

The Domain-General Multiple Demand (MD) Network Does Not Support Core Aspects of Language Comprehension: A Large-Scale fMRI Investigation

Evgeniia Diachek,^{1*}  Idan Blank,^{2*} Matthew Siegelman,^{3*} Josef Affourtit,⁴ and Evelina Fedorenko^{4,5,6}

¹Department of Psychology, Vanderbilt University, Nashville, Tennessee 37203, ²Department of Psychology, University of California at Los Angeles, Los Angeles, California 90095, ³Department of Psychology, Columbia University, New York, New York 10027, ⁴Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, ⁵McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and ⁶Department of Psychiatry, Massachusetts General Hospital, Charlestown, Massachusetts 02129

Aside from the language-selective left-lateralized frontotemporal network, language comprehension sometimes recruits a domain-general bilateral frontoparietal network implicated in executive functions: the multiple demand (MD) network. However, the nature of the MD network's contributions to language comprehension remains debated. To illuminate the role of this network in language processing in humans, we conducted a large-scale fMRI investigation using data from 30 diverse word and sentence comprehension experiments (481 unique participants [female and male], 678 scanning sessions). In line with prior findings, the MD network was active during many language tasks. Moreover, similar to the language-selective network, which is robustly lateralized to the left hemisphere, these responses were stronger in the left-hemisphere MD regions. However, in contrast with the language-selective network, the MD network responded more strongly (1) to lists of unconnected words than to sentences, and (2) in paradigms with an explicit task compared with passive comprehension paradigms. Indeed, many passive comprehension tasks failed to elicit a response above the fixation baseline in the MD network, in contrast to strong responses in the language-selective network. Together, these results argue against a role for the MD network in core aspects of sentence comprehension, such as inhibiting irrelevant meanings or parses, keeping intermediate representations active in working memory, or predicting upcoming words or structures. These results align with recent evidence of relatively poor tracking of the linguistic signal by the MD regions during naturalistic comprehension, and instead suggest that the MD network's engagement during language processing reflects effort associated with extraneous task demands.

Key words: domain-general network; fMRI; language network

Significance Statement

Domain-general executive processes, such as working memory and cognitive control, have long been implicated in language comprehension, including in neuroimaging studies that have reported activation in domain-general multiple demand (MD) regions for linguistic manipulations. However, much prior evidence has come from paradigms where language interpretation is accompanied by extraneous tasks. Using a large fMRI dataset (30 experiments/481 participants/678 sessions), we demonstrate that MD regions are engaged during language comprehension in the presence of task demands, but not during passive reading/listening, conditions that strongly activate the frontotemporal language network. These results present a fundamental challenge to proposals whereby linguistic computations, such as inhibiting irrelevant meanings, keeping representations active in working memory, or predicting upcoming elements, draw on domain-general executive resources.

Received Aug. 20, 2019; revised Mar. 2, 2020; accepted Apr. 5, 2020.

Author contributions: E.D., I.B., M.S., J.A., and E.F. performed research; E.D., I.B., M.S., J.A., and E.F. analyzed data; E.D. and E.F. wrote the first draft of the paper; E.D., I.B., M.S., J.A., and E.F. edited the paper; E.D. and E.F. wrote the paper; E.F. designed research.

*E.D., I.B., and M.S. contributed equally to this work.

E.F. was supported by National Institutes of Health Awards R00-HD057522, R01-DC016607, and R01-DC016950, Simons Center for the Social Brain at Massachusetts Institute of Technology Simons Foundation Grant, and Department of Brain and Cognitive Sciences and the McGovern Institute for Brain Research, Massachusetts Institute of Technology. We thank the Athinoula A. Martinos Imaging Center (McGovern Institute for Brain Research, Massachusetts Institute of Technology) and its support team, especially Steve Shannon and Atsushi Takahashi; former and current EvLab members, especially Zach Mineroff,

Brianna Pritchett, Dima Ayyash, and Zuzanna Balewski, for help with data collection; Alfonso Nieto-Castanón for help with the comparison between two preprocessing and analysis pipelines; Cory Shain and Ted Gibson for comments on earlier versions of the manuscript; and Swathi Kiran for discussions of the role of the MD network in language recovery in aphasia.

The authors declare no competing financial interests.

Correspondence should be addressed to Evgeniia Diachek at ediachek@mit.edu or Evelina Fedorenko at evelina9@mit.edu.

<https://doi.org/10.1523/JNEUROSCI.2036-19.2020>

Copyright © 2020 the authors

Introduction

Converging evidence from neuroimaging and patient studies suggests that a left-lateralized frontotemporal brain network is selective for language processing. These regions respond to linguistic input (visual or auditory) across diverse materials and tasks (e.g., Fedorenko et al., 2010, 2016; Vagharchakian et al., 2012; Scott et al., 2017; Deniz et al., 2019), but not to nonlinguistic cognitive tasks, such as arithmetic calculations, executive function tasks, music perception, action/gesture observation, and nonverbal social information (e.g., Fedorenko et al., 2011; Monti et al., 2012; Pritchett et al., 2018; Jouravlev et al., 2019; Paunov et al., 2019; for review, see Fedorenko and Varley, 2016).

In addition to this “core” language-selective network, numerous fMRI language studies have reported activation in what appear-to-be regions of a different network: a domain-general bilateral network of frontal, parietal, cingular, and opercular brain regions known as the multiple demand (MD) network (Duncan, 2010, 2013). This network supports diverse cognitive tasks (Duncan and Owen, 2000; Fedorenko et al., 2013; Hugdahl et al., 2015) and has been linked to constructs, such as working memory (WM), cognitive control, and goal-directed behavior (Cole and Schneider, 2007; Duncan, 2010). The MD network is dissociated from the language network (for review, see Fedorenko and Blank, 2020), as evidenced by brain imaging studies (Fedorenko et al., 2011; Blank et al., 2014; Mineroff et al., 2018), patient investigations (Woolgar et al., 2018), and gene expression patterns (Kong et al., 2020). Therefore, the two networks likely serve separable computational goals. However, many complex cognitive processes may rely on multiple distinct, and possibly interacting, cognitive mechanisms and their associated brain networks (Petersen and Sporns, 2015). Language comprehension may thus be supported by both the language-selective network and the domain-general MD network.

MD network engagement has been reported for diverse linguistic phenomena, including lexical/structural/referential ambiguity (e.g., Rodd et al., 2005; Novais-Santos et al., 2007; January et al., 2009; McMillan et al., 2013), high surprisal (e.g., Strijkers et al., 2019; compare Shain et al., 2020), including grammatical violations (e.g., Kuperberg et al., 2003; Nieuwland et al., 2012; Mollica et al., 2020), and syntactic complexity in unambiguous structures (e.g., Peelle et al., 2010). These results align with behavioral evidence for the role of WM/cognitive control in language comprehension (e.g., King and Just, 1991; Gernsbacher, 1993; Waters and Caplan, 1996; Gibson, 1998; Gordon et al., 2002; Fedorenko et al., 2006, 2007; Lewis et al., 2006; Novick et al., 2009). Some have therefore proposed that domain-general executive resources, implemented in the MD network, support core aspects of linguistic interpretation related to lexical access, syntactic parsing, or semantic composition (Hasson et al., 2018), such as inhibiting irrelevant meanings/parses (Novick et al., 2005), selecting the relevant representation from among alternatives (Thompson-Schill et al., 2002; Hirshorn and Thompson-Schill, 2006; Grindrod et al., 2008), supporting predictive coding for language processing (Strijkers et al., 2019), or keeping linguistic representations active in WM (Moser et al., 2007).

Others, however, have questioned the importance of domain-general executive resources/the MD network to language processing (for review, see Fedorenko, 2014; Campbell and Tyler, 2018). For example, Wright et al. (2011) showed that some frontal regions, plausibly MD areas, are only engaged during a lexical decision task, but not passive listening to the same materials. And Blank and Fedorenko (2017) demonstrated that MD regions do not closely track the linguistic signal during comprehension

of naturalistic stories, suggesting that they are unlikely to support computations that relate to the properties of the input (see also Wehbe et al., 2020).

To illuminate the role of the MD network in language processing, we conducted a large-scale investigation of diverse comprehension tasks. In particular, we used data from 30 fMRI experiments to examine the responses of MD and language regions, functionally defined in each participant using independent localizer paradigms (Fedorenko et al., 2010, 2013), to different linguistic stimuli and tasks. To foreshadow the key results, we found above-baseline responses in the MD network during many linguistic tasks. However, passive comprehension tasks, which robustly engage the language-selective network, elicited a response at the level of the fixation baseline in the MD network. These results argue against the role of the MD network in core aspects of sentence comprehension, at least across the materials tested here and in neurotypical young adults.

Materials and Methods

Because prior literature has not delivered a clear answer as to the role of the MD network (also sometimes referred to as the “executive/cognitive control network” or “task positive network”) in language comprehension, we here combined data from numerous diverse word and sentence comprehension experiments that have been conducted in our laboratory over the last decade. Given that each participant performed functional localizer tasks (e.g., Saxe et al., 2006) for the MD (and language) network, we could straightforwardly combine data from across experiments by pooling responses from functionally defined MD (or language) areas and have greater confidence that these constitute the “same” regions (i.e., functional units) across individuals compared with relying on anatomy alone (e.g., Brett et al., 2002; Saxe et al., 2006; Fedorenko and Kanwisher, 2009; Fedorenko et al., 2010; Nieto-Castañón and Fedorenko, 2012; Fedorenko and Blank, 2020). The fact that the linguistic experiments varied in the presence of an explicit task (13 passive reading/listening experiments, 17 experiments with a task), with the task further varying across experiments, allowed us to test the critical question of whether the MD network’s engagement is restricted to cases where an explicit task is present. A traditional whole-brain group-analytic approach (e.g., Holmes and Friston, 1998) would be unlikely to yield a clear answer in this study due to the combination of (1) high interindividual variability in the precise locations of functional areas and (2) the proximity of language and MD networks in the frontal and parietal cortex. These factors would lead to the blurring of language/MD network boundaries, critically undermining our ability to separate these networks in the analyses (for discussion, see, e.g., Nieto-Castañón and Fedorenko, 2012; Fedorenko and Blank, 2020). That said, as noted below, we make the individual contrast maps publicly available, allowing others to reanalyze the current dataset using any additional approaches, to complement the set of analyses reported here.

