finally, we examined the hemisphere-specific effects for the VOT cluster which was the cluster that showed the greatest pre to post-treatment WD increase (Section 3.4.3). There were 6 local hubs in the VOT (black nodes in Fig. 9), of which 4 were in the left and 2 were in the right hemisphere. In the LH, (see Fig. 10b) the average WD values calculated with the 4 LH hubs showed that, as for the bilateral cluster, WD increased, albeit marginally, from pre- to post-treatment ($t = 2.04$,

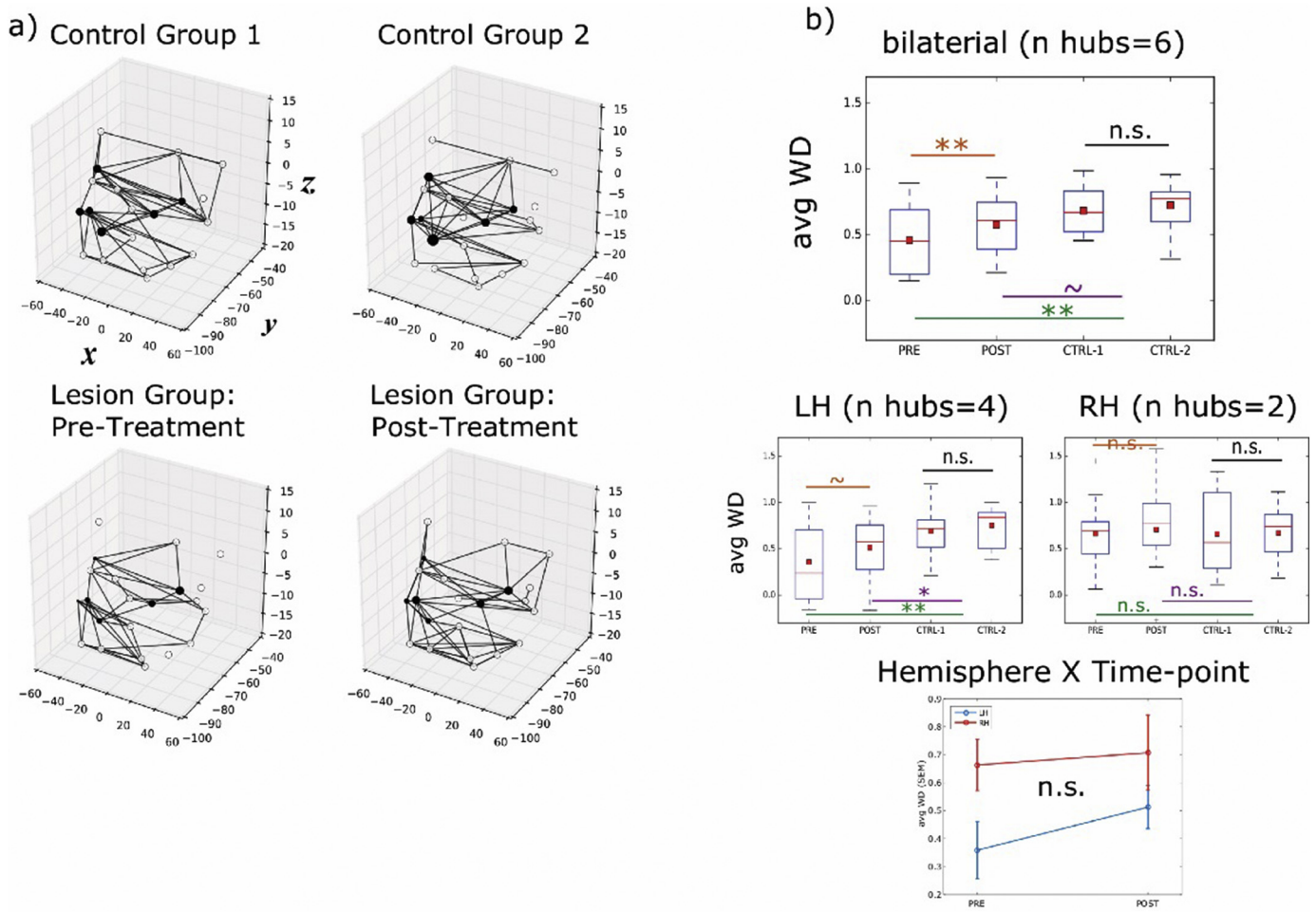


**Fig. 10.** (a) VOT cluster connectivity for groups and time-points. Nodes and connections of the VOT cluster are plotted in 3D space with x, y, z axes corresponding to MNI coordinates. Connections that are present in all participants of that group/time-point are shown. Filled circles represent local hubs (identified based on *within-module degree z-score* (WD) measures from Control Group 1) and the circle size is proportional to the WD value calculated with each group. (b) Comparisons of average WD values across groups and time-points. The comparisons are color-coded as in Fig. 5. The average WD values depicted in the boxplots correspond to the average WD values of the 6 hub nodes. The average WD values (i.e., local integration) of the VOT cluster are similar across the two Control Groups and normalizes in the Lesion Group from pre-to post-treatment. Results of the separate analysis of the left and right hemisphere local hubs are also shown. Although there are some significant LH effects, there is no significant interaction between time-point and hemisphere.

$p = .0612$), and average WD values at both time-points remained significantly lower than for the Control Group (pre: $t(36) = 3.54$, $p = .0011$, post: $t(36) = 2.40$, $p = .0216$). For the RH, we did not observe significant pre-post changes in average WD ($t(14) = 0.43$, $p = .6712$), and the individuals in the Lesion Group were not different from those in the Control Group (pre: $t(36) = 0.03$, $p = .9748$, post: $t(36) = -0.28$, $p = .7846$). However, we did not find a statistically significant interaction between hemisphere and time-point in the individuals in the Lesion Group ($F(1,1) = 0.47$, $p = .5045$). In sum, although there was a trend towards left-lateralization of the WD treatment-related changes in the VOT cluster, the non-significant interaction between the effects in the two hemispheres does not allow for strong conclusions in this regard.

