Cognitive Neuropsychology

ISSN: 0264-3294 (Print) 1464-0627 (Online) Journal homepage: http://www.tandfonline.com/loi/pcgn20

Can neuroimaging help aphasia researchers? Addressing generalizability, variability, and interpretability

Idan A. Blank, Swathi Kiran & Evelina Fedorenko

To cite this article: Idan A. Blank, Swathi Kiran & Evelina Fedorenko (2017): Can neuroimaging help aphasia researchers? Addressing generalizability, variability, and interpretability, Cognitive Neuropsychology, DOI: 10.1080/02643294.2017.1402756

To link to this article: https://doi.org/10.1080/02643294.2017.1402756

Published online: 30 Nov 2017.

COGNITIVE NEUROPSYCHOLOGY, 2017
https://doi.org/10.1080/02643294.2017.1402756

Routledge
Taylor & Francis Group

Can neuroimaging help aphasia researchers? Addressing generalizability, variability, and interpretability

Idan A. Blank[a], Swathi Kiran[b] and Evelina Fedorenko[c,d]

[a]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA; [b]Department of Speech Language and Hearing Sciences, Aphasia Research Laboratory, Sargent College, Boston University, Boston, MA, USA; [c]Department of Psychiatry, Massachusetts General Hospital, Charlestown, MA, USA; [d]Department of Psychiatry, Harvard Medical School, Boston, MA, USA

ABSTRACT

Neuroimaging studies of individuals with brain damage seek to link brain structure and activity to cognitive impairments, spontaneous recovery, or treatment outcomes. To date, such studies have relied on the critical assumption that a given anatomical landmark corresponds to the same functional unit(s) across individuals. However, this assumption is fallacious even across neurologically healthy individuals. Here, we discuss the severe implications of this issue, and argue for an approach that circumvents it, whereby: (i) functional brain regions are defined separately for each subject using fMRI, allowing for inter-individual variability in their precise location; (ii) the response profile of these subject-specific regions are characterized using various other tasks; and (iii) the results are averaged across individuals, guaranteeing generalizabliity. This method harnesses the complementary strengths of single-case studies and group studies, and it eliminates the need for post hoc "reverse inference" from anatomical landmarks back to cognitive operations, thus improving data interpretability.

ARTICLE HISTORY
Received 8 March 2017
Revised 14 October 2017
Accepted 1 November 2017

KEYWORDS
Aphasia; fMRI; functional localization; group studies; individual differences

Three desiderata for cognitive neuropsychology

Cognitive neuropsychologists study individuals with brain damage in pursuit of two central aims: One is to analyse the behaviour of such individuals in terms of impaired cognitive functions and thus develop or constrain theories of intact (non-damaged) information-processing architectures (Caramazza, 1992); the other is to characterize the pattern of spontaneous and treatment-induced recovery and thus inform or evaluate different interventions. The success of this research programme critically depends on meeting three desiderata: (a) *generalizability*: The analysis of impaired behaviour should capture universal aspects of cognitive architecture that are compatible with data across different individuals with brain damage, and treatment effects should be reliable across individuals with similar deficits; (b) *taking into account variability*: Theories of cognitive architecture, as well as treatments, should account for the wide diversity in symptoms across individuals diagnosed with the same syndrome; and (c) *interpretability*: The experimental design should yield data that allow for valid inferences regarding both cognitive theories (which, in turn, should be computationally explicit enough to be informed by such data) and treatments.

Here, we discuss how these desiderata can be met in studies of individuals whose brain damage results in impaired language comprehension and/or production —that is, persons with aphasia (PWAs). Although the ideas we present are pertinent to the study of any high-level cognitive deficit following brain damage, our focus on aphasia is motivated by historical, pragmatic, and discursive considerations. Historically, early aphasia research is acknowledged as the origin of the cognitive neuropsychological discipline (Broca, 1861/2006; Dax, 1863; Lichtheim, 1885; Wernicke, 1874/1969), and it may be interesting to consider what progress has been made from the time of those first studies towards meeting the desiderata articulated above. Pragmatically, aphasia affects approximately one third of stroke survivors (Berthier, 2005), and more than 30% of these exhibit long-term, severe deficits (e.g., Bakheit et al., 2007). Indeed, research on aphasia has been prominently featured in this journal throughout its 34-year

CONTACT Idan A. Blank iblank@mit.edu

history. Discursively, and most importantly, focusing on aphasia frames this commentary within a long-standing, passionate debate on whether one can rationally assume that studies of PWAs are, in principle, capable of meeting the outlined desiderata.

The critical concern of this debate has been that aphasia researchers are seeking knowledge about the language-processing architecture using data that might be inherently inadequate for providing such knowledge. This debate has challenged cognitive neuropsychologists to re-examine the methodological foundations of their discipline, leading to two conflicting conclusions: Some scientists called for rejecting the common approach of "group studies", in which data are averaged across a sample of PWAs with a shared syndrome, in favour of "single-case studies" that comprehensively characterize the deficits of single individuals (Badecker & Caramazza, 1985; Caramazza, 1984, 1986, 1991; Caramazza & Badecker, 1989; Caramazza & McCloskey, 1988; McCloskey & Caramazza, 1988; Shallice, 1979); others, instead, defended group studies and questioned the validity of inferences made in single-case studies (Bates, Appelbaum, & Allard, 1991; Bub & Bub, 1988; Caplan, 1988; Grodzinsky, Piñango, Zurif, & Drai, 1999; Newcombe & Marshall, 1988; Robertson, Knight, Rafal, & Shimamura, 1993; Whitaker & Slotnick, 1988; Zurif, Gardner, & Brownell, 1989; Zurif, Swinney, & Fodor, 1991).

Opponents of group studies, led by Alfonso Caramazza, have argued that this approach fails to meet two desiderata: It does not take into account the variability in symptoms across PWAs, and does not yield interpretable data. First, they have criticized the existing groupings of PWAs under diagnostic categories such as "agrammatism" for their reliance on subjective clinical intuitions (e.g., Badecker & Caramazza, 1985). Empirical data, they suggested, did not provide necessary and sufficient conditions for inclusion in such categories given the substantial, unexplained differences observed across individuals with a putatively shared deficit. Data interpretation was thus severely limited, with no principled way of distinguishing meaningless differences within a category (i.e., noise) from meaningful differences across categories (reflecting theoretically distinct impairments). Consequently, these categories appeared to be psychologically arbitrary.

Second, these critics emphasized that such arbitrariness cannot be resolved by further experiments. Rather, categorizations for group studies are arbitrary by construction (Caramazza, 1986): In a group study, a set of PWAs all showing some behavioural signature (the grouping criterion) are hypothesized to have a functional impairment in a certain, common component of the language-processing architecture; this hypothesis is evaluated by testing whether this group exhibits, on average, another behavioural signature (the predicted symptom) that should result from this postulated impairment. However, such averaging is valid only if the group is homogeneous with respect to the predicted symptom (and the underlying impairment), for which there is no a priori guarantee. Thus, when a statistical test indicates that the predicted symptom is present in this group, it does not specify whether the symptom is (a) present in every individual, supporting the tested hypothesis; or (b) present in most but not all individuals, arguably licensing the rejection of the hypothesis and the adoption of a new categorization scheme: PWAs exhibiting both the grouping criterion and the predicted symptom belong in one category, whereas those exhibiting only the grouping criterion (but not the predicted symptom) belong in another. These two scenarios can be distinguished only through an analysis of each individual in the sample (i.e., a series of single-case analyses).

In contrast, proponents of group studies have pointed out that inter-individual variability in the predicted symptom is naturally incorporated into statistical inference: Significance tests evaluate the size of a group-averaged effect (the signal) against this variability (the noise), which is similar to unexplained noise in data from neurologically healthy individuals. High variability would not yield significant effects, and the resulting readjustment of infelicitous grouping criteria should, over time, converge towards theoretically meaningful syndromes (Zurif et al., 1989).

Moreover, group studies have been advocated for overcoming the limitations of single-case studies, which have been criticized for failing to meet the desiderata of generalizability and interpretability. Single-case studies give excessive weight to the behavioural symptoms of one individual despite lacking a criterion for distinguishing between behaviours that reflect damage to language mechanisms and those that are idiosyncratic, "accidental", and irrelevant (e.g., Zurif et al., 1989). For instance, the pattern of impaired and spared language functions of a PWA may reflect

some atypical premorbid neurocognitive characteristics, unusual behavioural strategies for solving experimental tasks, or a non-representative and extreme manifestation of a certain deficit (e.g., Rosenbaum, Gilboa, & Moscovitch, 2014; Varley, 1998). Taking such idiosyncrasies as evidence for constraining the language-processing architecture is a potentially precarious practice, especially when those idiosyncrasies are incompatible with a cognitive model that has previously received wide empirical support (Caramazza, 1986). Linking behavioural data to an underlying cognitive architecture based on a single case is further complicated because patterns of behavioural impairment are complex: Rather than exhibiting impaired performance in one task and intact performance in another, a patient would often show somewhat impaired performance in one task and disproportionately more impaired performance in another (e.g., Bi, Han, Shu, & Caramazza, 2007; Caramazza & Hillis, 1991; Rapp & Caramazza, 2002; see also Van Orden, Pennington, & Stone, 2001). Even less defensible is generalization from a case study when characterizing patterns of spontaneous or treatment-based recovery. Overall, given that a particular behavioural pattern often admits several cognitive explanations—however a priori unlikely some of them might be—a cognitive neuropsychologist ought to evaluate the relative support that each explanation receives from the data. Because differences in such relative support across hypotheses increase with the addition of independent observations, studies that examine multiple PWAs are sometimes claimed to confer higher generalizability. Simply put, group studies are assumed to average out the effects of potential idiosyncrasies on task performance. Nonetheless, some proponents of single-case studies maintain that this statistical sense of "generalizability" is irrelevant to their paradigm, which is argued to provide "existence proofs" of cognitive deficits.