Participants

A total of 481 unique individuals (age 18–71 years, mean 26.4 years; 288 [~60%] females; for information about participants’ age, sex, and handedness, see Table SI-1 available at OSF: <https://osf.io/pdtk9/>) from the Cambridge/Boston community participated for payment across 30 fMRI language comprehension experiments, with 11–385 participants per experiment (for numbers of participants in each experiment, see Table 1; for information about participant overlap among experiments, see Table SI-4 at <https://osf.io/pdtk9/>). Each participant completed 1–14 critical experiments (median = 1), for a total of 678 critical experiment scanning sessions comprising the current dataset (see below for details). A total of 455 participants (~95%) were right-handed, as determined by the Edinburgh Handedness Inventory (Oldfield, 1971), or self-report; the remaining 26 left-handed participants showed typical left-lateralized language activations in the language localizer task (for arguments for including left-handers in cognitive neuroscience experiments, see Willems et al., 2014). A total of 402 participants (~83%) were native speakers of English; the remaining 79 participants

Table 1. Design, materials, and procedure details for Experiments 1–30

	Experiment 1	Experiment 2	Experiment 3
No. of subjects	387	12	12
Task	Visual sentence comprehension; task = passive reading + button press	Visual sentence comprehension; task = sentence rating	Visual sentence comprehension; task = sentence rating
Critical conditions	S	S	S
fMRI design	Blocked	Event-related	Blocked
Time per word (ms)	450	Whole sentence	Whole sentence
No. of words per trial	12	6–23	5–9
Trial length (s)	6	8	4
No. of trials per block	3	NA	4
No. of blocks/events per condition per run	8	52	20
Range of experimental runs	1–2	2–3	3–4
Associated publications/manuscripts	NA	Kline et al., 2018	NA
Version of the MD localizer used	Nonwords>sentences contrast of the language localizer	Spatial WM localizer	Spatial WM localizer
	Experiment 4	Experiment 5	Experiment 6
No. of subjects	16	13	13
Task	Visual sentence comprehension; task = same/different meaning judgment	Visual sentence comprehension; task = comprehension questions	Auditory sentence comprehension; task = sentence-picture matching
Critical conditions	S	S	S
fMRI design	Event-related	Event-related	Event-related
Time per word (ms)	Whole-sentence	350	variable (auditory presentation)
No. of words per trial	8–14	10–11	9
Trial length (s)	6	6	6
No. of trials per block	NA	NA	NA
No. of blocks/events per condition per run	40	60	28
Range of experimental runs	1–2	4–6	4
Associated publications/manuscripts	Siegelman et al. (2019); Fedorenko et al. (2020)	NA	Blank et al. (2016)
Version of the MD localizer used	Spatial WM localizer	Nonwords>sentences contrast of the language localizer	Spatial WM localizer
	Experiment 7	Experiment 8	Experiment 9
No. of subjects	22	12	13
Task	Visual sentence comprehension; task = memory probe	Visual sentence comprehension; task = passive reading	Visual sentence comprehension; task = plausibility judgment
Critical conditions	S	S	S
fMRI design	Event-related	Event-related	Blocked
Time per word (ms)	350	300	Whole-sentence
No. of words per trial	10	24	6–9
Trial length (s)	6	7.2	1.5
No. of trials per block	NA	NA	10
No. of blocks/events per condition per run	12	25	2
Range of experimental runs	4–5	4–6	2
Associated publications/manuscripts	Fedorenko et al. (2020)	NA	Ivanova et al. (2019)
Version of the MD localizer used	Spatial WM localizer	Spatial WM localizer	Spatial WM localizer
	Experiment 10	Experiment 11	Experiment 12
No. of subjects	13	16	19
Task	Visual sentence comprehension; task = comprehension questions	Visual sentence comprehension; task = passive reading	Visual sentence comprehension; task = passive reading
Critical conditions	S	S	S
fMRI design	Event-related	Event-related	Blocked
Time per word (ms)	Whole-sentence	Whole-sentence	Whole-sentence
No. of words per trial	8–11	4–11	11–16
Trial length (s)	6	3	3
No. of trials per block	NA	NA	6
No. of blocks/events per condition per run	60	90	2
Range of experimental runs	2–5	8–12	5
Associated publications/manuscripts	NA	Pereira et al. (2018)	Amit et al. (2017)
Version of the MD localizer used	Spatial WM localizer	Spatial WM localizer	Nonwords>sentences contrast of the language localizer
	Experiment 13	Experiment 14	Experiment 15
No. of subjects	21	12	79

(Table continues.)

Table 1. Continued

Task	Visual sentence comprehension; task = memory probe	Auditory passage comprehension; task = passive listening	Auditory passage comprehension; task = passive listening
Critical conditions	S	S	S
fMRI design	Event-related	Blocked	Blocked
Time per word (ms)	400	Variable (auditory presentation)	Variable (auditory presentation)
No. of words per trial	6	variable	variable
Trial length (s)	4	18	18
No. of trials per block	NA	1	1
No. of blocks/events per condition per run	10	8	4
Range of experimental runs	2–4	1–2	3
Associated publications/manuscripts	NA	Scott et al. (2017)	Ayyash (unpublished observation)
Version of the MD localizer used	Spatial WM localizer	Spatial WM localizer	Spatial WM localizer
	Experiment 16	Experiment 17	Experiment 18
No. of subjects	17	15	17
Task	Auditory passage comprehension; task = passive listening	Visual passage comprehension; task = inference about implied information	Visual passage comprehension; task = passive reading
Critical conditions	S	S	S
fMRI design	Blocked	Event-related	Blocked
Time per word (ms)	Variable (audit presentation)	Whole-sentence	Whole-sentence
No. of words per trial	Variable	Variable	Variable
Trial length (s)	7	8	24
No. of trials per block	3	NA	1
No. of blocks/events per condition per run	6	24	6
Range of experimental runs	4–5	4	2
Associated publications/manuscripts	Jouravlev et al. (2019)	NA	Jacoby and Fedorenko (2018)
Version of the MD localizer used	Spatial WM localizer	Spatial WM localizer	Spatial WM localizer
	Experiment 19	Experiment 20	Experiment 21
No. of subjects	17	16	16
Task	Visual sentence and word-list comprehension; task = memory probe	Visual sentence and word-list comprehension; task = passive reading	Visual sentence and word-list comprehension; task = memory probe
Critical conditions	S W	S W	S W
fMRI design	Event-related	Event-related	Blocked
Time per word (ms)	450	300	400
No. of words per trial	12	12	9
Trial length (s)	7	3.6	5
No. of trials per block	NA	NA	4
No. of blocks/events per condition per run	6–24	6	12
Range of experimental runs	3–5	3–5	2–3
Associated publications/manuscripts	Mollica et al. (2020)	Mollica et al. (unpublished observation)	NA
Version of the MD localizer used	Spatial WM localizer	Spatial WM localizer	Spatial WM localizer
	Experiment 22	Experiment 23	Experiment 24
No. of subjects	15	33	21
Task	Visual sentence and word-list comprehension; task = memory probe	Visual sentence and word-list comprehension; task = memory probe	Visual sentence and word-list comprehension; task = passive reading
Critical conditions	S	S W	S W
fMRI design	Blocked	Event-related	Blocked
Time per word (ms)	Whole-sentence	450	300
No. of words per trial	5–7	12	6–12
Trial length (s)	4	7	3.5–6.5
No. of trials per block	4	NA	4
No. of blocks/events per condition per run	4–8	5–20	6
Range of experimental runs	2–5	4–6	2–4
Associated publications/manuscripts	NA	Mollica et al. (2020)	NA
Version of the MD localizer used	Spatial WM localizer	Spatial WM localizer	Spatial WM localizer
	Experiment 25	Experiment 26	Experiment 27
No. of subjects	23	16	19
Task	Visual word comprehension; task = passive reading	Visual word comprehension; task = passive reading	Visual word comprehension; task = passive reading
Critical conditions	W	W	W
fMRI design	Blocked	Blocked	Blocked
Time per word (ms)	see Trial length	see Trial length	see Trial length

(Table continues.)

Table 1. Continued

No. of words per trial	1	1	1
Trial length (s)	800	1750	1750
No. of trials per block	25	8	8
No. of blocks/events per condition per run	6	18	18
Range of experimental runs	1–2	3–4	3–4
Associated publications/manuscripts	NA	NA	NA
Version of the MD localizer used	Spatial WM localizer	Spatial WM localizer	Spatial WM localizer
	Experiment 28	Experiment 29	Experiment 30
No. of subjects	13	11	30
Task	Visual word comprehension; task = semantic association	Visual word comprehension; task = semantic association	Visual word comprehension; task = semantic association
Critical conditions	W	W	W
fMRI design	Blocked	Event-related	Blocked
Time per word (ms)	see Trial length	see Trial length	see Trial length
No. of words per trial	4	3–5	1
Trial length (s)	5	4	2
No. of trials per block	4	NA	3
No. of blocks/events per condition per run	8	72	24
Range of experimental runs	2	3–4	1–3
Associated publications/manuscripts	Chai et al. (2016)	NA	NA
Version of the MD localizer used	Nonwords > sentences contrast of the language localizer	Spatial WM localizer	Spatial WM localizer

were native speakers of diverse languages and fluent speakers of English (for these participants, we examined responses to language processing in their native language; data from Ayyash et al., unpublished observation). All participants gave informed consent in accordance with the requirements of Massachusetts Institute of Technology's Committee on the Use of Humans as Experimental Subjects.

Experimental design

Design, stimuli, and procedure. In describing the dataset in more detail, it is helpful to define a few terms. A critical experiment dataset is a set of functional runs for a single participant for a critical experiment (total number of critical experiments = 30). A (scanning) session is a single visit of a participant to the MRI facility, during which one or more experiments are run. A critical experiment session is a session that contains one or more critical experiment datasets. An MD localizer session is a session that contains data for an MD localizer (one of two versions, as detailed below). A language localizer session is a session that contains data for a language localizer.

We have 939 critical experiment datasets (Table 1) across 678 critical experiment sessions. For 26 of the 30 experiments (507 of 939 critical experiment datasets), we functionally identified the MD network using a spatial WM localizer described below (e.g., Blank et al., 2014). For the remaining 4 experiments (432 of 939 critical experiment datasets), we used another difficulty manipulation based on a contrast between the reading of nonwords and the reading of sentences, as in Fedorenko et al. (2013). Furthermore, for the 26 experiments that used the spatial WM MD localizer, in 307 of the 507 critical experiment datasets, the MD localizer was administered in the same scanning session as the critical experiment; in the remaining 200 critical experiment datasets, the MD localizer came from an earlier scanning session (the activation patterns are highly stable within and across scanning sessions) (Assem et al., 2019; Shashidhara et al., 2020; E.F., unpublished data). Similarly, for the 4 experiments that used the *nonwords > sentences* MD localizer contrast, in 418 of the 432 critical experiment datasets, the MD localizer was administered in the same scanning session as the critical experiment; in the remaining 14 critical experiment datasets, the MD localizer came from an earlier scanning session.

All participants further completed a language localizer task (Fedorenko et al., 2010). The language functional regions of interest (fROIs) were used in some control analyses, as detailed below. One version of the language localizer served as one of the critical language experiments given that it included a passive sentence comprehension condition. In 748 of 939 critical experiment datasets, the language

localizer was administered in the same scanning session as the critical experiment; in the remaining 191 critical experiment datasets, the language localizer came from an earlier scanning session (the activation patterns are highly stable within and across scanning sessions) (Mahowald and Fedorenko, 2016). Most scanning sessions lasted ~2 h and included one or more other tasks for unrelated studies.

MD localizer. For 26 of 30 critical experiments (507 of 939 critical experiment datasets), regions of the MD network were localized using a spatial WM task contrasting a harder condition with an easier condition (e.g., Fedorenko et al., 2011, 2013; Blank et al., 2014). On each trial (8 s), participants saw a fixation cross for 500 ms, followed by a 3×4 grid within which randomly generated locations were sequentially flashed (1 s per flash) two at a time for a total of eight locations (hard condition) or one at a time for a total of four locations (easy condition). Then, participants indicated their memory for these locations in a two-alternative, forced-choice paradigm via a button press (the choices were presented for 1000 ms, and participants had up to 3 s to respond). Feedback, in the form of a green checkmark (correct responses) or a red cross (incorrect responses), was provided for 250 ms, with fixation presented for the remainder of the trial. Hard and easy conditions were presented in a standard blocked design (4 trials in a 32 s block, 6 blocks per condition per run) with a counterbalanced order across runs. Each run included four blocks of fixation (16 s each) and lasted a total of 448 s (Fig. 1). The *hard > easy* contrast targets brain regions engaged in cognitively demanding tasks. Fedorenko et al. (2013) have established that the regions activated by this task are also activated by a wide range of other demanding tasks (see also Duncan and Owen, 2000; Hugdahl et al., 2015). For the remaining 4 critical experiments in which not every participant performed the spatial WM task (432 of 939 of the critical experiment datasets), we used the *nonwords > sentences* contrast of the language localizer task (described below) to define the MD fROIs (Fedorenko et al., 2013).