## 4. General discussion

In this study, we investigated the properties of functional networks in chronic stroke-induced language deficits, specifically evaluating the graph-theoretic measures of *modularity* as well as global and local connectivity (*participation coefficient* (PC) and *within-module degree z-score* (WD), respectively). Fifteen individuals with chronic, acquired written language deficits (dysgraphia) subsequent to left-hemisphere stroke completed an intensive behavioral treatment program, with task-based functional MRI data collected before and after the treatment. Using task-based functional connectivity (Norman-Haignere, et al., 2012), we identified a reference modular structure as well as local and global hubs from age-matched healthy controls. On this basis, we investigated the three network properties both before and after treatment, comparing the Lesion and Control Groups and the relationship between the network properties and behavior. We found: (1) at the pre-treatment time-point, network properties (PC and, to a lesser extent, *modularity*) indexed deficit severity, such that higher PC and lower *modularity* were associated with more severe deficits; (2) at the pre-treatment time-point, WD was significantly lower in the individuals with lesions than in the healthy controls and higher WD values were associated with better future responsiveness to treatment; (3) there were significant increases in *modularity* and WD from pre- to post-treatment; (4) the magnitude of the *modularity* and WD increases were (negatively) correlated with treatment gains (as measured outside the scanner); (5) cluster-based analyses showed that the observed *modularity* increases were driven by enhanced WD, especially within the ventral occipitotemporal cortex (VOTC). Overall, the investigation revealed that more modular networks with higher local integration were associated with lower deficit severity and greater response to treatment; after treatment, both global network *modularity* and local integration increased, especially within intact ventral occipital-temporal regions of

the spelling network.

### 4.1. The impact of lesions on network properties and task performance

Previous work has shown that *modularity* values are associated with various aspects of cognitive performance in healthy populations such as motor skill learning (Bassett et al., 2011, 2013), working memory load and/or task demands (Kitzbichler et al., 2011; Braun et al., 2015a; Yue et al., 2017), cognitive training (Gallen et al., 2016), as well as normal aging (Meunier et al., 2009a; Betzel et al., 2014; Chan et al., 2014). Abnormal modular organization has also been found in populations with various neurological disorders, including Alzheimer's Disease (Stam et al., 2006; Buckner et al., 2009; Brier et al., 2014), schizophrenia (Bassett et al., 2008; Lynall et al., 2010; Yang et al., 2016), traumatic brain injury and strokes (Gratton et al., 2012; Arnemann et al., 2015; Caeyenberghs et al., 2017; Siegel et al., 2018). Although higher *modularity* values are typically associated with better performance (e.g., Gratton et al., 2012; Siegel et al., 2018), this is not always the case and one might well imagine that higher *modularity* might be beneficial for certain tasks but not others, as different tasks should benefit differently from local and global integration. In fact, in neurotypical adults Yue et al. (2017) reported that higher *modularity* values were associated with stronger performance on less complex tasks, while lower *modularity* (greater global interactivity) was associated with stronger performance on more complex tasks.

#### 4.1.1. Before treatment, modularity and global integration were comparable to heathy controls but local integration was below normal

First, we found that in terms of overall *modularity* and global integration values (mean PC of the global hubs), at pre-treatment the individuals with lesions (as a group) did not differ from the control participants. The generally normal levels of *modularity* for the Lesion Group are consistent with the findings of Siegel et al. (2018), who found that whereas *modularity* values were below normal shortly after a stroke, they had normalized by 3 months post-stroke.

Second, before treatment, the individuals in the Lesion Group exhibited significantly lower average *local integration* values (mean WD of the local hubs) than the control participants, indicating weaker within-module connectivity. Although comparisons with healthy controls specifically regarding PC and WD values have not been previously reported, the lower WD values we observed are generally consistent with the Siegel et al. (2016) finding of lower than normal segregation between specific network pairs within the ipsilesional hemisphere.

#### 4.1.2. Before treatment, greater modularity and local integration were associated with better spelling performance and future responsiveness to treatment

As depicted in Fig. 6a, for the individuals in the Lesion Group, lower global connectivity (hence higher global segregation) was associated with less severe deficits (better spelling performance). This indicates that, generally speaking, a more modular organization reflects a more functionally intact spelling network. It is worth noting, however, that in the literature higher *modularity* is not always associated with lower deficit severity. For example, higher than normal *modularity* has been found in Parkinson's disease and multiple sclerosis and higher *modularity* was related to worse behavioral performance (Baggio et al., 2014; Gamboa et al., 2014). Given the differing demands of different tasks and the damage etiologies, it is not surprising to find a mix of findings in this regard. In fact, this variability offers an opportunity to better understand the surely complex relationship between global network properties, task demands and different types of network disruption.

Pre-treatment average WD (local integration) values were significantly and positively correlated with the future behavioral responsiveness to treatment (Fig. 7), consistent with the general notion that greater pre-treatment modularity/segregation reflects greater network health, at least in the context of spelling and dysgraphia. Although the relationships of *modularity* and average PC with future treatment response were not statistically significant, they were in the directions that would have been predicted based on this general notion. These results are consistent with, and extend to include average WD, the findings of Arnemann et al. (2015) and Gallen et al. (2016) that baseline *modularity* levels predicted subsequent learning in brain injured and healthy older individuals, respectively. The focus of this study was on understanding the properties of neural networks and how these are changed by lesion and recovery rather than on more broadly identifying biomarkers of responsiveness to treatment. The latter would represent a different research focus and, to be effective, would entail comprehensive comparison of the predictive properties of network properties with those of other demographic, behavioral, and neural measures. We do note, however, that while average WD at pre-treatment was significantly correlated with treatment gains, lesion size was not.

### 4.2. Changes in network properties that support recovery of function

We found that both *modularity* and average WD significantly increased from before to after treatment while average PC remained stable. Given that average WD values were significantly below normal before treatment, it may have been expected that an increase in local integration would play a role in recovery of function. Despite this increase, after treatment, average WD values still remained significantly below normal, while *modularity* values increased to levels significantly above normal. Importantly, the significant correlation reported between changes in average WD values and *modularity* values indicates that the increases in local integration were driving the observed increases in overall *modularity*. Both Siegel et al. (2018) and Duncan and Small (2016), in the context of spontaneous and treatment-based recovery, respectively, also reported that increases in *modularity* values were associated with improvement in performance. Our work extends these previous results by specifically identifying the role that increases in local integration play in the overall *modularity* increases.