A second, related threat to generalizability in single-case studies is that their findings are potentially non-falsifiable in both practice and principle. In practice, a PWA is often chosen as a case study based not on some clinical syndrome but on their performance of the critical tasks in that study. Such screening renders the inclusion criterion and the hypothesis non-independent, virtually assuring support for the hypothesis. The resulting case reports are biased in favour of PWAs whose behaviours exhibit the effect of interest, even if this effect characterizes very few of the individuals originally screened (Robertson et al., 1993). Furthermore, when a single-case study reports that two linguistic functions are dissociable—one can be selectively impaired while the other remains intact—the findings are interpreted as a proof of concept that these functions are cognitively separable. Thus, it is in principle impossible to obtain new data from other PWAs that would falsify this conclusion (Caramazza, 1986). However, if scientific progress consists of updating theories based on observed data, then the amount of neuropsychological evidence in support of a theory should be viewed against the amount of neuropsychological evidence against it.

Finally, some critics of single-case studies have asserted that the interpretable unit of analysis in aphasia research is the distribution of effects across individuals rather than individual data (Grodzinsky et al., 1999). Specifically, if a group of PWAs with a given syndrome are impaired on a certain task, they are not expected to all perform at chance but, rather, to be distributed around chance performance. Thus, an individual sampled from this group as a case study might perform above chance, leading to the erroneous conclusion that the syndrome in question does not entail impairment on the tested task. In contrast, a group study would yield a distribution of performance scores that is more consistent with chance performance than with no impairment.

Evidently, the debate over single-case and group studies has urged advocates of each approach to articulate their conflicting views on using data from PWAs to obtain knowledge about the language-processing architecture. This reflective process has had a formative influence on cognitive neuropsychology. At a more pragmatic level, both single-case and group studies have been well justified through their complementary contributions to cognitive science as well as to treatment development, which have been numerous and profound (Caramazza, 1992). Nonetheless, and despite some attempts at a compromise between these approaches (e.g., Bates, McDonald, MacWhinney, & Appelbaum, 1991), a behavioural methodology that synergistically harnesses their respective strengths in order to meet all three desiderata—generalizability, taking into account inter-individual variability, and interpretability—is yet to be found.

Struggling with irreconcilable desiderata in neuroimaging studies of aphasia

Over the past two decades, cognitive neuropsychologists have expanded their methodological arsenal beyond traditional measures of behavioural task performance as they have increasingly recognized the potential of neuroimaging data to inform their research questions. First, insofar as distinct components of the language-processing architecture engage separable neural circuits, we can expect that damage to a given circuit will consistently and selectively impair a certain component across PWAs. Thus, testing whether only those PWAs who share a putative syndrome have a common locus of brain damage (e.g., Bates et al., 2003) could indicate whether this syndrome is a natural category and, consequently, inform cognitive theories (e.g., Mesulam, Thompson, Weintraub, & Rogalski, 2015). Such research is, in essence, an extension of the cognitive neuroscientific programme (Mather, Cacioppo, & Kanwisher, 2013) to the study of PWAs. Second, patterns of brain reorganization following damage may have a prognostic value (e.g., van Hees, McMahon, Angwin, de Zubicaray, & Copland, 2014; van Hees, McMahon, Angwin, de Zubicaray, Read, et al., 2014) and inform treatment design (for a review, see: Meinzer, Harnish, Conway, & Crosson, 2011). Specifically, identifying those brain regions or networks whose functional properties are associated with language rehabilitation would both provide targets for neuro-stimulation treatments (e.g., Geranmayeh, Brownsett, & Wise, 2014; Hamilton, Chrysikou, & Coslett, 2011; Naeser et al., 2005) and motivate behavioural interventions to capitalize on the particular cognitive processes that engage these regions/networks (e.g., van Hees, McMahon, Angwin, de Zubicaray, & Copland, 2014).

Excitement about these potential contributions of neuroimaging techniques has been reflected in an increasing number of studies on aphasia employing such techniques, mostly fMRI (Geranmayeh et al., 2014). However, as fMRI gained popularity among aphasia researchers, they have gradually recognized that adapting neuroimaging methodologies—originally developed for the study of neurologically intact populations—had to confront the same challenges faced by behavioural methods: establishing generalizability, taking into account inter-individual variability, and maintaining interpretability. These issues have prompted collaborative efforts to develop methodological guidelines for the use of fMRI in aphasia research (Crosson et al., 2007; Kiran et al., 2013; Meinzer et al., 2013; Price, Crinion, & Friston, 2006; Rapp, Caplan, Edwards, Visch-Brink, & Thompson, 2013; Veldsman, Cumming, & Brodtmann, 2015; Wilson, Bautista, Yen, Lauderdale, & Eriksson, 2016). Naturally, in striving to meet the three desiderata, a critical focus of these efforts has been to negotiate the relative benefits (and respective costs) of group and single-case studies.

This negotiation ought to take into account the limitations of both fMRI and the study of brain-damaged populations. First, the signal measured with fMRI is heavily contaminated with fluctuations that are unexplained by the cognitive processes of interest. Such noise compromises the sensitivity to task-induced signal changes, especially given their small effect size (~1–2% increase relative to baseline). Thus, identifying signal differences across experimental conditions in data from a single participant often mandates that the contrasted conditions be substantially distinct (e.g., sentence comprehension vs. fixation; Crosson et al., 2007), but such gross contrasts cannot isolate a single cognitive process. Subtle contrasts (e.g., listening to sentences that are locally ambiguous vs. unambiguous) require larger amounts of data for sufficient statistical power; and more data are easier to obtain by studying a group of PWAs than a single individual due to, for example, attrition, practice effects, and constraints on scan duration (especially for older individuals, some of whom find the scanner environment uncomfortable).

Second, lesions vary highly in extent and location across PWAs and typically do not neatly correspond to a single brain region or cognitive function, such that an accurate mapping of components in a language-processing architecture onto the brain requires pooling data across individuals (Bates et al., 2003). Third, establishing that patterns of brain reorganization associated with recovery are generalizable rather than idiosyncratic to a particular individual also requires such pooling. Finally, identifying brain activation patterns that are related to behaviour is contingent on observing variability in task performance that can be correlated with variability in fMRI data. Although such variability can be obtained in a longitudinal study of a single PWA throughout recovery,

in practice its measurement is more feasible across a group of PWAs.

Perhaps unsurprisingly then, neuroimaging studies of aphasia have overall favoured group studies over single-case studies. In such studies, summary statistics characterizing a sample of PWAs are obtained by collecting spatially localized measures of brain structure or activity for each individual, and then averaging these data across the group. Critically, this averaging requires a principled way for establishing correspondence across brains, which naturally differ in size, shape, and folding patterns. The common solution is to transform data from all individual brains such that they match a template of a "standard", or "average", brain. This normalization into a "common space" establishes anatomical correspondence across brains —that is, a one-to-one mapping between stereotactic coordinates in this space and anatomical landmarks, rendering data pooling straightforward.

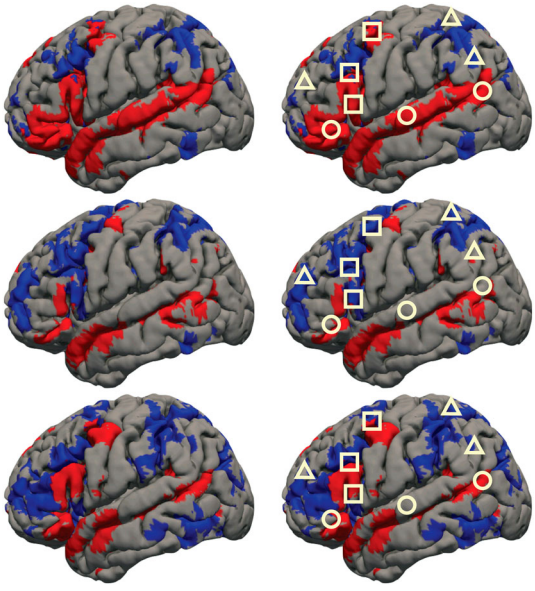

Figure 1. Left: language-selective regions (red) and domain-general regions engaged across many cognitively effortful tasks (blue) in the left hemisphere of three neurologically healthy adults. The former regions, identified with a passive reading task, show a stronger response to sentences than lists of nonwords (Fedorenko et al., 2011); the latter regions, identified with a working-memory task that requires keeping track of several spatial locations, are more engaged during trials that are harder (8 locations) than during those that are easier (4 locations; Fedorenko et al., 2013). Activations for both tasks are thresholded at the $p < .001$, uncorrected whole-brain level. Right: the same individual maps highlighting the variability in the precise locations of functional activity. Squares: sample regions that are language-selective in one individual but domain-general in another. Circles: sample regions that are language-selective in one individual but are not recruited during either task in another. Triangles: sample regions that are domain-general in one individual but are not recruited during either task in another. [To view this figure in colour, please see the online version of this Journal.]

This procedure crucially relies on the implicit assumption that, across individuals, a given anatomical location houses the same functional (i.e., cognitive) unit(s). However, this assumption is demonstrably invalid. The mapping of function onto macroanatomy exhibits high inter-individual variability even in the neurologically intact population (Duffau, 2017), in line with variability in the mapping between micro-anatomical structure and sulci/gyri, as has been long established (e.g., Brodmann, 1909/1994). This structural and functional variability is especially pronounced in associative cortices, which are of highest interest to aphasia researchers—that is, the temporal lobe (Gloor, 1997; Jones & Powell, 1970; Wise et al., 2001), the frontal lobe (Amunts et al., 1999; Juch, Zimine, Seghier, Lazeyras, & Fasel, 2005; Tomaiuolo et al., 1999), and the parietal lobe (Caspers et al., 2008; Caspers et al., 2006; Scheperjans et al., 2008). In these areas, regions with distinct functional profiles often lie side by side, but their precise anatomical locations are inconsistent across individuals. For instance, in Broca's area, regions that are selectively engaged in linguistic tasks are adjacent to regions that are engaged in cognitively effortful tasks across many domains (Fedorenko, Duncan, & Kanwisher, 2012); and in posterior superior temporal sulcus, language-selective regions are adjacent to regions that support mental state inference (Deen, Koldewyn, Kanwisher, & Saxe, 2015). In both of these areas, a given stereotactic coordinate may thus exhibit one functional profile in one individual and a different profile in another individual (Figure 1).