Language localizer (used in some control analyses, and as one of the critical experiments). This task is described in detail in Fedorenko et al. (2010). Briefly, participants read sentences and lists of unconnected, pronounceable nonwords in a blocked design. Stimuli were presented one word/nonword at a time. Each of the 481 unique participants completed one or more language localizer sessions ($n = 423$ completed a single localizer session; $n = 46$ completed 2 sessions; $n = 8$ completed three sessions; and $n = 4$ completed four sessions), for a total of 555 language localizer sessions included in the analyses. Across this dataset, five slightly different versions of the language localizer were used (Table 2). For 71 language localizer sessions, each trial ended with a memory probe and participants had to indicate, via a button press, whether or not that

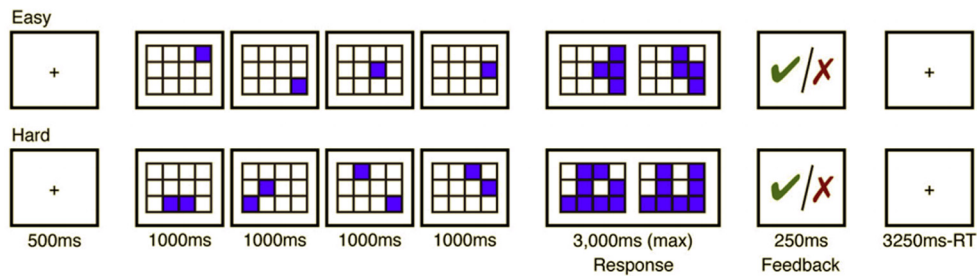


Figure 1. Procedure and timing for the spatial WM task used to localize the MD fROIs.

Table 2. Timing parameters for the different versions of the language localizer task

	Localizer 1	Localizer 2	Localizer 3	Localizer 4	Localizer 5
No. of unique scanning sessions	40	3	4	24	484
IPS	189	198	198	198	179
Conditions	Sentences (S), Nonwords (N)	Sentences (S), Word list (W), Nonwords (N)	Sentences (S), Word list (W), Nonwords (N)	Sentences (S), Word list (W), Nonwords (N)	Sentences (S), Nonwords (N)
Task	Memory probe	Memory probe	Memory probe	Memory probe	Button press
Materials	12 words/nonwords	12 words/nonwords; words - morphologic complexity manipulation	12 words/nonwords; words - morphologic complexity manipulation, sentences - content manipulation	12 words/nonwords	12 words/nonwords
Experiment block duration (s)	18	18	18	18	18
Trials per block	3	3	3	3	3
Trial duration (s)	6	6	6	6	6
Trial structure	300 ms trial-initial fixation; 12 words/nonwords presented for 350 ms each; 1000 ms probe; 500 ms trial-final fixation	300 ms trial-initial fixation; 12 words/nonwords presented for 350 ms each; 1000 ms probe; 500 ms trial-final fixation	300 ms trial-initial fixation; 12 words/nonwords presented for 350 ms each; 1000 ms probe; 500 ms trial-final fixation	300 ms trial-initial fixation; 12 words/nonwords presented for 350 ms each; 1000 ms probe; 500 ms trial-final fixation	100 ms trial-initial fixation; 12 words/nonwords presented for 450 ms each; 400 ms hand icon; 100 ms trial-final fixation
Experiment blocks per run	16	18	18	18	16
Experiment blocks per condition per run	8	6	6	6	8
Fix block duration (s)	18	18	18	18	14
Fix blocks per run	5	4	4	4	5
Run duration (s)	378	396	396	396	358

probe had appeared in the preceding sentence/nonword sequence. In the remaining 484 localizer sessions, participants read the materials passively and performed a simple button-press task at the end of each trial (included to help participants remain alert). The language localizer has been shown to be robust to changes in the materials, modality of presentation, and task (Fedorenko et al., 2010; Fedorenko, 2014; Scott et al., 2017; Ivanova et al., 2019).

Critical experiments

To broadly evaluate the role of the MD network in language comprehension, we examined neural responses across 30 diverse experiments conducted in the E.F. laboratory between 2010 and 2019, which included word-level and sentence-level materials. Details of all the experiments are reported in Table 1, but we here summarize the general approach to the selection of experimental conditions and the key dimensions of variation present across the experiments.

Each of the 30 experiments was originally designed to evaluate a specific hypothesis about (1) the sensitivity of the language and/or the MD network to some linguistic (lexical, syntactic, semantic, or pragmatic) manipulation, or (2) the selectivity of the two networks for versus non-linguistic conditions. For example, Experiment 2 compared responses to one-liner versus closely matched nonjoke controls (Kline et al., 2018). Experiment 6 compared responses to syntactically versus more complex sentences (sentences containing subject- vs object-extracted relative clauses) (Blank et al., 2016). Experiment 16 compared responses to spoken linguistic versus speech-accompanying gestures (Jouravlev et al., 2019). Experiment 24 contrasted sentences that contained a temporary

syntactic versus control unambiguous sentences (following the design of Snijders et al., 2009). Data from some of these experiments have been published or are reported in preprints and papers under review (Table 1); other experiments are parts of ongoing projects and have not yet been reported anywhere (we make the data used in the analyses below, estimates of neural responses to the conditions of the critical and localizer experiments in the MD and language fROIs in each participant [Table SI-2], as well as whole-brain activation maps for the critical and localizer contrasts, available at <https://osf.io/pdtk9/>; raw data are available from the senior author on request). For the purposes of this study, in each experiment, we (1) selected only the conditions where participants were asked to read or listen to words/word lists or plausible well-formed sentences (we excluded conditions that, e.g., contained syntactic violations), and, where necessary, (2) averaged the responses across the fine-grained linguistic manipulations to derive a single response magnitude for (a) word comprehension and/or (b) sentence comprehension.

Eighteen experiments involved sentence comprehension (3 of these involved passages, 14 unconnected sentences, and 1 both passages and unconnected sentences), 6 involved word-level comprehension, and the remaining 6 contained both sentence materials and matched word-list conditions. In 26 experiments, linguistic materials were presented visually, and in the remaining 4 auditorily. Critically, for the research question asked here, the experiments varied in the task used: in 13 experiments, participants read or listened to the materials passively (sometimes accompanied by a simple button-press task); and in the remaining 17 experiments, they were asked to perform a task (a memory probe task in 6 experiments, a semantic association task in 3

experiments, a sentence rating task in 2 experiments, a comprehension-question task in 2 experiments, a meaning similarity judgment task in 1 experiment, an inference task in 1 experiment, a plausibility judgment task in 1 experiment, and a sentence-picture matching task in 1 experiment).

To summarize some of the procedural/timing details (provided in Table 1), 16 experiments used a blocked design, and the other 14 an event-related design. In blocked design experiments, participants saw or heard between 4 and 72 blocks per condition (each between 8.5 and 26 s in duration). “Condition” here is the overarching sentence-comprehension or word-comprehension condition; so, for example, if an experiment had two conditions (syntactically easy and syntactically more complex sentences), we here report the number of blocks across the two conditions, given that we average the responses between those two conditions in the analyses, as described above. In event-related design experiments, participants saw or heard between 18 and 1080 trials per condition (each between 3 and 8 s in duration). The materials for all experiments are available from the senior author on request (those that come from published studies are typically available on the associated OSF pages, as indicated in the relevant publications).

Data acquisition, preprocessing, and first-level modeling

Data acquisition. Whole-brain structural and functional data were collected on a whole-body 3 Tesla Siemens Trio scanner with a 32-channel head coil at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at Massachusetts Institute of Technology. T1-weighted structural images were collected in 176 axial slices with 1 mm isotropic voxels (TR = 2530 ms; TE = 3.48 ms). Functional, BOLD data were acquired using an EPI sequence with a 90° flip angle and using GRAPPA with an acceleration factor of 2; the following parameters were used: 31 4.4-mm-thick near-axial slices acquired in an interleaved order (with 10% distance factor), with an in-plane resolution of 2.1 mm × 2.1 mm, FOV in the phase encoding (A ≫ P) direction 200 mm and matrix size 96 × 96 voxels, TR = 2000 ms and TE = 30 ms. The first 10 s of each run was excluded to allow for steady-state magnetization.

Preprocessing. Data preprocessing was conducted with SPM5 (using default parameters, unless specified otherwise) and supporting, custom MATLAB scripts (available from the senior author on request). We used an older version of SPM here because the current dataset is part of a larger dataset with ~900 participants and over 2000 scanning sessions, and we wanted to keep the preprocessing and first-level modeling consistent across that larger dataset; but see below. Preprocessing of anatomic data included normalization into a common space (MNI template), resampling into 2 mm isotropic voxels. Preprocessing of functional data included motion correction (realignment to the mean image using second-degree b-spline interpolation), normalization (estimated for the mean image using trilinear interpolation), resampling into 2 mm isotropic voxels, smoothing with a 4 mm FWHM Gaussian filter, and high-pass filtering at 200 s.

First-level modeling. For both the MD localizer task and the critical tasks, a standard mass univariate analysis was performed in SPM5, separately for each participant, whereby a GLM estimated, for each voxel, the effect size of each condition in each experimental run. These effects were each modeled with a boxcar function (representing entire blocks/events) convolved with the canonical HRF. The model also included first-order temporal derivatives of these effects, as well as nuisance regressors representing entire experimental runs and offline-estimated motion parameters.

To ensure that the results are robust to the version of SPM used for preprocessing and first-level modeling, we reanalyzed 4 of the 30 experiments in SPM12. Although the mean responses in the MD fROIs to the critical conditions differed slightly between the two pipelines, they (1) were overall very similar and (2) did not exhibit any systematic bias, which could have affected any of the critical comparisons (see Fig. SI-4 at <https://osf.io/pdtk9/>, showing a direct pipeline comparison for these experiments).

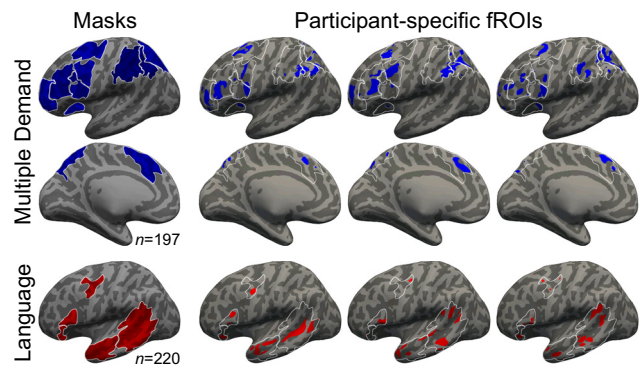


Figure 2. Masks and subject-specific fROIs. Data are shown for the MD network (top, middle; blue) and language network (bottom; red), and are approximate projections from functional volumes onto the cortical surface of an inflated average brain in common space. Only the left hemisphere is shown. The leftmost column shows masks derived from a group-level representation of data for the MD localizer contrast (Hard > Easy) and the language localizer contrast (Sentences > Nonwords), in an independent group of subjects, using watershed parcellation. These masks were used to constrain the selection of subject-specific fROIs. The other columns show approximate locations of example MD and language fROIs from 3 subjects. Apparent overlap across MD and language fROIs within an individual is illusory and due to projection onto the cortical surface. Because data were analyzed in volume (not surface) form, some parts of a given fROI that appear discontinuous in the figure (e.g., separated by a sulcus) are contiguous in volumetric space. White contours represent the borders of the masks.