In an examination of individual differences, unlike Duncan and Small (2016) and Siegel et al. (2018), we found that the amount of behavioral improvement was negatively correlated with *modularity* change, such that greater improvement was associated with smaller *modularity* increases (or even decreases in a few individuals). Note that this was the case despite the fact that pre-treatment accuracy on the Training items (as well as overall dysgraphia severity) was included in the regression model. Given the negative direction of the relationship and the fact that *modularity* values were higher than normal after treatment, we considered whether increased *modularity* after treatment was maladaptive or beneficial. If high *modularity* after treatment was maladaptive (e.g., modules might have become overly segregated) then higher post-treatment *modularity* values should have been associated with lower post-treatment spelling accuracy. However, we found no such relationship. Instead, the evidence supports the alternative interpretation that high *modularity* after treatment is generally beneficial but that the observed negative relationship results from the fact that individuals with high *modularity* at *pre*-treatment required less neural change to obtain greater behavioral benefits. This interpretation is consistent with the further finding that higher pre-treatment *modularity* values were associated with less severe deficits and smaller *modularity* increases.

While we think this is a reasonable interpretation that is consistent with the observations, we also acknowledge that the finding of the negative relationship between *modularity* (and WD) increases and behavioral gains may well indicate a level of complexity that we do not yet fully understand. It is important to clarify, however, that it is not the case that it is only participants whose *modularity* values actually decreased who made treatment gains. All participants made significant treatment gains and only 3/15 individuals exhibited decreases in *modularity* values (and 6/15 exhibited decreases in WD, see Fig. 8a).

Therefore, *modularity* (and WD) increases are generally associated with treatment gains, however the magnitude of the neural increases is negatively associated with treatment gains. The explanation we offered earlier provides an account of these general findings, but there are other possibilities such as sub-groups of subjects, non-linear relationships among network values, initial deficit levels, and recovery-associated changes. Examining these types of possibilities will require more complex analysis models that in turn will require a greater number of (carefully studied) participants.

In sum, we found that spelling treatment resulted in improvement of spelling accuracy for both trained and untrained words (as measured outside the scanner) and that this was accompanied by increases in network *modularity* which were largely driven by increased integration within local networks (WD). Individuals with more modular, globally segregated systems prior to treatment were able to benefit more from treatment and required less neural change to obtain greater behavioral benefits. It is worth noting that, consistent with our findings, Purcell, Wiley and Rapp (under review) also found, despite evaluating entirely different neural measures, that higher pre-treatment system integrity required less neural change to achieve greater treatment response.

### 4.3. Treatment-induced changes in ventral occipital-temporal cortex (VOTC)

While the analyses focused primarily on quantifying aspects of overall brain network organization (*modularity*, average PC and average WD), for WD we were able to consider if the observed changes in local integration were distributed uniformly across the 10 modules (i.e., clusters) or if they were concentrated in certain modules more than others. We found (Fig. 9) that the changes in average WD from pre- to post-treatment were not evenly distributed across the modules. Instead some modules showed virtually no change, others exhibited trending ($p < .1$) changes (dorsal frontal parietal and ventromedial prefrontal clusters), but only the ventral occipito-temporal (VOT) cluster exhibited a marginally significant (after multiple comparisons correction) change in average WD values from before to after treatment. Fig. 10 depicts the changes in the local connectivity from before to after treatment within this VOT module. It illustrates that not only does the density of local connectivity significantly increase with recovery, it clearly moves towards normalization. This specific finding is interesting for a number of reasons.

First, the ventral occipital-temporal region includes the left-hemisphere area often referred to as the VWFA (visual word form area). This area has been shown to form a part of the orthographic network in the healthy brain, playing an important role in both reading (e.g., Cohen et al., 2002; McCandliss et al., 2003) and spelling (Purcell et al., 2011; Planton et al., 2013). Furthermore, in terms of its specific role in spelling, it is thought to play a key role in orthographic long-term memory representation (the long-term storage of our knowledge of word spellings), used for both reading and spelling (Rapp and Dufor, 2011; Rapp et al., 2016). On that basis, we would suggest that in the group of individuals that participated in this study, the treatment may have involved reinstating or strengthening knowledge and/or retrieval processes for word spellings, within this region.

Second, although the treatment-induced changes were not entirely limited to this single region, it is nonetheless the case that they were not similarly distributed across all clusters. This indicates that although one may evaluate global network characteristics, one should not lose sight of the possibility that relevant network characteristics or changes might actually be fairly localized. Localized changes will impact global measures and it is important to understand when and under what circumstances these global values reflect more or less localized network characteristics (also see Gallen et al., 2015).

### 4.4. Absence of robust hemisphere-specific effects

Regarding the neurotopography of the brain changes that support recovery of function, previous studies have variously attributed positive recovery outcomes to: (1) ipsilesional-perilesional reorganization (e.g., Winhuisen et al., 2007; Fridriksson, 2010, 2012; Postman-Caucheteux et al., 2010); (2) modulation of contralesional areas homologous to the stroke (e.g., Thulborn et al., 1999; Gold and Kertesz, 2000; Blasi et al., 2002; Turkeltaub et al., 2012); and (3) reorganization of both left perilesional and homologous contralesional areas (Kuest and Karbe, 2002; Crosson et al., 2005; Fridriksson et al., 2006). To examine these possibilities in the context of our study, we repeated all of the network-based analyses on hemisphere-specific nodes and connections.

The results indicated only three areas in which there was some indication of differences between the hemispheres: (1) at pre-treatment, left hemisphere *modularity* values were significantly higher for the Lesion compared to the Control Group, while right hemisphere ones did not differ across groups (Fig. S5); (2) at pre-treatment, left-hemisphere *modularity* values were significantly correlated with lesion volume, while in the right hemisphere they were not, (3) the left hemisphere VOT local hubs showed a marginally significant ($p < .06$) increase in average WD from pre-to post-treatment, while the right hemisphere ones did not (Fig. 10). However, none of these effects showed a statistically significant interaction with hemisphere. Thus, while the results suggest some left-lateralization of network effects, they are not sufficiently robust to support strong conclusions. This is likely to be another area where future work, with larger numbers of participants will be needed to more decisively clarify the issues.