Because anatomy—at least at the current resolution of MRI—is not a precise predictor of function, group averaging of fMRI data based on anatomical alignment of brains is a precarious practice leading to erroneous inferences (Fischl et al., 2008; Frost & Goebel, 2012; Tahmasebi et al., 2012). First, when a functional region shows little spatial overlap across individuals, its activation might go undetected at the group level (compromised sensitivity). Thus, anatomical alignment is biased to reveal an effect of interest only in those brain regions whose activations are relatively strong, spatially extensive, or happen to be anatomically consistent across the individuals in a

particular sample; other regions that show the same effect are often missed. Such Type II errors might be mistakenly taken as evidence that the effect of interest is spatially restricted to a subset of some functional network or even localized to a single region. Moreover, an effect that appears localized to a certain region in some studies might instead appear localized to a different region in other studies (see, e.g., Blank, Balewski, Mahowald, & Fedorenko, 2016, for a discussion of such errors in the case of syntactic processing). Second, anatomically aligned data might conflate functionally distinct activations when they spatially overlap across individuals such that, at the group level, they appear to originate from a single region (low functional resolution; Fedorenko & Kanwisher, 2009). This illusory co-localization of effects (Type I error) might be taken as evidence against the functional specialization of a brain region for a particular cognitive process, even though, in each individual, the two effects arise in distinct, albeit adjacent, regions (see, e.g., Fedorenko, Duncan, et al., 2012, for a discussion of language specificity in Broca's area).

Third, even if one could easily determine which group-based effects are representative of individual brains rather than artefacts of anatomical alignment across participants, the interpretation of such effects in terms of mental functions is severely limited. Namely, when we detect group-level activity in a certain anatomical location (e.g., Broca's area) during some task (e.g., object naming), we must use post hoc "reverse inference" from stereotactic coordinates back to the underlying cognitive process (e.g., prior studies have reported that Broca's area is engaged in linguistic processing). However, such inferences are not deductively valid (e.g., activity in Broca's area might reflect either a language-specific operation or general attention/task engagement; Poldrack, 2006). Thus, deciding which reverse inferences happen to be veridical and which are false often amounts to a "gambling game" (but see Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011; Yarkoni, Poldrack, Van Essen, & Wager, 2010).

These issues are exacerbated when studying brain-damaged populations, due to several added sources of inter-individual variability: (a) Lesion extent and location are highly variable, compromising group-level detection of perilesional activations (Crosson et al., 2007); (b) PWAs with lesions in a common anatomical location can present with different symptoms (Basso, Bracchi, Capitani, & Laiacona, 1987; Hillis et al., 2004); and (c) PWAs with similar symptoms vary in the anatomical location of their respective lesions (Bonilha & Fridriksson, 2009; Dronkers, 2000; Mesulam et al., 2015; Newhart, Ken, Kleinman, Heidler-Gary, & Hillis, 2007).

Unfortunately, whereas cognitive neuropsychologists recognize these sources of variability across PWAs, they seldom account for the fundamental variability in functional-to-anatomical mapping in their neuroimaging studies (see also Rorden & Karnath, 2004); even when group studies are rejected in favour of single-case studies that can capture individual-level relationships between lesions, activity in spared brain regions, and behaviour, the cognitive interpretation of neuroimaging data proceeds through precarious reverse inference. The result is a literature with apparent inconsistencies that do not fit naturally within a unified framework. For instance, whereas spontaneous recovery in PWAs appears to be associated with both left-hemispheric and right-hemispheric recruitment, conflicting evidence suggests that the latter might play either a compensatory/restorative role or a maladaptive/detrimental role (Anglade, Thiel, & Ansaldo, 2014; Hamilton et al., 2011; Heiss & Thiel, 2006; Turkeltaub, Messing, Norise, & Hamilton, 2011), even within the same individual (Turkeltaub et al., 2012). In addition, within each hemisphere, many brain regions that engage in language processing in neurologically healthy individuals (Binder et al., 1997; Fedorenko, Behr, & Kanwisher, 2011; Fedorenko & Thompson-Schill, 2014; Jung-Beeman, 2005; Menenti, Gierhan, Segaert, & Hagoort, 2011) have been irregularly reported to also be engaged during spontaneous recovery, but few could be reliably predicted to do so (Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004; Meinzer et al., 2011; Mesulam et al., 2015; Mirman et al., 2015). Moreover, as described above, interpreting activations based on macroanatomical landmarks conflates language-specific mechanisms with domain-general resources related to cognitive effort or with mechanisms that support information processing in domains other than language (Geranmayeh et al., 2014). Similar conundrums are widespread in studies of treatment-induced brain changes (Crinion & Leff, 2007; Meinzer & Breitenstein, 2008; Thompson & den Ouden, 2008).

Neuroimaging studies of aphasia have so far struggled to offer a methodological approach that

meets all three desiderata of cognitive neuropsychological research. Namely, attempts at generalization beyond a single case have consistently ignored critical inter-individual variability in the precise anatomical location of functional regions, thus paradoxically hindering the replicability of results across studies. Furthermore, such variability renders the linking of anatomy back to cognitive processes logically flawed and, hence, potentially erroneous. These issues affect virtually all existing neuroimaging investigations of PWAs, including (a) voxel-based, lesion–symptom mapping analyses of anatomical data (Bates et al., 2003; Dronkers et al., 2004; Geva et al., 2011; Mesulam et al., 2015; Mirman et al., 2015; Wilson, 2016); (b) group-level analyses of functional data in a common space, for identifying stereotactic coordinates that show an effect of interest across the sample (e.g., studies contrasting the recruitment of the two hemispheres during spontaneous recovery, cited above); (c) group- and individual-level functional characterizations of particular brain regions that are chosen based on an independent, but group-based, criterion such as an independent task (Sharp, Turkheimer, Bose, Scott, & Wise, 2010) or data from neurologically healthy individuals (Bonner & Grossman, 2012; Fridriksson, Bonilha, Baker, Moser, & Rorden, 2010); and (d) comparisons of fMRI data across a series of single cases on the basis of anatomical alignment. The implications, for anyone who regards neuroimaging as a valid research method in cognitive neuropsychology, are alarming.

A possible reconciliation: Group-level analysis of subject-specific functional regions

One might wonder whether the criticisms articulated above are not limited only to fMRI studies of brain-damaged individuals and, thus, might more generally compromise the validity of most neuroimaging studies. Indeed, already two decades ago, cognitive neuroscientists became cognizant of the challenges associated with analysing functional data from anatomically aligned brains. An elegant methodology for circumventing these issues, originally developed for studying low-level visual processes (e.g., Sereno, Dale, Reppas, & Kwong, 1995; Tootell et al., 1995), was brought by Nancy Kanwisher to the study of higher level cognition in a seminal study of face perception (Kanwisher, McDermott, & Chun, 1997). In particular, she matched brain regions across individuals based directly on the observable functional profiles of those regions rather than through structural alignment, eschewing the reliance on anatomical landmarks as a proxy for function.

Abandoning the requirement for precise anatomical correspondence across brains in favour of functional correspondence has since become a heralded standard in neuroimaging studies of vision and is effectively required for targeting many visual processes (some critiques of this approach have been summarized in Friston, Rotshtein, Geng, Sterzer, & Henson, 2006; for a rebuttal of these early misunderstandings, see: Saxe, Brett, & Kanwisher, 2006). In contrast, most cognitive neuroscientists studying the language network have not embraced this method, barring few early adopters (Ben-Shachar, Hendler, Kahn, Ben-Bashat, & Grodzinsky, 2003; Ben-Shachar, Palti, & Grodzinsky, 2004; Hickok, Buchsbaum, Humphries, & Muftuler, 2003; January, Trueswell, & Thompson-Schill, 2009; Neville et al., 1998). More recently, however, the method has been adapted in a principled manner to the study of language processing by Fedorenko, Hsieh, Nieto-Castañón, Whitfield-Gabrieli, and Kanwisher (2010), who argued for its superiority over existing practices on both theoretical (Nieto-Castañón & Fedorenko, 2012; see also Saxe et al., 2006) and empirical (e.g., Blank et al., 2016) grounds. Consequently, it has been increasingly employed to revisit, refine and, often, challenge traditional views on the language-processing architecture (Axelrod, Bar, Rees, & Yovel, 2015; Basilakos, Smith, Fillmore, Fridriksson, & Fedorenko, 2017; Blank, Kanwisher, & Fedorenko, 2014; Chai, Mattar, Blank, Fedorenko, & Bassett, 2016; Deen et al., 2015; Fedorenko et al., 2011; Fedorenko, Duncan, et al., 2012; Fedorenko, Fillmore, Smith, Bonilha, & Fridriksson, 2015; Fedorenko, McDermott, Norman-Haignere, & Kanwisher, 2012; Fedorenko, Nieto-Castañón, & Kanwisher, 2012a; Fedorenko et al., 2016; Humphreys & Gennari, 2014; Hung et al., 2015; Mahowald & Fedorenko, 2016; Overath, McDermott, Zarate, & Poeppel, 2015; Prado, Mutreja, & Booth, 2013; Redcay, Velnoskey, & Rowe, 2016).