Definition of the MD fROIs

For each critical experiment dataset of each participant, we defined a set of MD functional ROIs using group-constrained, subject-specific localization (Fedorenko et al., 2010). In particular, as described above, for 26 of the 30 experiments (507 of 939 critical experiment datasets), we used the spatial WM MD localizer. Each individual map for the *hard* > *easy* spatial WM contrast was intersected with a set of 20 binary masks. These masks (Fig. 2; masks available for download from <https://osf.io/pdtk9/>) were derived from a probabilistic activation overlap map for the same contrast in a large set of participants ($n = 197$) using watershed parcellation, as described by Fedorenko et al. (2010), and corresponded to relatively large areas within which most participants showed activity for the target contrast. These masks covered the frontoparietal MD network (including what some treat as a separate, cingulo-opercular, subnetwork) (e.g., Power et al., 2011), and were highly overlapping with a set of anatomic masks used by Fedorenko et al. (2013). For the remaining 4 experiments (432 of 939 critical experiment datasets), we used a contrast between the reading of nonwords and the reading of sentences. Each individual map for the *nonwords* > *sentences* contrast was intersected with the same 20 masks (for evidence that this contrast yields similar activations to more typical executive function tasks, see Fedorenko et al., 2013). Within each mask, a participant-specific MD fROI was defined as the top 10% of voxels with the highest t values for the localizer contrast (*hard* > *easy* or *nonwords* > *sentences*).

In all the critical analyses reported here, we treat the MD fROIs as a functionally integrated system given that prior work has established that these regions not only share functional profiles, but also that the MD regions' time courses are strongly correlated during both rest and task performance (e.g., Blank et al., 2014; Paunov et al., 2019), and the effect sizes in task-based paradigms correlate strongly across participants (Mineroff et al., 2018; Assem et al., 2019). However, we acknowledge the possibility that subdivisions may exist within this network (e.g., Blank et al., 2014; Paunov et al., 2019). And treating this network as an integrated system need not imply that all of its regions are identical in their response patterns and functions.

Definition of the language (fROIs) (for control analyses)

To define the language fROIs, each individual map for the *sentences* > *nonwords* contrast from the language localizer was intersected with a set of five binary masks. These masks (Fig. 2; masks available for download

from <https://osf.io/pdtk9/>) were derived from a probabilistic activation overlap map for the language localizer contrast in a large set of participants ($n = 220$), following the method described by Fedorenko et al. (2010) for a smaller set of participants. These masks covered the frontotemporal language network in the left hemisphere.

Validation of the MD fROIs

To ensure that the MD fROIs behave as expected (i.e., show a reliably greater response to the hard spatial WM condition compared with the easy one, or a greater response to the nonwords condition compared with the sentences condition), we used an across-runs cross-validation procedure (e.g., Nieto-Castañón and Fedorenko, 2012). For the 26 experiments in which all participants completed the spatial WM MD localizer task, we identified the unique participants that completed two runs of the MD localizer, leaving us with 172 sessions for the cross-validation analysis. Similarly, for the 4 experiments where we used the *nonwords* > *sentences* MD localizer task, we identified the unique participants that completed two runs of the MD localizer, leaving us with 366 sessions for the cross-validation analysis. In this analysis, the first run of the localizer was used to define the fROIs, and the second run to estimate the responses (in percent BOLD signal change [PSC]) to the localizer conditions, ensuring independence (e.g., Kriegeskorte et al., 2009); then the second run was used to define the fROIs, and the first run to estimate the responses; finally, the extracted magnitudes were averaged across the two runs to derive a single response magnitude for each of the localizer conditions (see Table SI-3a at <https://osf.io/pdtk9/>; for the cross-validated responses to the language localizer conditions in the language fROIs, see Table SI-3b). Statistical analyses were performed on these extracted PSC values (the MATLAB code is available at <https://osf.io/pdtk9/>). For the 26 MD localizer sessions that only contained a single run of the spatial WM task, we used visual examination of whole-brain activation maps for the *hard* > *easy* contrast, to ensure that the expected pattern of activation is observed.

Statistical analysis

To estimate the responses in the MD fROIs to the conditions of the critical experiments, the data from all the runs of the MD localizer were used to define the fROIs, and the responses to each condition (sentence comprehension and/or word comprehension) were then estimated in these regions, and, in some cases, averaged across conditions to derive a single response magnitude for sentence comprehension and/or word comprehension, as described above. Statistical analyses were then performed on these extracted PSC values (see Table SI-2 and the associated R code at <https://osf.io/pdtk9/>).

To characterize the role of the MD network in language comprehension, we ran several linear mixed-effect models using the “lme4” package in R with p value approximation performed by the “lmerTest” package (Bates et al., 2015; Kuznetsova et al., 2017). In particular, we asked four questions. First, we asked whether, across experiments, the MD network is engaged (above the low-level fixation baseline) during language comprehension. Second, we asked whether the MD network, like the language network, shows stronger responses to language processing in the left compared with the right hemisphere. Third, we compared the MD network’s responses to sentences versus words/word lists. One robust signature of the language network is a stronger response to sentences compared with word lists (e.g., Snijders et al., 2009; Fedorenko et al., 2010, 2016; Pallier et al., 2011), presumably because processing sentences requires additional computations compared with processing individual word meanings (for a recent discussion, see, e.g., Mollica et al., 2020). We wanted to test whether the MD network shows a similar preference for sentences. Finally and critically, we asked whether the MD network’s engagement is stronger for experiments that had included an explicit task, compared with the ones where participants passively read or listened to stimuli.

Is the MD network engaged during language comprehension?

$$\text{Effect size} \sim \text{condition} + (1 + \text{condition}|ID) + (1 + \text{condition}|ROI) \\ + (1 + \text{condition}|experiment)$$

We fit a linear mixed-effect regression model, predicting the level of BOLD response in the MD fROIs across the 30 experiments. The model

included a fixed effect for condition (sentences vs words/word lists); the difference between these levels was not of interest to this particular question, and was included for appropriately modeling variance in the data. In addition, it included random intercepts and slopes for condition by participant, fROI, and experiment.

Does the MD network show left-lateralized responses?

$$\text{Effect size} \sim \text{hemisphere} + (1 + \text{hemisphere}|ID) \\ + (1 + \text{hemisphere}|ROI) + (1 + \text{hemisphere}|experiment)$$

We fit a linear mixed-effect regression model, predicting the level of BOLD response in the MD fROIs across experiments, separately for sentence conditions and word conditions. (We tested the model for sentence and word conditions separately because the hemisphere \times condition interaction was significant in a combined model.) The model included a fixed effect for hemisphere and random intercepts and slopes for hemisphere by participant, fROI, and experiment. The mean difference between the fixed effects for the two hemispheres was tested against zero using the “glht” command in “multcomp” package (Hothorn et al., 2008) in R.

Additionally, we performed the same analysis for the language network fROIs, which are expected to show robust left lateralization (e.g., Mahowald and Fedorenko, 2016).

It is worth noting that directly comparing parameter estimates from the homologous regions in the two hemispheres is potentially problematic given the interhemispheric differences in vascularization and hemodynamic response properties (e.g., Miezin et al., 2000; Handwerker et al., 2004; Hedna et al., 2013; compare Taylor et al., 2018). In particular, the use of a canonical HRF in modeling neural responses (albeit with time derivatives) may affect the degree of model fit in the two hemispheres, which would contribute to differences in the strengths of response, thus complicating interpretation. However, because no clear alternatives exist for comparing activity between the hemispheres, we chose to include these analyses.

Does the MD network respond differentially to versus words/word lists?

$$\text{Effect size} \sim \text{condition} + (1 + \text{condition}|ID) + (1 + \text{condition}|ROI) \\ + (1 + \text{condition}|experiment)$$

We fit a linear mixed-effect regression model, predicting the level of BOLD response in the MD fROIs across the 30 experiments. The model included a fixed effect for condition (sentences vs words/word lists), and random intercepts and slopes for condition by participant, fROI, and experiment. The mean difference between the fixed effects for versus words/word lists was tested against zero using the “glht” command in “multcomp” package (Hothorn et al., 2008) in R.

Additionally, we performed the same analysis for the language network fROIs, which are expected to show a robust *sentences* > *word lists* effect (e.g., Snijders et al., 2009; Fedorenko et al., 2010, 2016; Pallier et al., 2011).

Does the MD network respond differentially to language comprehension depending on whether an explicit task is used?

$$\text{Effect size} \sim \text{task} + (1|ID) + (1 + \text{task}|ROI) + (1|experiment)$$

We fit a linear mixed-effect regression model, predicting the level of BOLD response in the MD fROIs across the 30 experiments. The model included a fixed effect for the type of task that participants had to perform (passive reading/listening vs an active task), random intercepts by participant and experiment, as well as a random intercept and slope for task by fROI. The mean difference between the fixed effects for versus passive task was tested against zero using the “glht” command in “multcomp” package (Hothorn et al., 2008) in R.

Additionally, we performed the same analysis for the language network fROIs. Whether/how language regions are modulated by the presence of a task is debated (e.g., Roskies et al., 2001; Noesselt et al., 2003; Andoh and Paus, 2011), so we took an opportunity to use this rich dataset to shed light on this question.

Results

Validation of the MD fROIs

As expected, and replicating prior work (e.g., Fedorenko et al., 2013; Blank et al., 2014; Mineroff et al., 2018; Assem et al., 2019), each of the MD fROIs showed a highly robust *hard* > *easy* effect (all $t_{(171)} > 18.7$; $p < 10^{-72}$, FDR-corrected for the 20 ROIs; Cohen $d > 0.65$, based on a dependent-samples t test). Similarly, for the participants for whom the *nonwords* > *sentences* contrast was used to define the MD fROIs, each of the fROIs showed a robust *nonwords* > *sentences* effect ($t_{(365)} > 12$; $p < 10^{-68}$, FDR-corrected for the 20 ROIs; Cohen $d > 0.41$, based on a dependent-samples t test).

Critical results

Replicating numerous prior studies that have reported activation within the MD network for linguistic manipulations (e.g., Kuperberg et al., 2003; Rodd et al., 2005; Novais-Santos et al., 2007; January et al., 2009; Peelle et al., 2010; Nieuwland et al., 2012; McMillan et al., 2013; Mollica et al., 2020), we found that, across experiments, language comprehension tasks elicited an above baseline response in the MD network (Fig. 3) (sentences: $b = 0.27$, $SE = 0.09$, $z = 2.91$, $p = 0.003$; words: $b = 0.41$, $SE = 0.07$, $z = 6.32$, $p < 10^{-9}$). Additionally, we found that the MD fROIs in the left hemisphere responded more strongly than the MD fROIs in the right hemisphere, for both sentence and word-level comprehension (Fig. 3) (sentences: $b = 0.20$, $SE = 0.04$, $z = 4.72$, $p < 10^{-5}$; words: $b = 0.18$, $SE = 0.06$, $z = 2.98$, $p = 0.003$). As expected, this pattern was also robustly present in the language network (sentences: $b = 0.62$, $SE = 0.14$, $z = 4.45$, $p < 10^{-5}$; words: $b = 0.40$, $SE = 0.09$, $z = 4.23$, $p < 10^{-4}$).

However, in contrast to the language network, which responds more strongly during sentence comprehension compared with the processing of unconnected lists of words (e.g., Fedorenko et al., 2010), an effect we replicated here ($b = 0.37$, $SE = 0.10$, $z = 3.51$, $p = 0.0004$), the MD network showed the opposite pattern, with a stronger response to words/word lists than sentences ($b = 0.15$, $SE = 0.06$, $z = 2.55$, $p = 0.011$; Fig. 4; for responses of individual fROIs, see Fig. SI-1).