### 4.5. Task-based vs. resting-state networks

One of the main differences between this investigation and the majority of previous functional connectivity studies, is that we examined network structure as defined by background connectivity – the residual time-course of the GLM from task-based fMRI (Norman-Haignere et al., 2012; Al-Aidroos et al., 2012), using a spelling task carried out during scanning. The relationship between resting-state functional connectivity (RSFC) and the functional connectivity derived from the task-based residual time-course has been previously examined (Fair et al., 2007; Cole et al., 2014). Fair et al. (2007) compared RSFC with task-based background connectivity and concluded that the task-based background connectivity was qualitatively similar to RSFC but also had several significant differences. Likewise, Cole et al. (2014) compared RSFC with task-based background connectivity derived from a large set of tasks, and concluded that task-based functional connectivity included both the intrinsic neural activity fluctuations of RS-fMRI along with subtle, yet significant, task-related components.

Our findings are consistent with this interpretation in that we found certain important differences between the parcellation of nodes we identified (from healthy control task-based functional connectivity patterns) and typical RS parcellations. For example, RS parcellations typically yield a single frontal-parietal cluster (or module), that is assumed to participate in the default-mode network or the frontoparietal network. However, the parcellation obtained from the task-based (spelling) functional connectivity in this investigation assigns these areas to distinct clusters. This type of separation between frontal and parietal regions might be due to the specific task involved in the current study, as the frontoparietal network has been argued to be a “flexible hub” that displays increased connectivity to other specific networks during task performance (e.g., Bassett et al., 2011, 2013; Cole et al., 2013; Yue et al., 2017). For instance, Cole et al. (2013) showed that the frontoparietal network displayed the greatest variability in its connectivity pattern to other cortical networks across a large number of tasks. Overall, the separation between the frontal and parietal areas we observed is in line with the claim of Fair et al. (2007) and Cole et al. (2014) that measures such as background connectivity are sensitive to

task-based functional networks.

Relatedly, the global hubs identified in this study were concentrated in bilateral posterior parietal and occipital areas, including the precuneus. While these locations are typically associated with hubs in RS functional connectivity studies (e.g., Hagmann et al., 2008; Buckner et al., 2009; Meunier et al., 2009b; Bertolero et al., 2015), we did not see the frontal hubs that have been reported in many of these studies. Presumably, using a measure such as background connectivity that captures task-based properties, we would expect that different hubs would be identified from different task-based data sets.

Overall, the fact that several of the findings we report are consistent with those reported in resting-state based studies of brain damage and recovery increases confidence that task-based background connectivity is an appropriate and useful measure for investigating network properties in these contexts. Moving forward, it will be important and interesting to identify the extent to which resting state and task-based approaches provide similar or different understandings of these complex systems, and how these relate to the recovery of different cognitive functions.

### 4.6. Limitations

There are various limitations of this investigation. First, although all the participants had a single left-hemisphere stroke, there was nonetheless considerable heterogeneity in the lesion distribution. Lesion heterogeneity poses a challenge to studying neural mechanisms of recovery of function. For example, perilesional tissue is often argued to play an important role in recovery but this can be difficult to assess in a set of heterogeneous lesions. Second, there is also the issue of deficit heterogeneity. Although the fact that all participants suffered from chronic acquired dysgraphia provides a homogeneity of deficit type absent in many studies, it is still the case that the participants suffered from different sub-types of dysgraphia and various degrees of other language and cognitive deficits. Third, there are a large number of graph-theoretic measures that could have been included, and which might have provided different insights into the characteristics of the functional neural networks that support recovery. Furthermore, a more targeted analysis approach could have been adopted, evaluating specific networks and connections (e.g., Van Hees et al., 2014; Sandberg et al., 2015). While there are certain advantages to examining global measures such as those used in this investigation, specific local effects are at risk of being over-looked. Fourth, even within the context of the *modularity,* PC and WD measures we adopted, different analysis decisions could have been made. We chose to define a reference network on the basis of neurotypical brain responses, but approaches that define reference networks based on Lesion Group data at the pre-timepoint could also be considered. Among other things that type of approach may result in a different distribution of global and local hubs, with consequences for PC and WD values.

## 5. Conclusions

It has long been recognized that interactions among brain regions play a pivotal role in supporting various cognitive processes. Graph-theoretic approaches offer powerful statistical metrics to evaluate neural connectivity data and hence have been shown to be a promising tool for disease detection and prognosis (Bullmore and Bassett, 2011; Carter et al., 2012; Braun et al., 2015b; Fornito et al., 2015; Ulm et al., 2018; Wig, 2017). The investigation reported here extends previous work directed at understanding the network characteristics that support recovery of function by using functional connectivity (rather than resting-state connectivity) to examine a larger set of related network characteristics with an experimental design that specifically allowed for the evaluation of treatment-driven recovery and comparisons with control participants. The results reveal that the analysis of network properties related to the *modularity* of the system provides a deeper understanding of the neural conditions that support performance and recovery in the face of brain injury. Specifically, in the context of recovery from dysgraphia, we found that recovery was associated with increases in *modularity* that was driven largely by increases in local connectivity. It will be especially interesting in future work to understand possible similarities and differences in network properties associated with recovery of different language and cognitive functions.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nicl.2019.101865.

## Funding

This work was supported by the National Institute of Deafness and Communication Disorders (grant number DC012283) to Brenda Rapp that is part of a P50 award supporting a multi-site project examining the neurobiology of language recovery in aphasia.

## Acknowledgments

We are grateful to NIH support for its support for this research (DC012283) and we thank Jennifer Shea, Jeremy Purcell, Donna Gotsch and Robert Wiley for their many valuable contributions to data collection and analysis.