Establishing functional correspondence across brains proceeds as follows (Figure 2): First, fMRI data are collected while participants perform a task designed to identify, in each individual, the location of brain regions that exhibit a specific functional

signature. For example, to identify regions engaged in high-level language processing, the "localizer" task can contrast linguistically well-formed materials and linguistically "degraded" materials that are matched on low-level properties (e.g., sentences vs. lists of non-words or acoustically-degraded speech; Fedorenko et al., 2010; Scott, Gallée, & Fedorenko, 2016); and to define regions sensitive to general cognitive demands (e.g., Duncan, 2010), the localizer task can contrast a harder and an easier version of a demanding executive task (e.g., Fedorenko, Duncan, & Kanwisher, 2013). For each participant, these functional data are processed to produce an individual map showing the effect size of the localizer contrast for each location in the brain.

Next, each individual map is intersected with a set of spatial "masks" that denote, grossly, where the localizer task is expected to elicit activations, taking into account the anticipated inter-individual differences in the precise location of these activations. These masks effectively constrain the search for individual activation loci; still, within their borders, those loci are free to vary across participants [Desideratum (b): accounting for variability]. These masks can be based on macroanatomical landmarks, like a gyrus or a sulcus (or a portion thereof), chosen based on previous studies. However, activations do not always map neatly onto such landmarks. An alternative is thus to derive the set of masks algorithmically in a data-driven way from the localizer activation patterns in an (ideally, independent) set of participants: Regions of activation with the highest empirical overlap across participants are identified, and the borders around these local maxima are gradually extended until the resulting masks are large enough to encompass the activations of a sufficient percentage of individuals [Desideratum (a): generalizability] (for full details, see Fedorenko et al., 2010; Julian, Fedorenko, Webster, & Kanwisher, 2012). In studies of PWAs, masks can be generated in different ways depending on the research goals: First, masks based on data from neurologically healthy participants allow the localization of regions whose functional profile in intact brains is well characterized. As long as these regions still show some response to the localizer task in PWAs, one can examine whether the recruitment of these regions during other tasks changes following brain damage or recovery. Second, masks based on previous aphasia research allow the localization of regions that are consistently recruited during language processing only following brain damage or during recovery and whose functional profile in intact brains is yet to be determined. (How consistent in behaviour and lesion location a sample of PWAs should be in order to generate reliable masks remains to be empirically determined). Finally, we strongly recommend always examining whole-brain activation maps instead of merely testing activations within the constraints of pre-determined spatial masks; it is indeed possible (especially in cases of brain damage) that some new brain region would emerge outside the boundaries of these large masks. If that region appears to be present in a substantial proportion of PWAs, one would want to include it in the main analysis. Regardless of the process chosen for generating them, the resulting masks are then tailored to each individual PWA by excluding their particular lesion(s)—that is, limiting the search space for activations to intact tissue, capable of producing blood-oxygen-level-dependent (BOLD) signal.

The intersection of an individual's activation map with a mask for a particular region defines the participant-specific area within that mask that is responsive to the localizer contrast (based on, e.g., a statistical threshold). Because these areas are localized based on functional data, they are called functional regions of interest (fROIs). Critically, the obtained fROIs correspond to the same functional unit across individuals, insofar as they exhibit gross anatomical consistency and a characteristic response to the localizer task [Desideratum (c): interpretability]. Therefore, functional responses to the main, critical task(s) of the experiment can be extracted from these subject-specific fROIs and submitted to a group-level analysis, even if the fROI locations do not precisely overlap across the sample (for further discussion, see Saxe et al., 2006).

To establish that data from the localizer task allow for a reliable definition of fROIs, it is critical to confirm that the localizer activations are reproducible; otherwise, localization based on these activations might capture random noise instead of functional units with consistent response profiles [Desideratum (c): interpretability]. To this end, a cross-validation procedure should be employed: Here, a portion of the localizer data (e.g., even runs) can be used to define the fROIs, and the held-out data (odd runs) can then

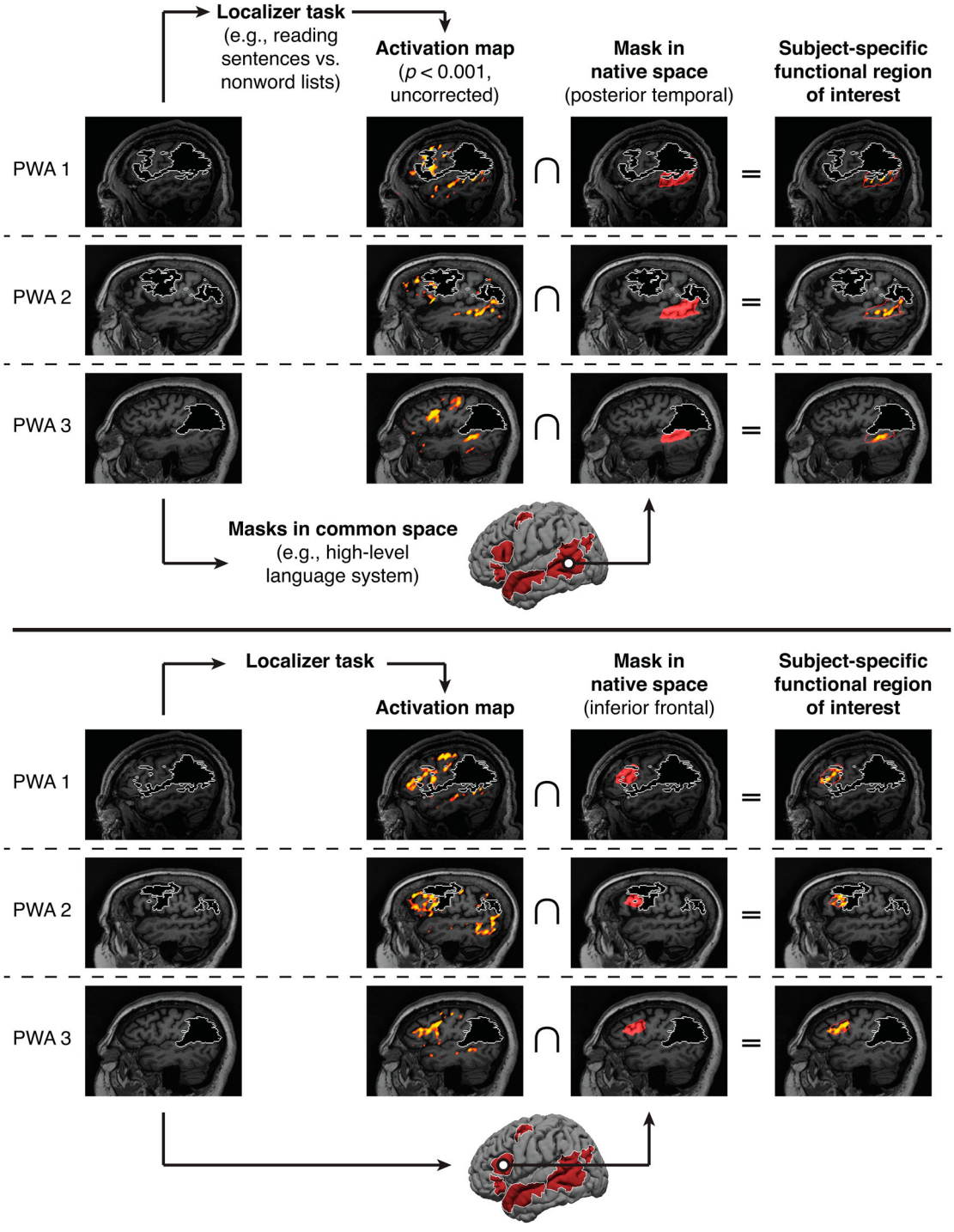

Figure 2. Illustration of steps for defining subject-specific functional regions of interest (fROIs). Anatomical scans for three individual persons with aphasia (PWAs) are shown in native space, with demarcated lesions. fROIs are defined by intersecting (a) individual activations during a localizer task (here, targeting the high-level language system); and (b) broad masks denoting where such activations are likely to occur, taking into account inter-individual variability in their precise locations (here, masks are derived from the localizer data of 220 neurologically healthy, young individuals, and are projected onto the native space of each PWA). Top: posterior temporal mask. Bottom: inferior frontal mask. For further details on the localizer task and masks, see Fedorenko et al. (2010). [To view this figure in colour, please see the online version of this Journal.]

be used to independently estimate responses to the localizer contrast in these fROIs. The estimates can then be tested across the sample to verify that the fROIs show the expected response profile (which served as the criterion for their definition). Importantly, such cross-validation also allows one to

retroactively define fROIs in experiments where localizer data have not been originally collected, by splitting the data from the critical task into halves; each half in turn serves as "localizer" data for defining fROIs that respond to the critical task, with the remaining half serving as the "main" data for testing whether the responses in these fROIs are robust and replicable (otherwise, they are false positives; for further discussion, see Nieto-Castañón & Fedorenko, 2012).

Whereas reliability is a necessary condition for functional localization in individual subjects, it is not sufficient: The localizer task should also be valid. Indeed, localizer validity might be compromised by two problems, but both can be empirically tested in a straightforward manner. To illustrate these problems, consider the following scenario: We hypothesize that in PWAs the right-hemispheric homologue of the language network is engaged in syntactic processing. Our critical task therefore tests whether the regions of this network respond more strongly to syntactically complex sentences than to syntactically simpler sentences in a sentence–picture matching task. Nonetheless, we must first functionally localize this network in each individual subject. What should our localizer task be?

We might choose a localizer task that is highly similar to the critical task—that is, requires syntactic processing. Perhaps we would search for regions that respond more strongly to sentences with syntactic errors than to well-formed sentences (assuming we have established that the activations for such a localizer are robust enough to be reliably detected in individual subjects given the amount of data collected). Here, a potential concern is that the localizer task might bias the data in favour of our hypothesis: It might identify only those regions within the right-hemispheric language network that are engaged in syntactic processing, but miss other regions within this network. For instance, if linguistic processing in PWAs were divided across regions that were each specialized for either lexical or combinatorial processing,[1] then the syntactic-violation-based localizer would fail to identify the former regions. Our data would thus erroneously indicate that all (rather than some or, even, few) right-hemispheric language regions are engaged in syntactic processing. To mitigate this concern, we could instead design a localizer task that contrasts two conditions differing not only in their combinatorial characteristics but also in their lexical characteristics (see suggestions earlier in this section).