Critically, we also found a strong effect of task, such that responses in the MD fROIs were stronger in the experiments with an explicit task than in the passive reading/listening paradigms ($b = 0.56$, $SE = 0.14$, $z = 4$, $p < 10^{-4}$; Fig. 5; for responses of individual fROIs, see Fig. SI-2). Indeed, some passive reading/listening experiments elicited a response at or below the fixation baseline in the MD network (Fig. SI-3). In contrast, in the language fROIs, the task did not affect the responses ($b = -0.18$, $SE = 0.14$, $z = -1.27$, $p = 0.203$), with robust responses elicited by both experiments with an explicit task ($b = 0.71$, $SE = 0.15$, $t_{(18,85)} = 4.6$, $p < 10^{-3}$) and passive reading/listening paradigms ($b = 0.89$, $SE = 0.16$, $t_{(22,38)} = 5.5$, $p < 10^{-4}$). If anything, the latter elicited numerically stronger responses.

Discussion

Across 30 fMRI language comprehension experiments (481 participants, 678 sessions), we examined how the brain regions of the domain-general MD network (Duncan, 2010, 2013), which

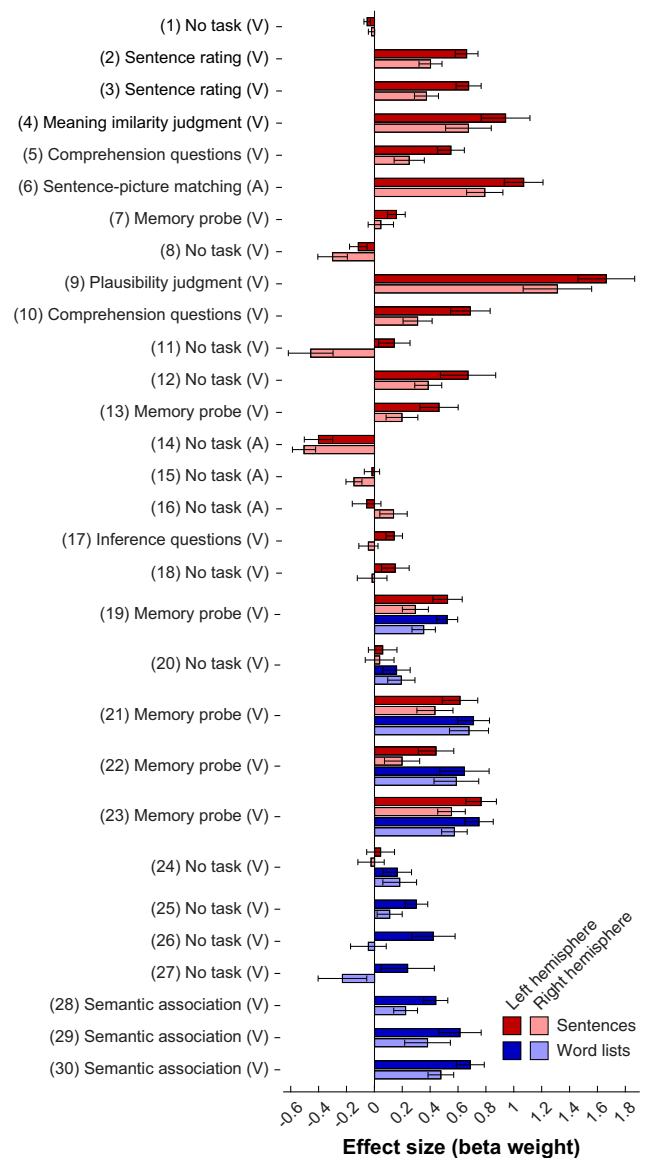


Figure 3. Responses to sentences and word lists in the MD network across experiments, and their laterality. Responses (β weights for the corresponding regressors in the GLM) are shown averaged across fROIs in the left (darker colors) and right (brighter colors) hemispheres, separately for sentence (red shades) and word list (blue shades) conditions. Responses are measured as β weights for the corresponding condition regressors in the GLM. Bars represent the average response across subjects. Error bars indicate SEM across subjects. Most experiments include either sentences or word lists, except for Experiments 19–24.

have been linked to executive demands, respond to language processing. Consistent with prior work, we found above-baseline MD responses during many linguistic tasks. Moreover, these responses were stronger in the left hemisphere, mirroring the lateralization observed in the frontotemporal language-selective network (Mahowald and Fedorenko, 2016). However, in sharp contrast to the language-selective network, which responds more strongly when participants process structured and meaningful stimuli (sentences) compared with individual words or lists of unconnected words, the MD network exhibited the opposite preference. This result already puts into question the key role of the MD network in any combinatorial linguistic operations, related to syntactic parsing or semantic composition, because one would expect brain regions that support such operations to

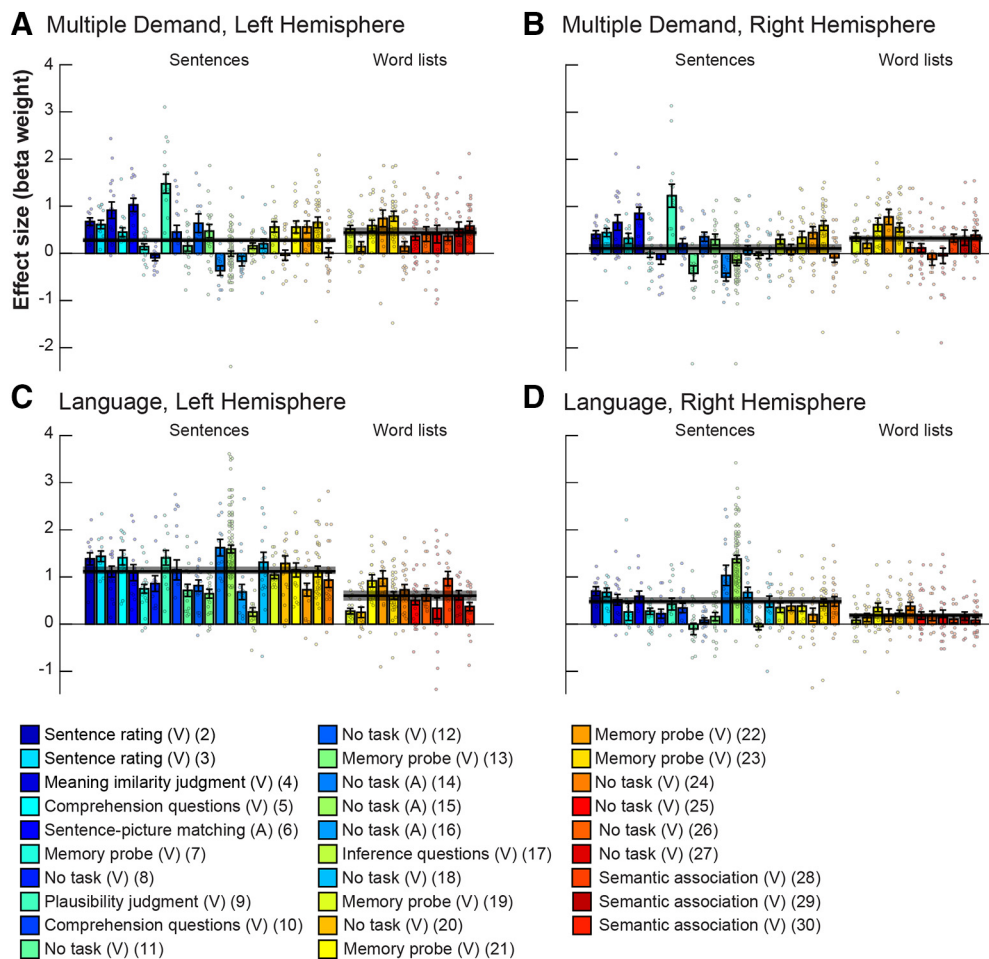


Figure 4. Responses of each network and hemisphere to the sentences and words/word lists conditions across experiments. Responses (β weights for the corresponding regressors in the GLM) are shown averaged across fROIs in the MD (**A, B**) and language (**C, D**) networks, separately for the left hemisphere (**A, C**) and right hemisphere (**B, D**). Data are presented for each of 29 experiments. Data for Experiment 1 are not shown because the number of individual data points was too large for the plot to be legible and informative (see Fig. 3). Dots represent data for individual subjects. Bars represent the average response across subjects. Error bars indicate SEM across subjects. Thick, horizontal black lines are averaged across experiments. Gray rectangles represent the corresponding 95% CIs. Conditions from the same experiment share the same color; specifically, the six experiments that each contained both sentences and word lists conditions are presented at the end (right) of the Sentences bar group and the beginning (left) of the Word lists bar group, for ease of comparison.

respond more strongly to stimuli where those operations are engaged (i.e., sentences). But even more importantly, MD responses strongly depended on the presence of an explicit task: passive reading/listening tasks, which elicit strong responses in the language areas, elicited a much weaker response in the MD network (at the level of the fixation baseline, on average; Fig. 5A,B) compared with experiments with a task.

Why might we, *a priori*, think that the domain-general MD network is important for language comprehension? There is a long tradition in the psycholinguistic literature to describe both lexical access and syntactic/semantic parsing using domain-general cognitive constructs. These include storing information in and retrieving it from WM, updating focal attention, inhibiting irrelevant information, selecting an option among alternatives, and predictive processing (e.g., Johnson-Laird, 1983; Abney and Johnson, 1991; King and Just, 1991; Resnik, 1992; Gernsbacher, 1993; Waters and Caplan, 1996; Gibson, 1998; McElree, 2000, 2001; Gordon et al., 2002; Fedorenko et al., 2006, 2007; Lewis et al., 2006; Novick et al., 2009; Rodd et al., 2010; Schuler et al., 2010; Vergauwe et al., 2010; Smith and Levy, 2013; van Schijndel et al., 2013; Rasmussen and Schuler, 2018). These kinds of mental operations may be implemented in domain-general circuits of the MD network, which has historically been linked to diverse

executive demands (Miller and Cohen, 2001; Duncan and Owen, 2000; Duncan, 2010). Indeed, prior neuroimaging studies have attributed core linguistic computations, such as the ones above, to (parts of) the MD network (e.g., Thompson-Schill et al., 2002; Novick et al., 2005; Hirshorn and Thompson-Schill, 2006; Moser et al., 2007; Grindrod et al., 2008; January et al., 2009; Strijkers et al., 2019). However, alternatively, computations, such as inhibiting irrelevant information or predictive processing, albeit similar across domains, may be implemented in domain-specific cortices that store the relevant knowledge representations (Hasson et al., 2015).