## Appendix A. Appendix

Table A1
Individual scores of the 6 standard tests shown in Fig. 2c. All scores are percent correct measures at pre-/post-timepoint respectively (*pre score/post score*). Values in parentheses are the normative values provided with each test.

| Subject ID | Oral reading (100) | Single-word comprehension (86) | Auditory comprehension (N/A) | Recognition memory (71) | Semantic comprehension (98–99) | Spoken picture naming (N/A) |
|---|---|---|---|---|---|---|
| 1 – ABS | 87/95 | 80/84 | 96/100 | 71/83 | 96/98 | 73/88 |
| 2 – AEF | 88/93 | 57/67 | 96/100 | 67/54 | 88/88 | 65/65 |
| 3 – DSK | 57/53 | 76/77 | 100/100 | 71/79 | 92/94 | 62/62 |
| 4 – DTE | 97/93 | 70/70 | 100/100 | 75/88 | 98/96 | 58/77 |
| 5 – ESG | 68/76 | 50/60 | 100/100 | 58/58 | 90/88 | 69/81 |
| 6 – FCE | 75/80 | 63/73 | 100/100 | 88/79 | 98/100 | 88/92 |
| 7 – JGL | 50/68 | 73/60 | 100/100 | 71/92 | 98/100 | 58/62 |
| 8 – KMN | 33/52 | 73/80 | 100/100 | 58/79 | 92/94 | 19/38 |
| 9 – KST | 40/93 | 73/77 | 100/100 | 58/67 | 94/85 | 42/34 |
| 10 – MSO | 85/82 | 57/60 | 100/95 | 79/75 | 94/100 | 77/73 |
| 11 – PQS | 90/93 | 87/93 | 100/100 | 58/75 | 98/98 | 96/92 |
| 12 – RFZ | 98/97 | 87/73 | 100/100 | 100/96 | 96/100 | 100/100 |
| 13 – RHH | 95/96 | 77/77 | 100/100 | 83/83 | 90/99 | 92/92 |

(*continued on next page*)

Table A1 (*continued*)

| Subject ID | Oral reading (100) | Single-word comprehension (86) | Auditory comprehension (N/A) | Recognition memory (71) | Semantic comprehension (98–99) | Spoken picture naming (N/A) |
|---|---|---|---|---|---|---|
| 14 – RHN | 100/98 | 93/87 | 100/100 | 75/79 | 98/100 | 96/100 |
| 15 – TCK | 85/83 | 73/70 | 100/100 | 75/100 | 100/100 | 73/77 |

Oral reading: single-word oral reading (PALPA 35; Kay et al., 1992).
Single-word comprehension: single-word written comprehension (PALPA 51; Kay et al., 1992).
Auditory comprehension: auditory single-word comprehension (Northwestern Naming Battery; Thompson et al., 2012).
Recognition memory: Doors and People test (Baddeley et al., 1995).
Semantic comprehension: Pyramids and Palm Trees (Howard and Patterson, 1992).
Spoken picture naming: oral picture naming (Northwestern Naming Battery; Thompson et al., 2012).