Following the reasoning above, we might be tempted to design a localizer that is as dissimilar as possible to the critical task. For instance, given that our main task contrasts sentences with complex versus simple syntactic structures, our localizer could avoid sentences altogether and search for brain regions that respond more strongly to words than to reading nonwords in an *n*-back task (again, assuming that the resulting activations are reliably detectable at the individual-subject level). However, a potential concern here is that the localizer might miss any regions that are engaged in only combinatorial linguistic processes but not lexical linguistic processes. If linguistic labour in PWAs followed a lexical–combinatorial division, our localizer would fail to identify the very regions that are critical for testing the syntactic-processing hypothesis (but we note that traditional approaches fare much worse in this respect).

To mitigate this concern, an independent run of the critical task could itself be used as a functional localizer, guaranteeing the identification of all regions that reliably scale their activity with syntactic complexity (note that this procedure guards against under-inclusion, unlike the aforementioned practice of behaviourally screening PWAs for single-case studies based on the critical deficit itself, which might contribute to under-inclusion). In this case, at least one additional task is required to demonstrate that the functionally localized regions indeed belong to the language network rather than, for example, scale their response with difficulty in any cognitive domain. Often, an elegant solution is to run an additional analysis where the critical task serves as the localizer, and the localizer task serves as the main task. In the current example, this analysis would test whether all regions that are localized based on their engagement in syntactic processing also respond more strongly to words than to nonwords during the *n*-back task (a response profile that is inconsistent with a general difficulty account, given that nonwords are harder to process).

As cognitive theories are refined, the functional localizers that are motivated by these theories evolve. Debates surrounding any particular localizer task are a natural part of this process and, as we have demonstrated above, are subject to empirical investigation. To the extent that scientific

controversies are inevitable, we hope that they would focus on designing a set of acceptable localizers for neuroimaging studies of PWAs, not on the value of this indispensable approach (discussions could also focus on developing ways to infer individual functional regions from patterns of anatomical connectivity via diffusion tensor imaging, e.g., Saygin et al., 2012 ; or from patterns of functional correlations via resting-state fMRI, e.g., Cohen et al., 2008; Eickhoff, Thirion, Varoquaux, & Bzdok, 2015; Fox, Liu, & Pascual-Leone, 2013; Li, Langley, Li, & Hu, 2015). For those hypotheses that are amenable to investigation via fMRI, this approach maximizes the likelihood of gaining accurate insights. In fact, beyond its well-established superiority over traditional approaches that are based on anatomical alignment across brains of neurologically intact individuals, the group-level analysis of subject-specific fROIs confers benefits that are specifically critical to the study of PWAs.

First, this method can detect an effect of a given magnitude (and estimate that magnitude with higher fidelity; Nieto-Castañón & Fedorenko, 2012) with smaller samples than would otherwise be necessary, owing to its increased power, so the often limited pool of PWAs is less of a concern. At the same time, subject-specific localization obviates the need for excessive smoothing that attempts to compensate for inter-individual variability in data from small samples (Mikl et al., 2008). Second, normalization of functional data into a common space is unnecessary when fROIs are subject-specific, so analyses can proceed in the "native space" of each individual to minimize data transformations. Although such analyses instead require that the masks constraining fROI location be transformed from the common space into the native space, this process is relatively insensitive to imperfect alignment because these masks cover large areas of the brain. Third, activity in subject-specific fROIs has been shown to predict behaviour or causally influence it, across several functional networks (Assem, Blank, Mineroff, Ademoglu, & Fedorenko, 2017; Pitcher, Charles, Devlin, Walsh, & Duchaine, 2009; Young, Camprodon, Hauser, Pascual-Leone, & Saxe, 2010), and establishing such brain–behaviour relationships is crucial for individualized neuro-stimulation treatments.

Interestingly, this approach also provides a way to estimate the response profile that fROIs in PWAs may have exhibited prior to brain damage. Such estimation is informative for inferring the mechanisms underlying recovery in a damaged brain—for example, reorganization of impaired functions, compensation by spared regions that are not part of the impaired network, or pre-existing functional redundancies. To help adjudicate between these hypotheses, we have recently developed a "virtual lesion" procedure (Blank, Rohter, Kiran, & Fedorenko, 2015): First, an image of the lesion from a given PWA is superimposed on the functional data from a matched control participant (or a sample of participants who are matched on average). Then, fROIs in the control participant(s) are re-defined under the constraint that they cannot fall within the virtual lesion and must instead be identified in the "intact" parts of each mask. These fROIs serve as proxies for the neural architecture surrounding and preceding the lesion, and their responses to a variety of tasks can be compared to those of fROIs defined in PWAs. Admittedly, any insights provided by this analysis maintain considerable uncertainty, but no current neuroimaging method offers even this limited inferential power.

The most important, and unprecedented, advantage of this approach is its synergistic harnessing of the respective strengths of group and single-case studies in order to meet all three desiderata of cognitive neuropsychology. Like group studies, it provides *generalizable results* because it tests for an effect of interest across a sample and, moreover, surpasses traditional analyses in sensitivity and functional resolution (Nieto-Castañón & Fedorenko, 2012). Like single-case studies, it *takes into account inter-individual variability* that either had preceded or was induced by the lesion, because it relies on subject-specific localization rather than on anatomical alignment across brains. In addition, the obtained results are *functionally interpretable* without recourse to reverse inference from macro-anatomical landmarks back to mental operations (Poldrack, 2006, 2011), because fROIs are defined by their engagement in the cognitive processes targeted by the localizer task. Further, results for a set of fROIs from different studies and labs that employ similar functional localizers can be related to one another in a straightforward manner, avoiding the invalid comparison of stereotactic coordinates across studies (e.g., Fedorenko, Nieto-Castañón, & Kanwisher, 2012b). Such integration of knowledge is required for a cumulative research enterprise.

Conclusion

We have emphasized that meeting the three desiderata of cognitive neuropsychology in neuroimaging studies is a methodologically complicated issue, and described at length the proposed solution: group-level analyses of data from subject-specific functional regions. This intricate discussion, however, might belie the simplicity of implementing this method. Specifically, the entire analysis pipeline—from data-driven generation of masks, through cross-validated fROI definition, to group-level analyses of activations in these fROIs during the critical experimental task—can be performed using a publicly available toolbox (https://www.nitrc.org/projects/spm_ss). Moreover, localizer tasks for defining fROIs that are engaged in high-level language processing, including passive listening tasks that are suitable for PWAs, have been developed and thoroughly validated (Fedorenko et al., 2010; Scott et al., 2016; Stoppelman, Harpaz, & Ben-Shachar, 2013; Tie et al., 2015; Wilson et al., 2016), and are available for download (https://evlab.mit.edu/funcloc/download-paradigms). In addition, these localizers are now available in 41 languages (https://evlab.mit.edu/alice).

To date, only a handful of studies on aphasia have employed subject-specific functional localization, all for the purpose of identifying individualized targets for neuro-stimulation (Baker, Rorden, & Fridriksson, 2010; Fridriksson, Richardson, Baker, & Rorden, 2011; Kakuda, Abo, Kaito, Watanabe, & Senoo, 2010). Nonetheless, we have argued that this method promises to significantly advance the rigorous use of neuroimaging for the study of PWAs in pursuit of other aims as well. Whether the goal is to understand the neurocognitive mechanisms underlying recovery, to apply such knowledge to the design of behavioural interventions, or to generally elucidate the language-processing architecture of the human mind, our conviction is one: It is imperative that we standardize the use of group-level analyses of subject-specific functional regions.

Note

1. We refer to the dissociation between lexical-semantics and syntax for illustrative purposes only. fMRI data from neurologically intact individuals (using subject-specific functional localization) demonstrate that each and every region in the left-hemispheric language network responds to both lexical-semantics and syntax across different experimental paradigms (Fedorenko, Mineroff, Siegelman, & Blank, 2017).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by National Institute of Child Health and Human Development [grant number K99/R00 award HD 057522].

References

Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H., & Zilles, K. (1999). Broca's region revisited: Cytoarchitecture and intersubject variability. *Journal of Comparative Neurology*, *412*(2), 319–341. <319::AID-CNE10>3.0.CO;2-7

Anglade, C., Thiel, A., & Ansaldo, A. I. (2014). The complementary role of the cerebral hemispheres in recovery from aphasia after stroke: A critical review of literature. *Brain Injury*, *28*(2), 138–145. doi:10.3109/02699052.2013.859734

Assem, M., Blank, I. A., Mineroff, Z., Ademoglu, A., & Fedorenko, E. (2017). Multiple demand system's activity predicts individual differences in working memory and fluid intelligence.