Our results support the latter possibility and argue against the role of the MD network in core aspects of language comprehension. If a brain region supports a computation that is part and parcel of language understanding, this computation should be performed regardless of whether we are processing language passively or whether language processing is accompanied by a secondary task, such as a memory-probe or comprehension-question task. This is exactly the pattern we observe in the language-selective network, which exhibits a task-independent response profile. However, the MD network's response during passive comprehension tasks is, on average, at the level of the fixation baseline. These findings suggest that the MD network's

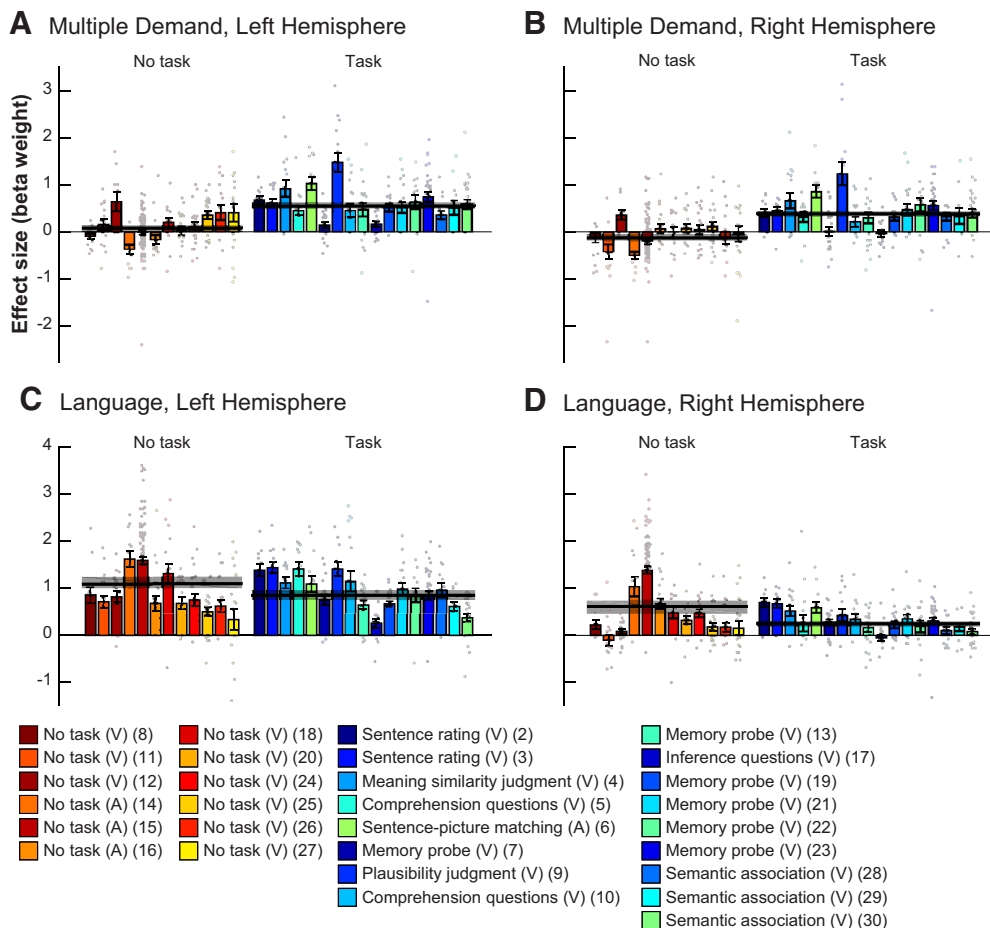


Figure 5. Responses of the MD (**A, B**) and language (**C, D**) networks for the left (**A, C**) and right (**B, D**) hemispheres to passive and task based paradigms across experiments. Same conventions as in Figure 4, but bars are now grouped by whether the experimental paradigm was passive comprehension or an active task. For experiments that contained both sentences and word lists, responses are averaged across these two conditions.

engagement likely reflects artificial task demands rather than language comprehension per se, and that all the core linguistic computations (compare section below, “Core” linguistic computations not tapped by our materials?) take place outside of MD areas, presumably in the language-selective areas (for converging evidence, see Blank and Fedorenko, 2017), which suggests that language, but not MD, regions “track” naturalistic linguistic input closely, and that the MD network’s computations are therefore unlikely to be related to the input features (Shain et al., 2020; Wehbe et al., 2020). Below, we raise four issues important to consider in light of the main conclusion we’re drawing here, that the MD network does not support core aspects of sentence comprehension.

“Core” linguistic computations not tapped by our materials?

We here construe core linguistic computations as computations that have to be engaged to extract a meaningful representation from the linguistic signal. By this definition, these operations should be engaged regardless of whether we are processing linguistic input passively, or whether we have to perform some additional task on the input. Core linguistic computations include computations related to lexical access and combinatorial processing (syntactic parsing and semantic composition), both of which strongly engage the frontotemporal language network (e.g., Fedorenko et al., 2010; Bautista and Wilson, 2016; Fedorenko and Blank, 2020). Might the materials used in the

current study, across the 30 experiments, not tap some core comprehension-related computations? Aside from the cases discussed in the section below (Noisy language comprehension?), the current set of materials is biased toward written language (26 of the 30 experiments use written materials) and does not include linguistic exchanges (e.g., dialogs or multiperson conversations). We therefore leave open the possibility that some linguistic operations engaged during auditory comprehension (e.g., prosody-related computations) (e.g., Kristensen et al., 2013) or during the processing of socially interactive materials may drive the MD network.

Noisy language comprehension?

The stimuli in the current study were clearly perceptible and well formed. This differs from naturalistic comprehension scenarios, which are characterized by both low-level perceptual and higher-level linguistic noise (speakers make false starts/errors, etc.). Long prominent in speech perception research (Mattys et al., 2012), noise has recently permeated models of sentence interpretation (e.g., Ferreira et al., 2002; Levy, 2008; Gibson et al., 2013; Traxler, 2014). Prior fMRI studies of acoustically (e.g., Adank, 2012; Hervais-Adelman et al., 2012; Wild et al., 2012; Scott and McGettigan, 2013; Vaden et al., 2013; Pelle, 2018) and linguistically (e.g., containing syntactic errors) (Kuperberg et al., 2003; Nieuwland et al., 2012; Mollica et al., 2020) noisy signals have reported activation in regions consistent with the topography of

the MD network. So, the MD network may be important for coping with signal corruption. This may also be the underlying cause of MD regions' responses during non-native (L2) language processing (e.g., Pliatsikas and Luk, 2016) because the representations of linguistic input are plausibly noisier in L2 speakers (Futrell and Gibson, 2017). That said, detection of errors in the input necessarily relies on the knowledge of the statistics of the relevant domain (language, in this case) and correction of noisy input relies on the knowledge of how noise operates (e.g., what kind of errors speakers are likely to make). Both of these kinds of knowledge are likely to be stored within the language-selective network, not the MD network (see also, e.g., Shain et al., 2020). Furthermore, in nonlinguistic domains, the MD network shows increased activity during any more cognitively demanding condition, not only conditions with noisy input (e.g., Duncan and Owen, 2000; Fedorenko et al., 2013; Crittenden and Duncan, 2014; Hugdahl et al., 2015). As a result, the mechanisms that support error detection and correction during language comprehension, and the MD network's contribution to processing noisy input remain to be characterized.

Other populations?

The current study focused on neurotypical young adults. However, our brains are notoriously plastic, and tissue not previously engaged in some function can assume that function in addition to its original function(s) or via repurposing (e.g., Feydy et al., 2002; Cramer, 2008; Kleim, 2011). The MD network might be especially plastic in this way, given that it flexibly supports diverse behaviors and modulates its responses based on current task demands (e.g., Freedman et al., 2001; Cromer et al., 2010; Kumano et al., 2016; Jackson et al., 2017). Recent behavioral (e.g., Martin and Allen, 2008; Corbett et al., 2009; El Hachoui et al., 2014; Bonini and Radanovic, 2015; Villard and Kiran, 2017; Wall et al., 2017; Simic et al., 2019) and neuroimaging (e.g., Brownsett et al., 2014; Geranmayeh et al., 2016, 2017; Meier et al., 2016; Sims et al., 2016; Stockert et al., 2020) studies have begun to suggest a possible role for the MD network in recovery from aphasia (for review, see Hartwigsen, 2018). Related evidence comes from increases in the MD network's activity during language processing in aging (e.g., Wingfield and Grossman, 2006). However, whether the MD engagement is functionally important, or simply reflects greater processing demands, remains to be discovered.

Language production?

The current study focused on comprehension. Might the MD network support core operations in language production? Executive processes have been implicated in both lexical access and syntactic planning based on behavioral (e.g., Alm and Nilsson, 2001; Roelofs and Piai, 2011; Strijkers et al., 2011; compare Ivanova et al., 2017), neuroimaging (e.g., Indefrey and Levelt, 2004; Shuster and Lemieux, 2005; Alario et al., 2006; Troiani et al., 2008; Eickhoff et al., 2009; Wilson et al., 2009; Adank, 2012; Geranmayeh et al., 2012; Grande et al., 2012; Heim et al., 2012), and patient (e.g., Ziegler et al., 1997; Nestor et al., 2003; Coelho et al., 2012; Endo, 2013) evidence. Although language production presumably relies on the same knowledge representations as comprehension, the computational demands differ. For example, syntactic operations are obligatory for producing correct linguistic output, but may be foregone during comprehension (Bock, 1995). In addition, production is more demanding and follows a developmental time course that resembles that of executive functions (e.g., Hartshorne and Germine, 2015). As a result, the MD network may support some aspects of

language production, although, as with comprehension, it will be important to dissociate core linguistic processes from extraneous task demands (Blanco-Elorrieta and Pykkänen, 2017).

In conclusion, we have ruled out a set of hypotheses about the contributions of the domain-general MD network to language comprehension. In particular, we showed that MD areas only respond in comprehension experiments in the presence of a secondary task. We have consequently argued that the MD network is unlikely to support core linguistic computations that relate to lexical access, syntactic parsing, or semantic composition. However, we leave open the possibilities that the MD network (1) supports linguistic computations not engaged by the current materials, including operations related to the processing of noisy linguistic input, (2) helps compensate for language loss after brain damage or in healthy aging, or (3) supports core linguistic computations during language production.

References

- Abney SP, Johnson M (1991) Memory requirements and local ambiguities of parsing strategies. *J Psycholinguist Res* 20:233–250.
- Adank P (2012) The neural bases of difficult speech comprehension and speech production: two activation likelihood estimation (ALE) meta-analyses. *Brain Lang* 122:42–54.
- Alario FX, Chainay H, Lehericy S, Cohen L (2006) The role of the supplementary motor area (SMA) in word production. *Brain Res* 1076:129–143.
- Alm H, Nilsson L (2001) The use of car phones and changes in driver behaviour. *Int J Vehicle Design* 26:4–11.
- Amit E, Hoeflin C, Hamzah N, Fedorenko E (2017) An asymmetrical relationship between verbal and visual thinking: converging evidence from behavior and fMRI. *Neuroimage* 152:619–627.
- Andoh J, Paus T (2011) Combining functional neuroimaging with off-line brain stimulation: modulation of task-related activity in language areas. *J Cogn Neurosci* 23:349–361.
- Assem M, Blank IA, Mineroff Z, Ademoglu A, Fedorenko E (2019) Activity in the Fronto-Parietal Multiple-demand Network is Robustly Associated with Individual Differences in Working Memory and Fluid Intelligence. *bioRxiv* 110270. doi: <https://doi.org/10.1101/110270>.
- Bates D, Machler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67:1–48.
- Bautista A, Wilson SM (2016) Neural responses to grammatically and lexically degraded speech. *Language, cognition and neuroscience* 31:567–574.
- Blanco-Elorrieta E, Pykkänen L (2017) Bilingual language switching in the laboratory versus in the wild: the spatiotemporal dynamics of adaptive language control. *J Neurosci* 37:9022–9036.
- Blank I, Balewski Z, Mahowald K, Fedorenko E (2016) Syntactic processing is distributed across the language system. *Neuroimage* 127:307–323.
- Blank I, Fedorenko E (2017) Domain-general brain regions do not track linguistic input as closely as language-selective regions. *J Neurosci* 37:9999–10011.
- Blank I, Kanwisher N, Fedorenko E (2014) A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *J Neurophysiol* 112:1105–1118.
- Bock K (1995) Sentence production: from mind to mouth. In: *Speech, language, and communication: handbook of perception and cognition* (Miller JL, Eimas PD, eds), pp 181–216. San Diego: Academic.
- Bonini MV, Radanovic M (2015) Cognitive deficits in post-stroke aphasia. *Arq Neuro-Psiquiatr* 73:840–847.
- Brett M, Anton JL, Valabregue R, Poline JB (2002). Region of interest analysis using an SPM toolbox. Paper presented at the 8th International Conference on Functional Mapping of the Human Brain. Sendai, Japan.
- Brownsett SL, Warren JE, Geranmayeh F, Woodhead Z, Leech R, Wise RJ (2014) Cognitive control and its impact on recovery from aphasic stroke. *Brain* 137:242–254.
- Campbell KL, Tyler LK (2018) Language-related domain-specific and domain-general systems in the human brain. *Curr Opin Behav Sci* 21:132–137.