## References


Abel, S., Weiller, C., Huber, W., Willmes, K., Specht, K., 2015. Therapy-induced brain reorganization patterns in aphasia. Brain *138*4, 1097–1112.
Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., ... Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn. Front. Neuroinform. 8, 14.
Achard, S., Salvador, R., Whitcher, B., Suckling, J., Bullmore, E.D., 2006. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. J. Neurosci. 261, 63–72.
Al-Aidroos, N., Said, C.P., Turk-Browne, N.B., 2012. Top-down attention switches coupling between low-level and high-level areas of human visual cortex. Proc. Natl. Acad. Sci. U. S. A. *109*36, 14675–14680.
Albert, R., Barabási, A.L., 2002. Statistical mechanics of complex networks. Rev. Mod. Phys. *74*1, 47.
Albert, R., Jeong, H., Barabási, A.L., 2000. Error and attack tolerance of complex networks. Nature *4066*794, 378.
Alstott, J., Breakspear, M., Hagmann, P., Cammoun, L., Sporns, O., 2009. Modeling the impact of lesions in the human brain. PLoS Comput. Biol. *5*6, e1000408.
Arnemann, K.L., Chen, A.J.W., Novakovic-Agopian, T., Gratton, C., Nomura, E.M., D'esposito, M., 2015. Functional brain network modularity predicts response to cognitive training after brain injury. Neurology *84*15, 1568–1574.
Baddeley, A.D., Wilson, B.A., Kopelman, M.D., 1995. Handbook of Memory Disorders. John Wiley and Sons Ltd., London, UK.
Baggio, H.C., Sala-Llonch, R., Segura, B., Marti, M.J., Valldeoriola, F., Compta, Y., ... Junqué, C., 2014. Functional brain networks and cognitive deficits in Parkinson's disease. Hum. Brain Mapp. 359, 4620–4634.
Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Treiman, R., 2007. The English lexicon project. Behav. Res. Methods 393, 445–459.
Bassett, D.S., Bullmore, E., Verchinski, B.A., Mattay, V.S., Weinberger, D.R., Meyer-Lindenberg, A., 2008. Hierarchical organization of human cortical networks in health and schizophrenia. J. Neurosci. 28 (37), 9239–9248.
Bassett, D.S., Wymbs, N.F., Porter, M.A., Mucha, P.J., Carlson, J.M., Grafton, S.T., 2011. Dynamic reconfiguration of human brain networks during learning. Proc. Natl. Acad. Sci. U. S. A. *108*18, 7641–7646.
Bassett, D.S., Wymbs, N.F., Rombach, M.P., Porter, M.A., Mucha, P.J., Grafton, S.T., 2013. Task-based core-periphery organization of human brain dynamics. PLoS Comput. Biol. *99*, e1003171.
Bertolero, M.A., Yeo, B.T., D'Esposito, M., 2015. The modular and integrative functional architecture of the human brain. Proc. Natl. Acad. Sci. U. S. A. *112*49, E6798–E6807.
Betzel, R.F., Byrge, L., He, Y., Goñi, J., Zuo, X.N., Sporns, O., 2014. Changes in structural and functional connectivity among resting-state networks across the human lifespan. Neuroimage 102, 345–357.
Blasi, V., Young, A.C., Tansy, A.P., Petersen, S.E., Snyder, A.Z., Corbetta, M., 2002. Word retrieval learning modulates right frontal cortex in patients with left frontal damage. Neuron 361, 159–170.
Braun, U., Schäfer, A., Walter, H., Erk, S., Romanczuk-Seiferth, N., Haddad, L., ... Meyer-Lindenberg, A., 2015a. Dynamic reconfiguration of frontal brain networks during executive cognition in humans. Proc. Natl. Acad. Sci. U. S. A. *112*37, 11678–11683.
Braun, U., Muldoon, S.F., Bassett, D.S., 2015b. On Human Brain Networks in Health and Disease. In: eLS. John Wiley & Sons, Ltd (Ed.). https://doi.org/10.1002/9780470015902.a0025783.
Brier, M.R., Thomas, J.B., Fagan, A.M., Hassenstab, J., Holtzman, D.M., Benzinger, T.L., ... Ances, B.M., 2014. Functional connectivity and graph theory in preclinical Alzheimer's disease. Neurobiol. Aging *354*, 757–768.
Buckner, R.L., Sepulcre, J., Talukdar, T., Krienen, F.M., Liu, H., Hedden, T., ... Johnson, K.A., 2009. Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. J. Neurosci. *296*, 1860–1873.
Bullmore, E., Bassett, D.S., 2011. Brain graphs: graphical models of the human brain connectome. Annu. Rev. Clin. Psychol. 7, 113–140.
Bullmore, E., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. *10*3, 186.
Bullmore, E., Sporns, O., 2012. The economy of brain network organization. Nat. Rev. Neurosci. *13*5, 336.
Caeyenberghs, K., Verhelst, H., Clemente, A., Wilson, P.H., 2017. Mapping the functional connectome in traumatic brain injury: what can graph metrics tell us? Neuroimage 160, 113–123.
Cao, Y., Vikingstad, E.M., George, K.P., Johnson, A.F., Welch, K.M., 1999. Cortical language activation in stroke patients recovering from aphasia with functional MRI. Stroke 3011, 2331–2340.
Carter, A.R., Astafiev, S.V., Lang, C.E., Connor, L.T., Rengachary, J., Strube, M.J., ... Corbetta, M., 2010. Resting interhemispheric functional magnetic resonance imaging connectivity predicts performance after stroke. Ann. Neurol. *67*3, 365–375.
Carter, A.R., Shulman, G.L., Corbetta, M., 2012. Why use a connectivity-based approach to study stroke and recovery of function? Neuroimage *62*4, 2271–2280.
Chan, M.Y., Park, D.C., Savalia, N.K., Petersen, S.E., Wig, G.S., 2014. Decreased segregation of brain systems across the healthy adult lifespan. Proc. Natl. Acad. Sci. U. S. A. *111*46, E4997–E5006.
Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., Dehaene, S., 2002. Language-specific tuning of visual cortex? Functional properties of the visual word form area. Brain *125*5, 1054–1069.
Cole, M.W., Reynolds, J.R., Power, J.D., Repovs, G., Anticevic, A., Braver, T.S., 2013. Multi-task connectivity reveals flexible hubs for adaptive task control. Nat. Neurosci. *16*9, 1348.
Cole, M.W., Bassett, D.S., Power, J.D., Braver, T.S., Petersen, S.E., 2014. Intrinsic and task-evoked network architectures of the human brain. Neuron *83*1, 238–251.
Crosson, B., Moore, A.B., Gopinath, K., White, K.D., Wierenga, C.E., Gaiefsky, M.E., ... Briggs, R.W., 2005. Role of the right and left hemispheres in recovery of function during treatment of intention in aphasia. J. Cogn. Neurosci. 173, 392–406.
Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31 (3), 968–980.
Dosenbach, N.U., Fair, D.A., Miezin, F.M., Cohen, A.L., Wenger, K.K., Dosenbach, R.A., ... Schlaggar, B.L., 2007. Distinct brain networks for adaptive and stable task control in humans. Proc. Natl. Acad. Sci. U. S. A. *104*26, 11073–11078.
Duncan, E.S., Small, S.L., 2016. Increased modularity of resting state networks supports improved narrative production in aphasia recovery. Brain Connect. *67*, 524–529.
Fair, D.A., Schlaggar, B.L., Cohen, A.L., Miezin, F.M., Dosenbach, N.U., Wenger, K.K., ... Petersen, S.E., 2007. A method for using blocked and event-related fMRI data to study "resting state" functional connectivity. Neuroimage *35*1, 396–405.
Fornito, A., Zalesky, A., Breakspear, M., 2015. The connectomics of brain disorders. Nat. Rev. Neurosci. *16*3, 159.
Fox, J., 2003. Effect displays in R for generalised linear models. J. Stat. Softw. 8 (15), 1–27. http://www.jstatsoft.org/v08/i15/.
Fox, J., Weisberg, S., 2018. Visualizing fit and lack of fit in complex regression models with predictor effect plots and partial residuals. Journal of Statistical Software 87 (9), 1–27.
Fridriksson, J., 2010. Preservation and modulation of specific left hemisphere regions is vital for treated recovery from anomia in stroke. J. Neurosci. 3035, 11558–11564.
Fridriksson, J., Morrow-Odom, L., Moser, D., Fridriksson, A., Baylis, G., 2006. Neural recruitment associated with anomia treatment in aphasia. Neuroimage 323, 1403–1412.
Fridriksson, J., Richardson, J.D., Fillmore, P., Cai, B., 2012. Left hemisphere plasticity and aphasia recovery. Neuroimage 602, 854–863.
Gallen, C.L., Baniqued, P.L., Chapman, S.B., Aslan, S., Keebler, M., Didehbani, N., D'Esposito, M., 2015. Modular brain network organization predicts response to cognitive training in older adults. PLoS One *111*2, e0169015.
Gamboa, O.L., Tagliazucchi, E., von Wegner, F., Jurcoane, A., Wahl, M., Laufs, H., Ziemann, U., 2014. Working memory performance of early MS patients correlates inversely with modularity increases in resting state functional connectivity networks. Neuroimage 94, 385–395.
Gold, B.T., Kertesz, A., 2000. Right hemisphere semantic processing of visual words in an aphasic patient: an fMRI study. Brain Lang. 733, 456–465.
Gratton, C., Nomura, E.M., Pérez, F., D'Esposito, M., 2012. Focal brain lesions to critical locations cause widespread disruption of the modular organization of the brain. J. Cogn. Neurosci. *246*, 1275–1285.
Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. Neuroimage *48*1, 63–72.
Guimera, R., Amaral, L.A.N., 2005. Functional cartography of complex metabolic networks. Nature *433*7028, 895.
Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O., 2008. Mapping the structural core of human cereb cortex. PLoS Biol. *67*, e159.