Axelrod, V., Bar, M., Rees, G., & Yovel, G. (2015). Neural correlates of subliminal language processing. *Cerebral Cortex*, *25*(8), 2160–2169. doi:10.1093/cercor/bhu022

Badecker, W., & Caramazza, A. (1985). On considerations of method and theory governing the use of clinical categories in neurolinguistics and cognitive neuropsychology: The case against agrammatism. *Cognition*, *20*(2), 97–125. doi:10.1016/0010-0277(85)90049-6

Baker, J. M., Rorden, C., & Fridriksson, J. (2010). Using transcranial direct-current stimulation to treat stroke patients with aphasia. *Stroke*, *41*(6), 1229–1236. doi:10.1161/STROKEAHA.109.576785

Bakheit, A., Shaw, S., Barrett, L., Wood, J., Carrington, S., Griffiths, S., … Koutsi, F. (2007). A prospective, randomized, parallel group, controlled study of the effect of intensity of speech and language therapy on early recovery from poststroke aphasia. *Clinical Rehabilitation*, *21*(10), 885–894. doi:10.1177/0269215507078486

Basilakos, A., Smith, K. G., Fillmore, P., Fridriksson, J., & Fedorenko, E. (2017). Functional characterization of the human speech articulation network. *Cerebral Cortex*, 1–15. doi:10.1093/cercor/bhx100

Basso, A., Bracchi, M., Capitani, E., & Laiacona, M. (1987). Age and evolution of language area functions: A study of adult stroke patients. *Cortex*, *23*, 475–483. doi:10.1016/S0010-9452(87)80008-4

Bates, E., Appelbaum, M., & Allard, L. (1991). Statistical constraints on the use of single cases in neuropsychological

research. *Brain and Language*, *40*(3), 295–329. doi:10.1016/0093-934X(91)90132-K

Bates, E., McDonald, J., MacWhinney, B., & Appelbaum, M. (1991). A maximum likelihood procedure for the analysis of group and individual data in aphasia research. *Brain and Language*, *40*(2), 231–265. doi:10.1016/0093-934X(91)90126-L

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion–symptom mapping. *Nature Neuroscience*, *6*(5), 448–450. https://doi.org/10.1101/110270

Ben-Shachar, M., Hendler, T., Kahn, I., Ben-Bashat, D., & Grodzinsky, Y. (2003). The neural reality of syntactic transformations evidence from functional magnetic resonance imaging. *Psychological Science*, *14*(5), 433–440. doi:10.1111/1467-9280.01459

Ben-Shachar, M., Palti, D., & Grodzinsky, Y. (2004). Neural correlates of syntactic movement: Converging evidence from two fMRI experiments. *NeuroImage*, *21*(4), 1320–1336. doi:10.1016/j.neuroimage.2003.11.027

Berthier, M. L. (2005). Poststroke aphasia: Epidemiology, pathophysiology and treatment. *Drugs & Aging*, *22*(2), 163–182. doi:10.2165/00002512-200522020-00006

Bi, Y., Han, Z., Shu, H., & Caramazza, A. (2007). Nouns, verbs, objects, actions, and the animate/inanimate effect. *Cognitive Neuropsychology*, *24*(5), 485–504. doi:10.1080/02643290701502391

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *The Journal of Neuroscience*, *17*(1), 353–362.

Blank, I. A., Balewski, Z., Mahowald, K., & Fedorenko, E. (2016). Syntactic processing is distributed across the language system. *Neuroimage*, *127*, 307–323. doi:10.1016/j.neuroimage.2015.11.069

Blank, I. A., Kanwisher, N., & Fedorenko, E. (2014). A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *Journal of Neurophysiology*, *112*(5), 1105–1118. doi:10.1152/jn.00884.2013

Blank, I. A., Rohter, S. V., Kiran, S., & Fedorenko, E. (2015). *Functional reorganization of the large-scale brain networks that support high-level cognition following brain damage in aphasia*. Academy of Aphasia 53rd annual meeting, Tuscon, AZ.

Bonilha, L., & Fridriksson, J. (2009). Subcortical damage and white matter disconnection associated with non-fluent speech. *Brain*, *132*(6), e108–e108. doi:10.1093/brain/awn200

Bonner, M. F., & Grossman, M. (2012). Gray matter density of auditory association cortex relates to knowledge of sound concepts in primary progressive aphasia. *Journal of Neuroscience*, *32*(23), 7986–7991. doi:10.1523/JNEUROSCI.6241-11.2012

Broca, P. (1861/2006). Comments regarding the seat of the faculty of spoken language, followed by an observation of aphemia (loss of speech). In Y. Grodzinsky & K. Amunts (Eds.), *Broca's region* (pp. 291–304). New York, NY: Oxford University Press.

Brodmann, K. (1909/1994). *Localisation in the cerebral cortex: The principles of comparative localisation in the cerebral cortex based on cytoarchitectonics (L. J. Garey, Trans.)*. London: Smith-Gordon Company Ltd.

Bub, J., & Bub, D. (1988). On the methodology of single-case studies in cognitive neuropsychology. *Cognitive Neuropsychology*, *5*(5), 565–582. doi:10.1080/02643298808253275

Caplan, D. (1988). On the role of group studies in neuropsychological and pathopsychological research. *Cognitive Neuropsychology*, *5*(5), 535–547. doi:10.1080/02643298808253273

Caramazza, A. (1984). The logic of neuropsychological research and the problem of patient classification in aphasia. *Brain and Language*, *21*(1), 9–20. doi:10.1016/0093-934X(84)90032-4

Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: The case for single-patient studies. *Brain and Cognition*, *5*(1), 41–66. doi:10.1016/0278-2626(86)90061-8

Caramazza, A. (1991). Data, statistics, and theory: A comment on Bates, McDonald, MacWhinney, and Applebaum's "A maximum likelihood procedure for the analysis of group and individual data in aphasia research". *Brain and Language*, *41*(1), 43–51. doi:10.1016/0093-934X(91)90109-E

Caramazza, A. (1992). Is cognitive neuropsychology possible? *Journal of Cognitive Neuroscience*, *4*(1), 80–95. doi:10.1162/jocn.1992.4.1.80

Caramazza, A., & Badecker, W. (1989). Patient classification in neuropsychological research. *Brain and Cognition*, *10*(2), 256–295. doi:10.1016/0278-2626(89)90056-0

Caramazza, A., & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature*, *349*(6312), 788–790. doi:10.1038/349788a0

Caramazza, A., & McCloskey, M. (1988). The case for single-patient studies. *Cognitive Neuropsychology*, *5*(5), 517–527. doi:10.1080/02643298808253271

Caspers, S., Eickhoff, S. B., Geyer, S., Scheperjans, F., Mohlberg, H., Zilles, K., & Amunts, K. (2008). The human inferior parietal lobule in stereotaxic space. *Brain Structure and Function*, *212*(6), 481–495. doi:10.1007/s00429-008-0195-z

Caspers, S., Geyer, S., Schleicher, A., Mohlberg, H., Amunts, K., & Zilles, K. (2006). The human inferior parietal cortex: Cytoarchitectonic parcellation and interindividual variability. *NeuroImage*, *33*(2), 430–448. doi:10.1016/j.neuroimage.2006.06.054

Chai, L. R., Mattar, M. G., Blank, I. A., Fedorenko, E., & Bassett, D. S. (2016). Functional network dynamics of the language system. *Cerebral Cortex*, *26*(11), 4148–4159. doi:10.1093/cercor/bhw238

Cohen, A. L., Fair, D. A., Dosenbach, N. U. F., Miezin, F. M., Dierker, D., Van Essen, D. C., … Petersen, S. E. (2008). Defining functional areas in individual human brains using resting functional connectivity MRI. *NeuroImage*, *41*(1), 45–57. doi:10.1016/j.neuroimage.2008.01.066

Crinion, J. T., & Leff, A. P. (2007). Recovery and treatment of aphasia after stroke: Functional imaging studies. *Current*

Opinion in Neurology, *20*(6), 667–673. doi:10.1097/WCO.0b013e3282f1c6fa

Crosson, B., McGregor, K., Gopinath, K. S., Conway, T. W., Benjamin, M., Chang, Y.-L., … Sherod, M. G. (2007). Functional MRI of language in aphasia: A review of the literature and the methodological challenges. *Neuropsychology Review*, *17*(2), 157–177. doi:10.1007/s11065-007-9024-z

Dax, M. (1863). Observations tendant à prouver la coïncidence constante des dérangements de la parole avec une lésion de l'hémisphère gauche du cerveau. *Comptes rendus de l'Académie des Sciences*, *56*, 536.

Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional organization of social perception and cognition in the superior temporal sulcus. *Cerebral Cortex*, *25*(11), 4596–4609. doi:10.1093/cercor/bhv111

Dronkers, N. F. (2000). The gratuitous relationship between Broca's aphasia and Broca's area. *Behavioral and Brain Sciences*, *23*(01), 30–31. doi:10.1017/S0140525X00322397

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*(1), 145–177. doi:10.1016/j.cognition.2003.11.002

Duffau, H. (2017). A two-level model of interindividual anatomo-functional variability of the brain and its implications for neurosurgery. *Cortex*, *86*, 303–313. doi:10.1016/j.cortex.2015.12.009

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179. doi:10.1016/j.tics.2010.01.004

Eickhoff, S. B., Thirion, B., Varoquaux, G., & Bzdok, D. (2015). Connectivity-based parcellation: Critique and implications. *Human Brain Mapping*, *36*(12), 4771–4792. doi:10.1002/hbm.22933

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, *108*(39), 16428–16433. doi:10.1073/pnas.1112937108

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology*, *22*(21), 2059–2062. doi:10.1016/j.cub.2012.09.011

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, *110*(41), 16616–16621. doi:10.1073/pnas.1315235110

Fedorenko, E., Fillmore, P., Smith, K., Bonilha, L., & Fridriksson, J. (2015). The superior precentral gyrus of the insula does not appear to be functionally specialized for articulation. *Journal of Neurophysiology*, *113*(7), 2376–2382. doi:10.1152/jn.00214.2014

Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. *Journal of Neurophysiology*, *104*(2), 1177–1194. doi:10.1152/jn.00032.2010

Fedorenko, E., & Kanwisher, N. (2009). Neuroimaging of language: Why hasn't a clearer picture emerged? *Language and Linguistics Compass*, *3*(4), 839–865. doi:10.1111/j.1749-818X.2009.00143.x

Fedorenko, E., McDermott, J. H., Norman-Haignere, S., & Kanwisher, N. (2012). Sensitivity to musical structure in the human brain. *Journal of Neurophysiology*, *108*(12), 3289–3300. doi:10.1152/jn.00209.2012

Fedorenko, E., Mineroff, Z., Siegelman, M., & Blank, I. (2017). *The distinction between lexico-semantic and syntactic processing is not an organizing dimension of the human language system*. CUNY sentence processing conference, Cambridge, MA.

