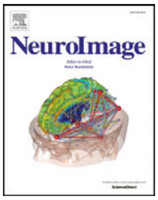
NeuroImage 139 (2016) 74–93

Contents lists available at ScienceDirect

NeuroImage

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



Reliable individual-level neural markers of high-level language processing: A necessary precursor for relating neural variability to behavioral and genetic variability

Kyle Mahowald [a,*], Evelina Fedorenko [b,c]

[a] *MIT, United States*
[b] *MGH, United States*
[c] *HMS, United States*

A R T I C L E I N F O

Article history:
Accepted 27 May 2016
Available online 31 May 2016

Keywords:
Language system
Individual differences
Neural markers of language activity
Reliability

A B S T R A C T

The majority of functional neuroimaging investigations aim to characterize an average human brain. However, another important goal of cognitive neuroscience is to understand the ways in which individuals differ from one another and the significance of these differences. This latter goal is given special weight by the recent reconceptualization of neurological disorders where sharp boundaries are no longer drawn either between health and neuropsychiatric and neurodevelopmental disorders, or among different disorders (e.g., Insel et al., 2010). Consequently, even the variability in the healthy population can inform our understanding of brain disorders. However, because the use of functional neural markers is still in its infancy, no consensus presently exists about which measures (e.g., effect size?, extent of activation?, degree of lateralization?) are the best ones to use. We here attempt to address this question with respect to one large-scale neural system: the set of brain regions in the frontal and temporal cortices that jointly support high-level linguistic processing (e.g., Binder et al., 1997; Fedorenko, Hsieh, Nieto-Castanon, Whitfield-Gabrieli, & Kanwisher, 2010). In particular, using data from 150 individuals all of whom had performed a language "localizer" task contrasting sentences and nonword sequences (Fedorenko et al., 2010), we: a) characterize the distributions of the values for four key neural measures of language activity (region effect sizes, region volumes, lateralization based on effect sizes, and lateralization based on volumes); b) test the reliability of these measures in a subset of 32 individuals who were scanned across two sessions; c) evaluate the relationship among the different regions of the language system; and d) evaluate the relationship among the different neural measures. Based on our results, we provide some recommendations for future studies of brain-behavior and brain-genes relationships. Although some of our conclusions are specific to the language system, others (e.g., the fact that effect-size-based measures tend to be more reliable than volume-based measures) are likely to generalize to the rest of the brain.

Introduction

The majority of studies in cognitive neuroscience seek to discover properties that are common to all individuals, to characterize an "average" human mind and brain. However, ways in which individuals differ from one another can also inform our understanding of human cognition. In psychology and cognitive science, investigations of individual differences in behavior have helped reveal the structure of – and the relationships among – many domains, including intelligence (e.g., Duncan et al., 2000; Gardner and Hatch, 1989; Kane and Engle, 2002; Spearman, 1904; Spearman, 1927), executive functions (e.g., Carlson et al., 2004; Colom, 2004; Conway, 1996; Mischel et al., 1989; Miyake et al., 2000), visual processing (e.g., Childers et al., 1985; Colombo et al., 1991; Vogel and Machizawa, 2004), social cognition (e.g., Herrmann et al., 2007; Miller and Saygin, 2013), speech perception (e.g., Surprenant and Watson, 2001), language comprehension (e.g., Daneman and Carpenter, 1980; Gernsbacher, 1991; Just and Carpenter, 1992; Pakulak and Neville, 2010; Traxler et al., 2012), music processing (e.g., Grahn and Schuit, 2012; Perrachione et al., 2013), and so on.

In addition to their importance for addressing questions in basic research, investigations of individual differences can shed light on neurological disorders. In particular, recent years have witnessed a shift in how mental illness is conceptualized, from the traditional, categorical, approach where sharp boundaries were drawn between health and neuropsychiatric and neurodevelopmental disorders, as well as among different neurological conditions (American Psychiatric Association,

---

[*] Corresponding author at: 77 Massachusetts Ave, Bldg 46-3037, Cambridge, MA 02139, United States.

E-mail address: kylemaho@mit.edu (K. Mahowald).

http://dx.doi.org/10.1016/j.neuroimage.2016.05.073
1053-8119/

2013) to a more probabilistic approach (e.g., Insel et al., 2010; Krug et al., 2010). Such a shift was inspired by a long-standing observation of variability present both within the healthy population and among individuals diagnosed with neurological disorders, combined with substantial overlap in the symptoms and genetic risk factors among disorders. This new way of thinking about mental illness calls for a shift in research practices: from group comparisons between, for example, individuals diagnosed with autism and neurotypical controls, to explorations of variability across large populations, to discover true endophenotypes.

Although a number of studies have attempted to link behavioral variability to genetic variability directly, including in the domain of language (e.g., Ocklenburg et al., 2013; Scerri et al., 2011; Whitehouse et al., 2011), *neural* markers are plausibly an important intermediate link given that genes shape the anatomy and functional organization of the brain, and these structural and functional characteristics of the brain in turn give rise to the observable behaviors. Indeed, neural markers – both anatomical and functional – are being used increasingly often in individual differences investigations of human cognition, including language (Cope et al., 2012; Krug et al., 2010; Landi et al., 2013; Pinel et al., 2012; Whalley et al., 2011). See Dubois and Adolphs (2016) for a thorough discussion of this approach.