He, B.J., Snyder, A.Z., Vincent, J.L., Epstein, A., Shulman, G.L., Corbetta, M., 2007. Breakdown of functional connectivity in frontoparietal networks underlies behavioral deficits in spatial neglect. Neuron *536*, 905–918.
Heiss, W.D., Kessler, J., Thiel, A., Ghaemi, M., Karbe, H., 1999. Differential capacity of left and right hemispheric areas for compensation of poststroke aphasia. Ann. Neurol. 454, 430–438.
Honey, C.J., Sporns, O., 2008. Dynamical consequences of lesions in cortical networks. Hum. Brain Mapp. *297*, 802–809.
Howard, D., Patterson, K.E., 1992. The Pyramids and Palm Trees Test: A Test of Semantic Access from Words and Pictures. Thames Valley Test Company, Bury St Edmunds.
Jarso, S., Li, M., Faria, A., Davis, C., Leigh, R., Sebastian, R., ... Hillis, A.E., 2013. Distinct mechanisms and timing of language recovery after stroke. Cogn. Neuropsychology. 307–308 (454–475).
Jenkinson, M., Bannister, P.R., Brady, J.M., Smith, S.M., 2002. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. NeuroImage 172, 825–841.
Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. Fsl. NeuroImage 62, 782–790.
Kay, J., Lesser, R., Coltheart, M., 1992. Psycholinguistic Assessments of Language Processing in Aphasia PALPA. Erlbaum, Hove, UK.
Kiran, S., Meier, E.L., Kapse, K.J., Glynn, P.A., 2015. Changes in task-based effective connectivity in language networks following rehabilitation in post-stroke patients with aphasia. Front. Hum. Neurosci. 9, 316.
Kitzbichler, M.G., Henson, R.N., Smith, M.L., Nathan, P.J., Bullmore, E.T., 2011. Cognitive effort drives workspace configuration of human brain functional networks. J. Neurosci. *3122*, 8259–8270.
Kuest, J., Karbe, H., 2002. Cortical activation studies in aphasia. Curr. Neurol. Neurosci. Rep. 26, 511–515.
Lynall, M.E., Bassett, D.S., Kerwin, R., McKenna, P.J., Kitzbichler, M., Muller, U., Bullmore, E., 2010. Functional connectivity and brain networks in schizophrenia. J. Neurosci. *3028*, 9477–9487.
Marangolo, P., Fiori, V., Sabatini, U., De Pasquale, G., Razzano, C., Caltagirone, C., Gili, T., 2016. Bilateral transcranial direct current stimulation language treatment enhances functional connectivity in the left hemisphere: preliminary data from aphasia. J. Cogn. Neurosci. *285*, 724–738.
Marcotte, K., Adrover-Roig, D., Damien, B., de Preaumont, M., Genereux, S., Hubert, M., Ansaldo, A.I., 2012. Therapy-induced neuroplasticity in chronic aphasia. Neuropsychologia 50 (8), 1776–1786.
McCandliss, B.D., Cohen, L., Dehaene, S., 2003. The visual word form area: expertise for reading in the fusiform gyrus. Trends Cogn. Sci. *77*, 293–299.
Meier, E.L., Kapse, K.J., Kiran, S., 2016. The relationship between frontotemporal effective connectivity during picture naming, behavior, and preserved cortical tissue in chronic aphasia. Front. Hum. Neurosci. 10, 109.
Meunier, D., Achard, S., Morcom, A., Bullmore, E., 2009a. Age-related changes in modular organization of human brain functional networks. Neuroimage *443*, 715–723.
Meunier, D., Lambiotte, R., Fornito, A., Ersche, K., Bullmore, E.T., 2009b. Hierarchical modularity in human brain functional networks. Front Neuroinform 3, 37.
Newman, M.E., 2004. Fast algorithm for detecting community structure in networks. Phys. Rev. E 69 (6), 066133.
Newman, M.E., 2006. Modularity and community structure in networks. Proc. Natl. Acad. Sci. U. S. A. *1032*3, 8577–8582.
Newman, M.E., Girvan, M., 2004. Finding and evaluating community structure in networks. Phys. Rev. *692* (0), 26113.
Nomura, E.M., Gratton, C., Visser, R.M., Kayser, A., Perez, F., D'Esposito, M., 2010. Double dissociation of two cognitive control networks in patients with focal brain lesions. Proc. Natl. Acad. Sci. U. S. A. *1072*6, 12017–12022.
Norman-Haignere, S.V., McCarthy, G., Chun, M.M., Turk-Browne, N.B., 2011. Category-selective background connectivity in ventral visual cortex. Cereb. Cortex *222*, 391–402.
Planton, S., Jucla, M., Roux, F.E., Démonet, J.F., 2013. The "handwriting brain": a meta-analysis of neuroimaging studies of motor versus orthographic processes. Cortex *491*0, 2772–2787.
Postman-Caucheteux, W.A., Birn, R.M., Pursley, R.H., Butman, J.A., Solomon, J.M., Picchioni, D., ... Braun, A.R., 2010. Single-trial fMRI shows contralesional activity linked to overt naming errors in chronic aphasic patients. J. Cogn. Neurosci. 226, 1299–1318.
Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., ... Petersen, S.E., 2011. Functional network organization of the human brain. Neuron *724*, 665–678.
Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage *593*, 2142–2154.
Purcell, J., Turkeltaub, P.E., Eden, G.F., Rapp, B., 2011. Examining the central and peripheral processes of written word production through meta-analysis. Front. Psychol. 2, 239.
R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rapp, B., Dufor, O., 2011. The neurotopography of written word production: an fMRI investigation of the distribution of sensitivity to length and frequency. J. Cogn. Neurosci. 23 (12), 4067–4081.
Rapp, B., Kane, A., 2002. Remediation of deficits affecting different components of the spelling process. Aphasiology *164* (6), 439–454.
Rapp, B., Lipka, K., 2010. The literate brain: The relationship between reading and spelling. Journal of Cognitive Neuroscience.