Fedorenko, E., Nieto-Castañón, A., & Kanwisher, N. (2012a). Lexical and syntactic representations in the brain: An fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia*, *50*(4), 499–513. doi:10.1016/j.neuropsychologia.2011.09.014

Fedorenko, E., Nieto-Castañón, A., & Kanwisher, N. (2012b). Syntactic processing in the human brain: What we know, what we don't know, and a suggestion for how to proceed. *Brain and Language*, *120*(2), 187–207. doi:10.1016/j.bandl.2011.01.001

Fedorenko, E., Scott, T. L., Brunner, P., Coon, W. G., Pritchett, B., Schalk, G., & Kanwisher, N. (2016). Neural correlate of the construction of sentence meaning. *Proceedings of the National Academy of Sciences*, *113*(41), E6256–E6262. doi:10.1073/pnas.1612132113

Fedorenko, E., & Thompson-Schill, S. L. (2014). Reworking the language network. *Trends in Cognitive Sciences*, *18*(3), 120–126. doi:10.1016/j.tics.2013.12.006

Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B. T. T., … Zilles, K. (2008). Cortical folding patterns and predicting cytoarchitecture. *Cerebral Cortex*, *18*(8), 1973–1980. doi:10.1093/cercor/bhm225

Fox, M. D., Liu, H., & Pascual-Leone, A. (2013). Identification of reproducible individualized targets for treatment of depression with TMS based on intrinsic connectivity. *NeuroImage*, *66*, 151–160. doi:10.1016/j.neuroimage.2012.10.082

Fridriksson, J., Bonilha, L., Baker, J. M., Moser, D., & Rorden, C. (2010). Activity in preserved left hemisphere regions predicts anomia severity in aphasia. *Cerebral Cortex*, *20*(5), 1013–1019. doi:10.1093/cercor/bhp160

Fridriksson, J., Richardson, J. D., Baker, J. M., & Rorden, C. (2011). Transcranial direct current stimulation improves naming reaction time in fluent aphasia. *Stroke*, *42*(3), 819–821. doi:10.1161/STROKEAHA.110.600288

Friston, K. J., Rotshtein, P., Geng, J. J., Sterzer, P., & Henson, R. N. (2006). A critique of functional localisers. *Neuroimage*, *30*(4), 1077–1087. doi:10.1016/j.neuroimage.2005.08.012

Frost, M. A., & Goebel, R. (2012). Measuring structural–functional correspondence: Spatial variability of specialised brain regions after macro-anatomical alignment. *NeuroImage*, *59*(2), 1369–1381. doi:10.1016/j.neuroimage.2011.08.035

Geranmayeh, F., Brownsett, S. L., & Wise, R. J. (2014). Task-induced brain activity in aphasic stroke patients: What is

driving recovery? *Brain*, *137*(10), 2632–2648. doi:10.1093/brain/awu163

Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J.-C., & Warburton, E. A. (2011). The neural correlates of inner speech defined by voxel-based lesion–symptom mapping. *Brain*, *134*(10), 3071–3082. doi:10.1093/brain/awr232

Gloor, P. (1997). *The temporal lobe and limbic system*. New York, NY: Oxford University Press.

Grodzinsky, Y., Piñango, M. M., Zurif, E., & Drai, D. (1999). The critical role of group studies in neuropsychology: Comprehension regularities in Broca's aphasia. *Brain and Language*, *67*(2), 134–147. doi:10.1006/brln.1999.2050

Hamilton, R. H., Chrysikou, E. G., & Coslett, B. (2011). Mechanisms of aphasia recovery after stroke and the role of noninvasive brain stimulation. *Brain and Language*, *118*(1), 40–50. doi:10.1016/j.bandl.2011.02.005

Heiss, W.-D., & Thiel, A. (2006). A proposed regional hierarchy in recovery of post-stroke aphasia. *Brain and Language*, *98*(1), 118–123. doi:10.1016/j.bandl.2006.02.002

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory–motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, *15*(5), 673–682. doi:10.1162/089892903322307393

Hillis, A. E., Work, M., Barker, P. B., Jacobs, M. A., Breese, E. L., & Maurer, K. (2004). Re-examining the brain regions crucial for orchestrating speech articulation. *Brain*, *127*(7), 1479–1487. doi:10.1093/brain/awh172

Humphreys, G. F., & Gennari, S. P. (2014). Competitive mechanisms in sentence processing: Common and distinct production and reading comprehension networks linked to the prefrontal cortex. *NeuroImage*, *84*, 354–366. doi:10.1016/j.neuroimage.2013.08.059

Hung, Y.-H., Pallier, C., Dehaene, S., Lin, Y.-C., Chang, A., Tzeng, O. J.-L., & Wu, D. H. (2015). Neural correlates of merging number words. *NeuroImage*, *122*, 33–43. doi:10.1016/j.neuroimage.2015.07.045

January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience*, *21*(12), 2434–2444. doi:10.1162/jocn.2008.21179

Jones, E. G., & Powell, T. P. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain*, *93*(4), 793–820. doi:10.1093/brain/93.4.793

Juch, H., Zimine, I., Seghier, M. L., Lazeyras, F., & Fasel, J. H. (2005). Anatomical variability of the lateral frontal lobe surface: Implication for intersubject variability in language neuroimaging. *NeuroImage*, *24*(2), 504–514. doi:10.1016/j.neuroimage.2004.08.037

Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, *60*(4), 2357–2364. doi:10.1016/j.neuroimage.2012.02.055

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences*, *9*(11), 512–518. doi:10.1016/j.tics.2005.09.009

Kakuda, W., Abo, M., Kaito, N., Watanabe, M., & Senoo, A. (2010). Functional MRI-based therapeutic rTMS strategy for aphasic stroke patients: A case series pilot study. *International Journal of Neuroscience*, *120*(1), 60–66. doi:10.3109/00207450903445628

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311.

Kiran, S., Ansaldo, A., Bastiaanse, R., Cherney, L. R., Howard, D., Faroqi-Shah, Y., … Thompson, C. K. (2013). Neuroimaging in aphasia treatment research: Standards for establishing the effects of treatment. *Neuroimage*, *76*, 428–435. doi:10.1016/j.neuroimage.2012.10.011

Li, K., Langley, J., Li, Z., & Hu, X. P. (2015). Connectomic profiles for individualized resting state networks and regions of interest. *Brain Connectivity*, *5*(2), 69–79. doi:10.1089/brain.2014.0229

Lichtheim, L. (1885). On aphasia. *Brain*, *7*, 433–484. doi:10.1093/brain/7.4.433

Mahowald, K., & Fedorenko, E. (2016). Reliable individual-level neural markers of high-level language processing: A necessary precursor for relating neural variability to behavioral and genetic variability. *NeuroImage*, *139*, 74–93. doi:10.1016/j.neuroimage.2016.05.073

Mather, M., Cacioppo, J. T., & Kanwisher, N. (2013). How fMRI can inform cognitive theories. *Perspectives on Psychological Science*, *8*(1), 108–113. doi:10.1177/1745691612469037

McCloskey, M., & Caramazza, A. (1988). Theory and methodology in cognitive neuropsychology: A response to our critics. *Cognitive Neuropsychology*, *5*(5), 583–623. doi:10.1080/02643298808253276

Meinzer, M., Beeson, P. M., Cappa, S., Crinion, J., Kiran, S., Saur, D., … Thompson, C. K. (2013). Neuroimaging in aphasia treatment research: Consensus and practical guidelines for data analysis. *Neuroimage*, *73*, 215–224. doi:10.1016/j.neuroimage.2012.02.058

Meinzer, M., & Breitenstein, C. (2008). Functional imaging studies of treatment-induced recovery in chronic aphasia. *Aphasiology*, *22*(12), 1251–1268. doi:10.1080/02687030802367998

Meinzer, M., Harnish, S., Conway, T., & Crosson, B. (2011). Recent developments in functional and structural imaging of aphasia recovery after stroke. *Aphasiology*, *25*(3), 271–290. doi:10.1080/02687038.2010.530672

Menenti, L., Gierhan, S. M. E., Segaert, K., & Hagoort, P. (2011). Shared language overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science*, *22*(9), 1173–1182. doi:10.1177/0956797611418347

Mesulam, M.-M., Thompson, C. K., Weintraub, S., & Rogalski, E. J. (2015). The Wernicke conundrum and the anatomy of language comprehension in primary progressive aphasia. *Brain*, *138*(8), 2423–2437. doi:10.1093/brain/awv154

Mikl, M., Mareček, R., Hluštík, P., Pavlicová, M., Drastich, A., Chlebus, P., … Krupa, P. (2008). Effects of spatial smoothing

on fMRI group inferences. *Magnetic Resonance Imaging*, *26* (4), 490–503. doi:10.1016/j.mri.2007.08.006

Mirman, D., Chen, Q., Zhang, Y., Wang, Z., Faseyitan, O. K., Coslett, H. B., & Schwartz, M. F. (2015). Neural organization of spoken language revealed by lesion-symptom mapping. *Nature Communications*, *6*, 6762. doi:10.1038/ncomms7762

Naeser, M. A., Martin, P. I., Nicholas, M., Baker, E. H., Seekins, H., Kobayashi, M., … Kurland, J. (2005). Improved picture naming in chronic aphasia after TMS to part of right Broca's area: An open-protocol study. *Brain and Language*, *93*(1), 95–105. doi:10.1016/j.bandl.2004.08.004

Neville, H. J., Bavelier, D., Corina, D., Rauschecker, J., Karni, A., Lalwani, A., … Turner, R. (1998). Cerebral organization for language in deaf and hearing subjects: Biological constraints and effects of experience. *Proceedings of the National Academy of Sciences*, *95*(3), 922–929. doi:10.1073/pnas.95.3.922