- Chai LR, Mattar MG, Blank IA, Fedorenko E, Bassett DS (2016) Functional network dynamics of the language system. *Cereb Cortex* 26:4148–4159.
- Coelho C, Lê K, Mozeiko J, Krueger F, Grafman J (2012) Discourse production following injury to the dorsolateral prefrontal cortex. *Neuropsychologia* 50:3564–3572.
- Cole MW, Schneider W (2007) The cognitive control network: integrated cortical regions with dissociable functions. *Neuroimage* 37:343–360.
- Corbett F, Jefferies E, Ehsan S, Lambon Ralph MA (2009) Different impairments of semantic cognition in semantic dementia and semantic aphasia: evidence from the non-verbal domain. *Brain* 132:2593–2608.
- Cramer SC (2008) Repairing the human brain after stroke: I. Mechanisms of spontaneous recovery. *Ann Neurol* 63:272–287.
- Crittenden BM, Duncan J (2014) Task difficulty manipulation reveals multiple demand activity but no frontal lobe hierarchy. *Cereb Cortex* 24:532–540.
- Cromer JA, Roy JE, Miller EK (2010) Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron* 66:796–807.
- Deniz F, Nunez-Elizalde AO, Huth AG, Gallant JL (2019) The representation of semantic information across human cerebral cortex during listening versus reading is invariant to stimulus modality. *J Neurosci* 39:7722–7736.
- Duncan J (2010) The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn Sci* 14:172–179.
- Duncan J (2013) The structure of cognition: attentional episodes in mind and brain. *Neuron* 80:35–50.
- Duncan J, Owen AM (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci* 23:475–483.
- Eickhoff SB, Heim S, Zilles K, Amunts K (2009) A systems perspective on the effective connectivity of overt speech production. *Mov Disord* 24:S187.
- El Hachoui H, Visch-Brink EG, Lingsma HF, van de Sandt-Koenderman MW, Dippel DW, Koudstaal PJ, Middelkoop HA (2014) Nonlinguistic cognitive impairment in poststroke aphasia: a prospective study. *Neurorehabil Neural Repair* 28:273–281.
- Endo R (2013) Realities, rewards, and risks of heritage-language education: perspectives from Japanese immigrant parents in a Midwestern community. *Biling Res J* 36:278–294.
- Fedorenko E (2014) The role of domain-general cognitive control in language comprehension. *Front Psychol* 5:335.
- Fedorenko E, Kanwisher N (2009) Neuroimaging of language: why hasn't a clearer picture emerged? *Lang Linguistics Compass* 3:839–865.
- Fedorenko E, Varley R (2016) Language and thought are not the same thing: evidence from neuroimaging and neurological patients. *Ann NY Acad Sci* 1369:132–153.
- Fedorenko E, Blank IA (2020) Broca's area is not a natural kind. *Trends Cogn Sci* 24:270–284.
- Fedorenko E, Gibson E, Rohde D (2006) The nature of working memory capacity in sentence comprehension: evidence against domain-specific working memory resources. *J Mem Lang* 54:541–553.
- Fedorenko E, Gibson E, Rohde D (2007) The nature of working memory in linguistic, arithmetic and spatial integration processes. *J Mem Lang* 56:246–269.
- Fedorenko E, Hsieh PJ, Nieto-Castañón A, Whitfield-Gabrieli S, Kanwisher N (2010) New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *J Neurophysiol* 104:1177–1194.
- Fedorenko E, Behr MK, Kanwisher N (2011) Functional specificity for high-level linguistic processing in the human brain. *Proc Natl Acad Sci USA* 108:16428–16433.
- Fedorenko E, Duncan J, Kanwisher N (2013) Broad domain generality in focal regions of frontal and parietal cortex. *Proc Natl Acad Sci USA* 110:16616–16621.
- Fedorenko E, Blank I, Siegelman M, Mineroff Z (2020) Lack of selectivity for syntax relative to word meanings throughout the language network. *bioRxiv* 477851. doi: <https://doi.org/10.1101/477851>.
- Fedorenko E, Scott TL, Brunner P, Coon WG, Pritchett B, Schalk G, Kanwisher N (2016) Neural correlate of the construction of sentence meaning. *Proc Natl Acad Sci USA* 113:E6256–E6262.
- Ferreira F, Bailey KG, Ferraro V (2002) Good-enough representations in language comprehension. *Curr Dir Psychol Sci* 11:11–15.
- Feydy A, Carlier R, Roby-Brami A, Bussel B, Cazalis F, Pierot L, Burnod Y, Maier MA (2002) Longitudinal study of motor recovery after stroke: recruitment and focusing of brain activation. *Stroke* 33:1610–1617.
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291:312–316.
- Futrell R, Gibson E (2017) L2 processing as noisy channel language comprehension. *Bilingualism* 20:683–684.
- Geranmayeh F, Brownsett SL, Leech R, Beckmann CF, Woodhead Z, Wise RJ (2012) The contribution of the inferior parietal cortex to spoken language production. *Brain Lang* 121:47–57.
- Geranmayeh F, Leech R, Wise RJ (2016) Network dysfunction predicts speech production after left hemisphere stroke. *Neurology* 86:1296–1305.
- Geranmayeh F, Chau TW, Wise RJ, Leech R, Hampshire A (2017) Domain-general subregions of the medial prefrontal cortex contribute to recovery of language after stroke. *Brain* 140:1947–1958.
- Gernsbacher MA (1993) Less skilled readers have less efficient suppression mechanisms. Los Angeles: Sage.
- Gibson E (1998) Linguistic complexity: locality of syntactic dependencies. *Cognition* 68:1–76.
- Gibson E, Bergen L, Piantadosi ST (2013) Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proc Natl Acad Sci USA* 110:8051–8056.
- Gordon PC, Hendrick R, Levine WH (2002) Memory-load interference in syntactic processing. *Proc Psychol Sci* 13:425–430.
- Grande M, Meffert E, Schoenberger E, Jung S, Frauenrath T, Huber W, Hussmann K, Moormann M, Heim S (2012) From a concept to a word in a syntactically complete sentence: an fMRI study on spontaneous language production in an overt picture description task. *Neuroimage* 61:702–714.
- Grindrod CM, Bilenko NY, Myers EB, Blumstein SE (2008) The role of the left inferior frontal gyrus in implicit semantic competition and selection: an event-related fMRI study. *Brain Res* 1229:167–178.
- Handwerker DA, Ollinger JM, D'Esposito M (2004) Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage* 21:1639–1651.
- Hartshorne JK, Germine LT (2015) When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychol Sci* 26:433.
- Hartwigsen G (2018) Flexible redistribution in cognitive networks. *Trends Cogn Sci* 22:687–698.
- Hasson U, Chen J, Honey CJ (2015) Hierarchical process memory: memory as an integral component of information processing. *Trends Cogn Sci* 19:304–313.
- Hasson U, Egidio G, Marelli M, Willems RM (2018) Grounding the neurobiology of language in first principles: the necessity of non-language-centric explanations for language comprehension. *Cognition* 180:135–157.
- Hedna VS, Bodhit AN, Ansari S, Falchhook AD, Stead L, Heilman KM, Waters MF (2013) Hemispheric differences in ischemic stroke: is left-hemisphere stroke more common? *Clin Neurol* 9:97–102.
- Heim S, Amunts K, Hensel T, Grande M, Huber W, Binkofski F, Eickhoff SB (2012) The role of human parietal area 7A as a link between sequencing in hand actions and in overt speech production. *Front Psychol* 3:534.
- Hervais-Adelman AG, Carlyon RP, Johnsrude IS, Davis MH (2012) Brain regions recruited for the effortful comprehension of noise-vocoded words. *Lang Cogn Process* 27:1145–1166.
- Hirshorn EA, Thompson-Schill SL (2006) Role of the left inferior frontal gyrus in covert word retrieval: neural correlates of switching during verbal fluency. *Neuropsychologia* 44:2547–2557.
- Holmes A, Friston K (1998) Generalisability, random effects and population inference. *Neuroimage* 7:S754.
- Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. *Biomed J* 50:346–363.
- Hugdahl K, Raichle ME, Mitra A, Specht K (2015) On the existence of a generalized non-specific task-dependent network. *Front Hum Neurosci* 9:430.
- Indefrey P, Levelt WJ (2004) The spatial and temporal signatures of word production components. *Cognition* 92:101–144.
- Ivanova AA, Mineroff Z, Zimmerer V, Kanwisher N, Varley R, Fedorenko E (2019) The language network is recruited but not required for non-verbal semantic processing. *bioRxiv* 696484. doi: <https://doi.org/10.1101/696484>.
- Ivanova I, Ferreira VS, Gollan TH (2017) Form overrides meaning when bilinguals monitor for errors. *J Mem Lang* 94:75–102.