Rapp, B., Purcell, J., Hillis, A.E., Capasso, R., Miceli, G., 2016. Neural bases of orthographic long-term memory and working memory in dysgraphia. Brain *1392*, 588–604.
Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. Neuroimage 523, 1059–1069.
Sandberg, C.W., Bohland, J.W., Kiran, S., 2015. Changes in functional connectivity related to direct training and generalization effects of a word finding treatment in chronic aphasia. Brain Lang. 150, 103–116.
Satterthwaite, T.D., Wolf, D.H., Ruparel, K., Erus, G., Elliott, M.A., Eickhoff, S.B., ... Hakonarson, H., 2013. Heterogeneous impact of motion on fundamental patterns of developmental changes in functional connectivity during youth. Neuroimage 83, 45–57.
Saur, D., Lange, R., Baumgaertner, A., Schraknepper, V., Willmes, K., Rijntjes, M., Weiller, C., 2006. Dynamics of language reorganization after stroke. Brain 1296, 1371–1384.
Sebastian, R., Long, C., Purcell, J., Faria, A.V., Lindquist, M., Jarso, S., ... Hillis, A.E., 2016. Imaging network level language recovery after left PCA stroke. Restor. Neurol. Neurosci. 344, 473–489.
Seghier, M.L., Zeidman, P., Neufeld, N.H., Leff, A.P., Price, C.J., 2010. Identifying abnormal connectivity in patients using dynamic causal modelling of fMRI responses. Front. Syst. Neurosci. 4, 142.
Seghier, M.L., Neufeld, N.H., Zeidman, P., Leff, A.P., Mechelli, A., Nagendran, A., ... Price, C.J., 2012. Reading without the left ventral occipito-temporal cortex. Neuropsychologia *501*4, 3621–3635.
Siegel, J.S., Ramsey, L.E., Snyder, A.Z., Metcalf, N.V., Chacko, R.V., Weinberger, K., ... Corbetta, M., 2016. Disruptions of network connectivity predict impairment in multiple behavioral domains after stroke. Proc. Natl. Acad. Sci. U. S. A. *1133*0, E4367–E4376.
Siegel, J.S., Seitzman, B.A., Ramsey, L.E., Ortega, M., Gordon, E.M., Dosenbach, N.U., ... Corbetta, M., 2018. Re-emergence of modular brain networks in stroke recovery. Cortex 101, 44–59.
Sporns, O., 2013. Network attributes for segregation and integration in the human brain. Curr. Opin. Neurobiol. *232*, 162–171.
Sporns, O., Honey, C.J., Kötter, R., 2007. Identification and classification of hubs in brain networks. PLoS One *210*, e1049.
Stam, C.J., Jones, B.F., Nolte, G., Breakspear, M., Scheltens, P., 2006. Small-world networks and functional connectivity in Alzheimer's disease. Cereb. Cortex *171*, 92–99.
Thirion, B., Varoquaux, G., Dohmatob, E., Poline, J.B., 2014. Which fMRI clustering gives good brain parcellations? Front. Neurosci. 8, 167.
Thompson, C.K., den Ouden, D.B., Bonakdarpour, B., Garibaldi, K., Parrish, T.B., 2010. Neural plasticity and treatment-induced recovery of sentence processing in agrammatism. Neuropsychologia *481*1, 3211–3227.
Thompson, C.K., Lukic, S., King, M.C., Mesulam, M.M., Weintraub, S., 2012. Verb and noun deficits in stroke-induced and primary progressive aphasia: the Northwestern naming battery. Aphasiology *265*, 632–655.
Thulborn, K.R., Carpenter, P.A., Just, M.A., 1999. Plasticity of language-related brain function during recovery from stroke. Stroke 304, 749–754.
Turkeltaub, P.E., Coslett, H.B., Thomas, A.L., Faseyitan, O., Benson, J., Norise, C., Hamilton, R.H., 2012. The right hemisphere is not unitary in its role in aphasia recovery. Cortex 489, 1179–1186.
Ulm, L., Copland, D., Meinzer, M., 2018. A new era of systems neuroscience in aphasia? Aphasiology *327*, 742–764.
Van Hees, S., McMahon, K., Angwin, A., de Zubicaray, G., Read, S., Copland, D.A., 2014. A functional MRI study of the relationship between naming treatment outcomes and resting state functional connectivity in post-stroke aphasia. Hum. Brain Mapp. *358*, 3919–3931.
Warren, J.E., Crinion, J.T., Lambon Ralph, M.A., Wise, R.J., 2009. Anterior temporal lobe connectivity correlates with functional outcome after aphasic stroke. Brain 13212, 3428–3442.
Warren, D.E., Power, J.D., Bruss, J., Denburg, N.L., Waldron, E.J., Sun, H., ... Tranel, D., 2014. Network measures predict neuropsychological outcome after brain injury. Proc. Natl. Acad. Sci. U. S. A. *1113*9, 14247–14252.
Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world'networks. Nature *393*6684, 440.
Wig, G.S., 2017. Segregated systems of human brain networks. Trends Cogn. Sci. 2112, 981–996.
Winhuisen, L., Thiel, A., Schumacher, B., Kessler, J., Rudolf, J., Haupt, W.F., Heiss, W.D., 2007. The right inferior frontal gyrus and poststroke aphasia: a follow-up investigation. Stroke 384, 1286–1292.
Yang, G.J., Murray, J.D., Wang, X.J., Glahn, D.C., Pearlson, G.D., Repovs, G., ... Anticevic, A., 2016. Functional hierarchy underlies preferential connectivity disturbances in schizophrenia. Proc. Natl. Acad. Sci. U. S. A. *1132*, E219–E228.
Yeo, B.T., Krienen, F.M., Eickhoff, S.B., Yaakub, S.N., Fox, P.T., Buckner, R.L., ... Chee, M.W., 2014. Functional specialization and flexibility in human association cortex. Cereb. Cortex *251*0, 3654–3672.
Yue, Q., Martin, R.C., Fischer-Baum, S., Ramos-Nuñez, A.I., Ye, F., Deem, M.W., 2017. Brain modularity mediates the relation between task complexity and performance. J. Cogn. Neurosci. *299*, 1532–1546.
Zhu, D., Chang, J., Freeman, S., Tan, Z., Xiao, J., Gao, Y., Kong, J., 2014. Changes of functional connectivity in the left frontoparietal network following aphasic stroke. Front. Behav. Neurosci. 8, 167.