Newcombe, F., & Marshall, J. C. (1988). Idealisation meets psychometrics: The case for the right groups and the right individuals. *Cognitive Neuropsychology*, *5*(5), 549–564. doi:10.1080/02643298808253274

Newhart, M., Ken, L., Kleinman, J. T., Heidler-Gary, J., & Hillis, A. E. (2007). Neural networks essential for naming and word comprehension. *Cognitive and Behavioral Neurology*, *20*(1), 25–30. doi:10.1097/WNN.0b013e31802dc4a7

Nieto-Castañón, A., & Fedorenko, E. (2012). Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *NeuroImage*, *63*(3), 1646–1669. doi:10.1016/j.neuroimage.2012.06.065

Overath, T., McDermott, J. H., Zarate, J. M., & Poeppel, D. (2015). The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nature Neuroscience*, *18*(6), 903–911. doi:10.1038/nn.4021

Pitcher, D., Charles, L., Devlin, J. T., Walsh, V., & Duchaine, B. (2009). Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Current Biology*, *19*(4), 319–324. doi:10.1016/j.cub.2009.01.007

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63. doi:10.1016/j.tics.2005.12.004

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, *72*(5), 692–697. doi:10.1016/j.neuron.2011.11.001

Prado, J., Mutreja, R., & Booth, J. R. (2013). Fractionating the neural substrates of transitive reasoning: Task-dependent contributions of spatial and verbal representations. *Cerebral Cortex*, *23*(3), 499–507. doi:10.1093/cercor/bhr389

Price, C. J., Crinion, J., & Friston, K. J. (2006). Design and analysis of fMRI studies with neurologically impaired patients. *Journal of Magnetic Resonance Imaging*, *23*(6), 816–826. doi:10.1002/jmri.20580

Rapp, B., Caplan, D., Edwards, S., Visch-Brink, E., & Thompson, C. K. (2013). Neuroimaging in aphasia treatment research: Issues of experimental design for relating cognitive to neural changes. *Neuroimage*, *73*, 200–207. doi:10.1016/j.neuroimage.2012.09.007

Rapp, B., & Caramazza, A. (2002). Selective difficulties with spoken nouns and written verbs: A single case study. *Journal of Neurolinguistics*, *15*(3), 373–402. doi:10.1016/S0911-6044(01)00040-9

Redcay, E., Velnoskey, K. R., & Rowe, M. L. (2016). Perceived communicative intent in gesture and language modulates the superior temporal sulcus. *Human Brain Mapping*, *37*(10), 3444–3461. doi:10.1002/hbm.23251

Robertson, L. C., Knight, R. T., Rafal, R., & Shimamura, A. P. (1993). Cognitive neuropsychology is more than single-case studies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 710–717. doi:10.1037/0278-7393.19.3.710

Rorden, C., & Karnath, H.-O. (2004). Using human brain lesions to infer function: A relic from a past era in the fMRI age? *Nature Reviews Neuroscience*, *5*(10), 812–819. doi:10.1038/nrn1521

Rosenbaum, R. S., Gilboa, A., & Moscovitch, M. (2014). Case studies continue to illuminate the cognitive neuroscience of memory. *Annals of the New York Academy of Sciences*, *1316*(1), 105–133. doi:10.1111/nyas.12467

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: A defense of functional localizers. *NeuroImage*, *30*(4), 1088–1096. doi:10.1016/j.neuroimage.2005.12.062

Saygin, Z. M., Osher, D. E., Koldewyn, K., Reynolds, G., Gabrieli, J. D., & Saxe, R. R. (2012). Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nature neuroscience*, *15*(2), 321–327.

Scheperjans, F., Eickhoff, S. B., Hömke, L., Mohlberg, H., Hermann, K., Amunts, K., & Zilles, K. (2008). Probabilistic maps, morphometry, and variability of cytoarchitectonic areas in the human superior parietal cortex. *Cerebral Cortex*, *18*(9), 2141–2157. doi:10.1093/cercor/bhm241

Scott, T. L., Gallée, J., & Fedorenko, E. (2016). A new fun and robust version of an fMRI localizer for the frontotemporal language system. *Cognitive Neuroscience*. doi:0.1080/17588928.2016.1201466

Sereno, M. I., Dale, A., Reppas, J., & Kwong, K. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889–893. doi:10.1126/science.7754376

Shallice, T. (1979). Case study approach in neuropsychological research. *Journal of Clinical Neuropsychology*, *1*(3), 183–211. doi:10.1080/01688637908414450

Sharp, D. J., Turkheimer, F. E., Bose, S. K., Scott, S. K., & Wise, R. J. (2010). Increased frontoparietal integration after stroke and cognitive recovery. *Annals of Neurology*, *68*(5), 753–756. doi:10.1002/ana.21866

Stoppelman, N., Harpaz, T., & Ben-Shachar, M. (2013). Do not throw out the baby with the bath water: Choosing an effective baseline for a functional localizer of speech processing. *Brain and Behavior*, *3*(3), 211–222. doi:10.1002/brb3.129

Tahmasebi, A. M., Davis, M. H., Wild, C. J., Rodd, J. M., Hakyemez, H., Abolmaesumi, P., & Johnsrude, I. S. (2012). Is the link between anatomical structure and function equally strong at all cognitive levels of processing? *Cerebral Cortex*, *22*(7), 1593–1603. doi:10.1093/cercor/bhr205

Thompson, C. K., & den Ouden, D.-B. (2008). Neuroimaging and recovery of language in aphasia. *Current Neurology and Neuroscience Reports*, *8*(6), 475–483. doi:10.1007/s11910-008-0076-0

Tie, Y., Rigolo, L., Ozdemir Ovalioglu, A., Olubiyi, O., Doolin, K. L., Mukundan, S., & Golby, A. J. (2015). A new paradigm for individual subject language mapping: Movie-watching fMRI. *Journal of Neuroimaging*, *25*(5), 710–720. doi:10.1111/jon.12251

Tomaiuolo, F., MacDonald, J. D., Caramanos, Z., Posner, G., Chiavaras, M., Evans, A. C., & Petrides, M. (1999). Morphology, morphometry and probability mapping of the pars opercularis of the inferior frontal gyrus: An in vivo MRI analysis. *European Journal of Neuroscience*, *11*(9), 3033–3046. doi:10.1046/j.1460-9568.1999.00718.x

Tootell, R. B., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., … Belliveau, J. W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience*, *15*(4), 3215–3230.

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*, *48*(9), 1179–1186. doi:10.1016/j.cortex.2011.06.010

Turkeltaub, P. E., Messing, S., Norise, C., & Hamilton, R. H. (2011). Are networks for residual language function and recovery consistent across aphasic patients? *Neurology*, *76*(20), 1726–1734. doi:10.1212/WNL.0b013e31821a44c1

van Hees, S., McMahon, K., Angwin, A., de Zubicaray, G., & Copland, D. A. (2014). Neural activity associated with semantic versus phonological anomia treatments in aphasia. *Brain and Language*, *129*, 47–57. doi:10.1016/j.bandl.2013.12.004

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. *Human Brain Mapping*, *35*(8), 3919–3931. doi:10.1002/hbm.22448

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? *Cognitive Science*, *25*(1), 111–172. doi:10.1207/s15516709cog2501_5

Varley, R. (1998). Aphasia language, aphasic thought: An investigation of propositional thinking in an a-propositional aphasic. In P. Carruthers, & J. Boucher (Eds.), *Language and thought: Interdisciplinary themes* (pp. 128–145). Cambridge: Cambridge University Press.

Veldsman, M., Cumming, T., & Brodtmann, A. (2015). Beyond BOLD: Optimizing functional imaging in stroke populations. *Human Brain Mapping*, *36*(4), 1620–1636. doi:10.1002/hbm.22711

Wernicke, C. (1874/1969). The symptom complex of aphasia: A psychological study on an anatomical basis. In R. S. Cohen, & M. W. Wartofsky (Eds.), *Boston studies in the philosophy of science* (pp. 34–97). Dordrecht: D. Reidel Publishing.

Whitaker, H. A., & Slotnick, H. B. (1988). Comments on "the case for single patient studies": Is (neuro) psychology possible? *Cognitive Neuropsychology*, *5*(5), 529–533. doi:10.1080/02643298808253272

Wilson, S. M. (2016). Lesion-symptom mapping in the study of spoken language understanding. *Language, Cognition and Neuroscience*, *32*, 891–899. doi:10.1080/23273798.2016.1248984

Wilson, S. M., Bautista, A., Yen, M., Lauderdale, S., & Eriksson, D. K. (2016). Validity and reliability of four language mapping paradigms. *NeuroImage: Clinical*, *16*, 399–408. doi:10.1016/j.nicl.2016.03.015

Wise, R. J. S., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural subsystems within "Wernicke's area". *Brain*, *124*(1), 83–95. doi:10.1093/brain/124.1.83

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670. doi:10.1038/nmeth.1635

Yarkoni, T., Poldrack, R. A., Van Essen, D. C., & Wager, T. D. (2010). Cognitive neuroscience 2.0: Building a cumulative science of human brain function. *Trends in Cognitive Sciences*, *14*(11), 489–496. doi:10.1016/j.tics.2010.08.004

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences*, *107*(15), 6753–6758. doi:10.1073/pnas.0914826107

Zurif, E., Gardner, H., & Brownell, H. H. (1989). The case against the case against group studies. *Brain and Cognition*, *10*(2), 237–255. doi:10.1016/0278-2626(89)90055-9

Zurif, E., Swinney, D., & Fodor, J. A. (1991). An evaluation of assumptions underlying the single-patient-only position in neuropsychological research: A reply. *Brain and Cognition*, *16*(2), 198–210. doi:10.1016/0278-2626(91)90006-T