- Jackson J, Rich AN, Williams MA, Woolgar A (2017) Feature-selective attention in frontoparietal cortex: multivoxel codes adjust to prioritize task-relevant information. *J Cogn Neurosci* 29:310–321.
- Jacoby N, Fedorenko E (2018) Discourse-level comprehension engages medial frontal Theory of Mind brain regions even for expository texts. *Lang Cogn Neurosci*. Advance online publication. Retrieved September 26, 2018. doi: 10.1080/23273798.2018.1525494.
- January D, Trueswell JC, Thompson-Schill SL (2009) Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: implications for the neural basis of sentence processing. *J Cogn Neurosci* 21:2434–2444.
- Johnson-Laird PN (1983) *Mental models: towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard UP.
- Jouravlev O, Zheng D, Balewski Z, Le Arnz Pongos A, Levan Z, Goldin-Meadow S, Fedorenko E (2019) Speech-accompanying gestures do not engage high-level language processing brain regions. *Neuropsychologia* 132:107132.
- King J, Just MA (1991) Individual differences in syntactic processing: the role of working memory. *J Mem Lang* 30:580–602.
- Kleim JA (2011) Neural plasticity and neurorehabilitation: teaching the new brain old tricks. *J Commun Disord* 44:521–528.
- Kline M, Gallee J, Balewski Z, Fedorenko E (2018) Understanding jokes draws most heavily on the Theory of Mind brain network. *PsyArXiv*. Available at <https://psyarxiv.com/h2nyx/>.
- Kong XZ, Tzourio-Mazoyer N, Joliot M, Fedorenko E, Liu J, Fisher SE, Francks C (2020) Gene expression correlates of the cortical network underlying sentence processing. *Neurobiology of Language* 1:77–103.
- Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 12:535–540.
- Kristensen LB, Wang L, Petersson KM, Hagoort P (2013) The interface between language and attention: prosodic focus marking recruits a general attention network in spoken language comprehension. *Cereb Cortex* 23:1836–1848.
- Kumano H, Suda Y, Uka T (2016) Context-dependent accumulation of sensory evidence in the parietal cortex underlies flexible task switching. *J Neurosci* 36:12192–12202.
- Kuperberg GR, Holcomb PJ, Sitnikova T, Greve D, Dale AM, Caplan D (2003) Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *J Cogn Neurosci* 15:272–293.
- Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest package: tests in linear mixed effects models. *J Stat Softw* 82:1–26.
- Levy R (2008) A noisy-channel model of rational human sentence comprehension under uncertain input. Paper presented at the 2008 Conference on Empirical Methods in Natural Language Processing, Honolulu.
- Lewis RL, Vasishth S, Van Dyke JA (2006) Computational principles of working memory in sentence comprehension. *Trends Cogn Sci* 10:447–454.
- 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.
- Martin RC, Allen CM (2008) A disorder of executive function and its role in language processing. *Semin Speech Lang* 29:201–210.
- Mattys SL, Davis MH, Bradlow AR, Scott SK (2012) Speech recognition in adverse conditions: a review. *Lang Cogn Process* 27:953–978.
- McElree B (2000) Sentence comprehension is mediated by content-addressable memory structures. *J Psycholinguist Res* 29:111–123.
- McElree B (2001) Working memory and focal attention. *J Exp Psychol Learn Mem Cogn* 27:817–835.
- McMillan CT, Coleman D, Clark R, Liang TW, Gross RG, Grossman M (2013) Converging evidence for the processing costs associated with ambiguous quantifier comprehension. *Front Psychol* 4:153.
- Meier EL, Kapse KJ, Kiran S (2016) The relationship between frontotemporal effective connectivity during picture naming, behavior, and preserved cortical tissue in chronic aphasia. *Front Hum Neurosci* 10:109.
- Miezin FM, Maccotta L, Ollinger JM, Petersen SE, Buckner RL (2000) Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage* 11:735–759.
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202.
- Mineroff ZA, Blank I, Mahowald K, Fedorenko E (2018) A robust dissociation among the language, multiple demand, and default mode networks: evidence from inter-region correlations in effect size. *Neuroimage* 119:501–511.
- Mollica F, Siegelman M, Diachek E, Piantadosi ST, Mineroff Z, Futrell R, Fedorenko E (2020) Composition is the core driver of the language-selective network. *Neurobiology of Language* 1:104–134.
- Monti MM, Parsons LM, Osherson DN (2012) Thought beyond language: neural dissociation of algebra and natural language. *Psychol Sci* 23:914–922.
- Moser DC, Fridriksson J, Healy EW (2007) Sentence comprehension and general working memory. *Clin Linguist Phon* 21:147–156.
- Nestor PJ, Graham NL, Fryer TD, Williams GB, Patterson K, Hodges JR (2003) Progressive non-fluent aphasia is associated with hypometabolism centred on the left anterior insula. *Brain* 126:2406–2418.
- Nieto-Castañón A, Fedorenko E (2012) Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *Neuroimage* 63:1646–1669.
- Nieuwland MS, Martin AE, Carreiras M (2012) Brain regions that process case: evidence from Basque. *Hum Brain Mapp* 33:2509–2520.
- Noesselt T, Shah NJ, Jäncke L (2003) Top-down and bottom-up modulation of language related areas: an fMRI study. *BMC Neurosci* 4:13.
- Novais-Santos S, Gee J, Shah M, Troiani V, Work M, Grossman M (2007) Resolving sentence ambiguity with planning and working memory resources: evidence from fMRI. *Neuroimage* 37:361–378.
- Novick JM, Kan IP, Trueswell JC, Thompson-Schill SL (2009) A case for conflict across multiple domains: memory and language impairments following damage to ventrolateral prefrontal cortex. *Cogn Neuropsychol* 26:527–567.
- Novick JM, Trueswell JC, Thompson-Schill SL (2005) Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. *Cogn Affect Behav Neurosci* 5:263–281.
- Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh Inventory. *Neuropsychologia* 9:97–113.
- Pallier C, Devauchelle AD, Dehaene S (2011) Cortical representation of the constituent structure of sentences. *Proc Natl Acad Sci USA* 108:2522–2527.
- Paunov AM, Blank IA, Fedorenko E (2019) Functionally distinct language and Theory of Mind networks are synchronized at rest and during language comprehension. *J Neurophysiol* 121:1244–1265.
- Peelle JE (2018) Listening effort: how the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear Hear* 39:204–214.
- Peelle JE, Johnsrude I, Davis MH (2010) Hierarchical processing for speech in human auditory cortex and beyond. *Front Hum Neurosci* 4:51.
- Pereira F, Lou B, Pritchett B, Ritter S, Gershman SJ, Kanwisher N, Botvinick M, Fedorenko E (2018) Toward a universal decoder of linguistic meaning from brain activation. *Nat Commun* 9:963.
- Petersen SE, Sporns O (2015) Brain networks and cognitive architectures. *Neuron* 88:207–219.
- Pliatsikas C, Luk G (2016) Executive control in bilinguals: a concise review on fMRI studies. *Bilingualism* 19:699–705.
- Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, Petersen SE (2011) Functional network organization of the human brain. *Neuron* 72:665–678.
- Pritchett B, Hoeflin C, Koldewyn K, Dechter E, Fedorenko E (2018) High-level language processing regions are not engaged in action observation or imitation. *J Neurophysiol* 120:2555–2570.
- Rasmussen NE, Schuler W (2018) Left-corner parsing with distributed associative memory produces surprisal and locality effects. *Cogn Sci* 42:1009–1042.
- Resnik P (1992) Left-corner parsing and psychological plausibility. Paper presented at the Proceedings of the 14th Conference on Computational Linguistics, Vol 1. Nantes, France.
- Rodd JM, Davis MH, Johnsrude IS (2005) The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cereb Cortex* 15:1261–1269.
- Rodd JM, Johnsrude IS, Davis MH (2010) The role of domain-general frontal systems in language comprehension: evidence from dual-task interference and semantic ambiguity. *Brain Lang* 115:182–188.
- Roelofs A, Piai V (2011) Attention demands of spoken word planning: a review. *Front Psychol* 2:307.

- Roskies AL, Fiez JA, Balota DA, Raichle ME, Petersen SE (2001) Task-dependent modulation of regions in the left inferior frontal cortex during semantic processing. *J Cogn Neurosci* 13:829–843.
- Saxe R, Brett M, Kanwisher N (2006) Divide and conquer: a defense of functional localizers. *Neuroimage* 30:1088–1096.
- Schuler W, AbdelRahman S, Miller T, Schwartz L (2010) Broad-coverage parsing using human-like memory constraints. *Comput Linguistics* 36:1–30.
- Scott TL, Galle J, Fedorenko E (2017) A new fun and robust version of an fMRI localizer for the frontotemporal language system. *Cogn Neurosci* 8:167–110.
- Scott SK, McGettigan C (2013) The neural processing of masked speech. *Hear Res* 303:58–66.
- Shain C, Blank IA, van Schijndel M, Schuler W, Fedorenko E (2020) fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia* 138:107307.
- Shashidhara S, Spronkers FS, Erez Y (2020) Individual-subject functional localization increases univariate activation but not multivariate pattern discriminability in the ‘multiple-demand’ frontoparietal network. *J Cogn Neurosci* 28:1–21.
- Shuster LI, Lemieux SK (2005) An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain Lang* 93:20–31.
- Siegelman M, Blank IA, Mineroff Z, Fedorenko E (2019) An attempt to conceptually replicate the dissociation between syntax and semantics during sentence comprehension. *Neuroscience* 413:219–229.
- Simic T, Rochon E, Greco E, Martino R (2019) Baseline executive control ability and its relationship to language therapy improvements in post-stroke aphasia: a systematic review. *Neuropsychol Rehabil* 29:395–439.
- Sims JA, Kapse K, Glynn P, Sandberg C, Tripodis Y, Kiran S (2016) The relationships between the amount of spared tissue, percent signal change, and accuracy in semantic processing in aphasia. *Neuropsychologia* 84:113–126.
- Smith NJ, Levy R (2013) The effect of word predictability on reading time is logarithmic. *Cognition* 128:302–319.
- Snijders TM, Vosse T, Kempen G, Van Berkum JJ, Petersson KM, Hagoort P (2009) Retrieval and unification of syntactic structure in sentence comprehension: an fMRI study using word-category ambiguity. *Cereb Cortex* 19:1493–1503.
- Stockert A, Wawrzyniak M, Klingbeil J, Wrede K, Kümmerer D, Hartwigsen G, Kaller CP, Weiller C, Saur D (2020) Dynamics of language reorganization after left temporo-parietal and frontal stroke. *Brain* 143:844–861.
- Strijkers K, Holcomb PJ, Costa A (2011) Conscious intention to speak proactively facilitates lexical access during overt object naming. *J Mem Lang* 65:345–362.
- Strijkers K, Chanoine V, Munding D, Dubarry AS, Trébuchon A, Badier JM, Alario FX (2019) Grammatical class modulates the (left) inferior frontal gyrus within 100 milliseconds when syntactic context is predictive. *Sci Rep* 9:4830.
- Taylor AJ, Kim JH, Ress D (2018) Characterization of the hemodynamic response function across the majority of human cerebral cortex. *Neuroimage* 173:322–331.
- Thompson-Schill SL, Jonides J, Marshuetz C, Smith EE, D’Esposito M, Kan IP, Knight RT, Swick D (2002) Effects of frontal lobe damage on interference effects in working memory. *Cogn Affect Behav Neurosci* 2:109–120.
- Traxler MJ (2014) Trends in syntactic parsing: anticipation, Bayesian estimation, and good-enough parsing. *Trends Cogn Sci* 18:605–611.
- Trojani V, Fernández-Seara MA, Wang Z, Detre JA, Ash S, Grossman M (2008) Narrative speech production: an fMRI study using continuous arterial spin labeling. *Neuroimage* 40:932–939.
- Vaden KI, Kuchinsky SE, Cute SL, Ahlstrom JB, Dubno JR, Eckert MA (2013) The cingulo-opercular network provides word-recognition benefit. *J Neurosci* 33:18979–18986.
- Vagharchakian L, Dehaene-Lambertz G, Pallier C, Dehaene S (2012) A temporal bottleneck in the language comprehension network. *J Neurosci* 32:9089–9102.
- Van Schijndel M, Exley A, Schuler W (2013) A model of language processing as hierarchic sequential prediction. *Top Cogn Sci* 5:522–540.
- Vergauwe E, Barrouillet P, Camos V (2010) Do mental processes share a domain-general resource? *Psychol Sci* 21:384–390.
- Villard S, Kiran S (2017) To what extent does attention underlie language in aphasia? *Aphasiology* 31:1226–1245.
- Wall KJ, Cumming TB, Copland DA (2017) Determining the association between language and cognitive tests in poststroke aphasia. *Front Neurol* 8:149.
- Waters GS, Caplan D (1996) The measurement of verbal working memory capacity and its relation to reading comprehension. *Q J Exp Psychol A* 49:51–79.
- Webbe L, Blank IA, Shain C, Futrell R, Levy R, von der Malsburg T, Smith N, Gibson E, Fedorenko E (2020) Incremental language comprehension difficulty predicts activity in the language network but not the multiple demand network. *bioRxiv*. doi: <https://doi.org/10.1101/2020.04.15.043844>.
- Wild CJ, Yusuf A, Wilson DE, Peelle JE, Davis MH, Johnsrude IS (2012) Effortful listening: the processing of degraded speech depends critically on attention. *J Neurosci* 32:14010–14021.
- Willems RM, Van der Haegen L, Fisher SE, Francks C (2014) On the other hand: including left-handers in cognitive neuroscience and neurogenetics. *Nat Rev Neurosci* 15:193–201.
- Wilson SM, Isenberg AL, Hickok G (2009) Neural correlates of word production stages delineated by parametric modulation of psycholinguistic variables. *Hum Brain Mapp* 30:3596–3608.
- Wingfield A, Grossman M (2006) Language and the aging brain: patterns of neural compensation revealed by functional brain imaging. *J Neurophysiol* 96:2830–2839.
- Woolgar A, Duncan J, Manes F, Fedorenko E (2018) Fluid intelligence is supported by the multiple-demand system not the language system. *Nat Hum Behav* 2:200–204.
- Wright P, Randall B, Marslen-Wilson WD, Tyler LK (2011) Dissociating linguistic and task-related activity in the left inferior frontal gyrus. *J Cogn Neurosci* 23:404–413.
- Ziegler JC, Montant M, Jacobs AM (1997) The feedback consistency effect in lexical decision and naming. *J Mem Lang* 37:533–554.