At present, the use of *anatomical* markers is more common, plausibly due to the availability of large datasets, with hundreds, and sometimes thousands, of participants. Such datasets accumulate because most cognitive neuroscience labs routinely collect high-resolution structural scans from every participant. However, anatomical markers based purely on macroanatomy (e.g., the cortical thickness and/or volume of a macroanatomically defined brain area) have their limitations. In particular, the relationship between structure and function is a complex one, especially in the higher-order association cortices, where functional activations do not align well with the macroanatomical landmarks (e.g., Fischl et al., 2008; Frost and Goebel, 2012; Tahmasebi et al., 2012). For example, a well-characterized face-selective brain region – the fusiform face area (FFA; Kanwisher et al., 1997) – cannot be defined anatomically (e.g., Frost and Goebel, 2012). Consequently, markers of brain activity may provide a stronger link between genes and behavior, especially for higher-level cognitive processes. Furthermore, they can increase the power of anatomical investigations (e.g., studies examining cortical thickness) by enabling researchers to delineate the relevant brain regions more accurately than sulcal/gyral landmarks alone allow.

To successfully relate functional neural markers to genetic and behavioral variability, however, it is important – for each relevant cognitive function – to determine a) which markers are reliable (i.e., stable within individuals over time), and b) how different markers relate to one another. At present, in the domain of language research, different groups use different language tasks (e.g., semantic verbal fluency, verb generation, sentence completion, rhyme judgment tasks), focus on different brain regions (e.g., inferior frontal regions, regions in the middle temporal gyrus, or even regions outside of the core fronto-temporal language network), and examine different markers of neural activity (e.g., effect size for the relevant contrast in some region of interest, volume of an activated region, degree of lateralization of a region). Any one of prior studies individually can potentially reveal something important about language or cognition more broadly. However, the real power would come from the ability *to compare and replicate findings* across studies and research groups, to discover truly robust relationships. This could only be achieved if we, as a field, agreed on a set of tasks and measures that are reasonable, and adopted a set of guidelines for how to use those. For example, in increasingly more domains of study researchers use “functional localizer” tasks, which quickly and reliably identify a subset of the brain engaged in a particular mental activity (e.g., face-selective regions, Kanwisher et al., 1997; voice-selective regions, Belin et al., 2000; or regions engaged in theory of mind, Saxe and Kanwisher, 2003). Because labs that use functional localizers include a localizer scan in every participant, large datasets are eventually accumulated, as needed for brain-genes investigations. Further, because the same or comparable localizers are used across research groups, findings can be replicated across groups in a straightforward way.

Indeed, it has become increasingly clear that, in order to begin to link functional and behavioral data to underlying genetic variation, we will need large datasets involving hundreds or even thousands of participants. For instance, Stein et al. (2012), in a large-scale meta-analysis ($n = 7795$) of how genomic variation affects total brain volume, intracranial volume, and hippocampal volume, found that the largest observed effect (which was for hippocampal volume) explained only a tiny fraction of the variance. Hoogman et al. (2014) found similarly small effect sizes for the FOXP2 gene. Of course, it is possible that part of the difficulty in detecting these relationships between genetic and neural variation stems from the reliance on macroanatomical landmarks, which may fail to identify the “natural kinds” of the mind and brain, as discussed above. For example, the hippocampus is structurally and functionally diverse (e.g., Poppenk et al., 2013; Schoene-Bake et al., 2014; Travis et al., 2014) and perhaps detecting relationships between genetic variability and the volumes of its different subregions would be easier. However, even setting this issue aside, the effect sizes of the relationships between genetic and neural (anatomical or functional) variation are likely to be small because any given trait is a product of a vast number of genetic factors. Between small effect sizes and the huge space of possible variation in the genome, a well-powered study needs a large number of participants, such as the data now available from widely used functional localizer tasks.

We have recently developed methods for identifying the fronto-temporal system engaged in high-level linguistic processing using a contrast between sentences and sequences of nonwords (Fedorenko et al., 2010). This and similar contrasts have been used in many prior studies (e.g., words vs. fixation or tones: Binder et al., 1997; Diaz and McCarthy, 2009; words vs. pseudowords: Petersen et al., 1990; sentences vs. fixation: Kuperberg et al., 2003; sentences vs. false font or consonant strings: Bavelier et al., 1998; Noppeney and Price, 2004; Robertson et al., 2000; sentences vs. lists of words: Fedorenko and Kanwisher, 2011; Fedorenko et al., 2010; Snijders et al., 2009; speech vs. backwards or degraded speech: Bedny et al., 2011; Scott et al., in press), and we established that this contrast works robustly at the individual-subject level. We also demonstrated that this fronto-temporal language system exhibits a high degree of functional specificity: its regions respond robustly during language processing, but not during other complex cognitive tasks, like arithmetic processing, general working memory tasks or music perception (Fedorenko et al., 2011, 2012b). This system is thus functionally distinct from another large-scale brain network, which has a strong presence in the left prefrontal cortex: the bilateral fronto-parietal executive, or cognitive control, system (Duncan, 2013; Fedorenko et al., 2012a), and this dissociation holds even during naturalistic language comprehension (Blank et al., 2014).

The goal of the current study, which targets the fronto-temporal language system, is three-fold. *First*, using a large dataset of healthy adult participants ($n = 79$), we characterize activity in the language system in a number of ways: focusing on eight key language regions (Fig. 1) and their right-hemisphere homologs, we report the distributions of values for effect sizes, volumes, and lateralization (computed based on either effect sizes or volumes). These distributions clearly show that there is substantial variability to be explained even in the healthy population with respect to language activations. In addition, any new population can now be evaluated with respect to these normative distributions, be it older or younger individuals, left handers, learners of English as a second language, bi/multi-linguals, or individuals with neurodevelopmental or acquired disorders. The data for this set of participants are available at https://web.archive.org/web/20160608155930/https://evlab.mit.edu/papers/Mahowald_NI. *Second*, we evaluate the reliability of these functional measures in a subset of