Title

The domain-general multiple demand (MD) network does not support core aspects of language comprehension: a large-scale fMRI investigation

Abbreviated title

MD network does not support language comprehension

Authors

Evgeniia Diachek*[1], Idan Blank*[3,4], Matthew Siegelman*[2], and Evelina Fedorenko[3,5,6]

Affiliations

[1]Department of Psychology, Vanderbilt University, Nashville, TN 37203, USA
[2]Department of Psychology, Columbia University, New York, NY 10027, USA
[3]Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139, USA
[4]Department of Psychology, UCLA, Los Angeles, CA 90095, USA
[5]McGovern Institute for Brain Research, MIT, Cambridge, MA 02139, USA
[6]Department of Psychiatry, MGH, Charlestown, MA 02129, USA

*Equal contributors

Corresponding Authors

Evgeniia Diachek or Ev Fedorenko
ediachek@mit.edu/evelina9@mit.edu; 43 Vassar Street, Room 46-3037, Cambridge, MA, 02139

Materials included

40 pages | 5 figures | 2 tables

Word count

Abstract: 250 | Significance Statement: 120 | Intro: 650 | Discussion: 1,500

Acknowledgements

We would like to acknowledge the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT, and its support team (Steve Shannon, and Atsushi Takahashi). We thank former and current EvLab members (especially Zach Mineroff, Brianna Pritchett, Dima Ayyash, and Zuzanna Balewski) for their help with data collection, Cory Shain and Ted Gibson for helpful comments on earlier versions of the manuscript, and Swathi Kiran for discussions of the role of the MD network in language recovery in aphasia. EF was supported by NIH awards R00-HD057522, R01-DC016607, and R01-DC016950, and a grant from the Simons Foundation to the Simons Center for the Social Brain at MIT.

Conflict of interest

The authors declare no competing financial interests.

Abstract (250/250 words)

Aside from the core language-specific left-lateralized fronto-temporal network, language comprehension sometimes additionally recruits a domain-general bilateral fronto-parietal 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, we conducted a large-scale fMRI investigation using data from 30 diverse word and sentence comprehension experiments (481 unique participants, 678 scanning sessions). In line with prior findings, the MD network was active during many language tasks. Moreover, similar to the language-specific network, which is robustly lateralized to the left hemisphere, these responses were stronger in the left-hemisphere MD regions. However, in stark contrast with the language-specific network, the MD network responded more strongly (i) to lists of unconnected words than to sentences, and critically, (ii) in paradigms with an explicit task compared to passive comprehension paradigms. In fact, 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-specific network. In tandem, these results argue against a role for the MD network in core aspects of sentence comprehension like 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.

Significance Statement (120/120 words)

Domain-general executive processes, like 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 elicit strong responses in the core fronto-temporal language network. These results present a fundamental challenge to proposals whereby linguistic computations, like inhibiting irrelevant meanings, keeping representations active in working memory, or predicting upcoming elements, draw on domain-general executive resources.

Introduction (650/650 words)

Converging evidence from neuroimaging and patient studies suggests that a left fronto-temporal brain network is selective for language processing. These regions respond to linguistic input (visual or auditory) across diverse materials and tasks (Buchweitz et al., 2009; Fedorenko et al., 2010, 2016; Vagharchakian et al., 2012; Scott et al., 2017**)**, but not to non-linguistic cognitive tasks, like arithmetic calculations, executive function tasks, music perception, action/gesture observation, and non-verbal social information (Fedorenko et al., 2011; Monti et al., 2012; Pritchett et al., 2018; Jouravlev et al., 2019; Paunov, 2019; see Fedorenko & Varley, 2016, for a review).

In addition to this "core" language 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 & Owen, 2000; Fedorenko et al., 2013) and has been linked to constructs like working memory, cognitive control, and goal-directed behavior (Cole & Shneider, 2007; Duncan, 2010). The MD network is dissociated from the language network, 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., 2018). 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 & Sporns, 2015). Language comprehension may thus be supported by both the language-specific network and the domain-general MD network.

MD network engagement has been reported for diverse linguistic phenomena, including lexical/structural/referential ambiguity (Rodd et al., 2005; Novais-Santos et al., 2007; January et al., 2009; McMillan et al., 2013), high surprisal (Strijkers et al., 2019; cf. Shain et al., 2019) including grammatical violations (Kuperberg et al., 2003; Nieuwland et al., 2012), and syntactic complexity in unambiguous structures (Peelle et al., 2010). These results align with behavioral evidence for the role of working memory/cognitive control in language comprehension (King & 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* (Hasson et al., 2018), like inhibiting irrelevant meanings/parses (Novick et al., 2005), selecting the relevant representation from among alternatives (Thompson-Schill et al., 2002; Hirshorn & Thompson-Schill, 2006; Grindrod et al., 2008), supporting predictive coding for language processing (Strijkers et al., 2019), or keeping linguistic representations active in working memory (Moser et al., 2007).

Others, however, have questioned the importance of domain-general executive resources / the MD network to language processing (for reviews, see Fedorenko, 2014; Campbell & 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 & 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., submitted).

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 thirty fMRI experiments to examine the responses of language and MD 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-specific network, failed to elicit a response in the MD network. These results argue against the role of the MD network in core aspects of sentence comprehension.

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 lab 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 to relying on anatomy alone (e.g., Brett et al., 2002; Saxe et al., 2006; Fedorenko & Kanwisher, 2009; Fedorenko et al., 2010; Nieto-Castañon & Fedorenko, 2012; Fedorenko & Blank, submitted). 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.

Participants

Four hundred and eighty-one unique individuals (age 18-71, mean 26.4; 288 (~60%) females; see **Table SI-4** available at OSF: https://osf.io/pdtk9/ for information about participants' age, sex, and handedness) from the Cambridge/Boston, MA community participated for payment across 30 fMRI language comprehension experiments, with 11-385 participants per experiment (see **Table 1** for numbers of participants in each experiment; see **Table SI-1** available at OSF: https://osf.io/pdtk9/ for information about participant overlap among experiments). 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). Four hundred and fifty-five 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 (see Willems et al., 2014, for arguments for including left-handers in cognitive neuroscience experiments). Four hundred and two participants (~83%) were native speakers of English; the remaining 79 participants 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., in prep.). All participants gave informed consent in accordance with the requirements of MIT's Committee on the Use of Humans as Experimental Subjects (COUHES).

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 (see **Table 1**) across 678 critical experiment sessions. For 26 of the 30 experiments (507/939 critical experiment datasets), we functionally identified the MD network using a spatial working memory (WM) localizer described below (e.g., Blank et al., 2014). For the remaining 4 experiments (432/939 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., 2017, unpublished data from the Fedorenko lab). 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 748/939 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 & Fedorenko, 2016).

Most scanning sessions lasted approximately 2 hours and included one or more other tasks for unrelated studies.

MD localizer. For 26/30 critical experiments (507/939 critical experiment datasets), regions of the MD network were localized using a spatial working memory (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 1,000 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 4 blocks of fixation (16 s each) and lasted a total of 448 s (**Figure 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 & Owen, 2000; Hugdahl et al., 2015). For the remaining 4 critical experiments in which not every participant performed the spatial WM task (432/939 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 3 sessions; and n=4 completed 4 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 (see **Table 2** for details). 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 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 in order 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; Siegelman et al., in prep.).

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 Fedorenko lab between 2010 and 2018, 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 (i) the sensitivity of the language and/or the MD network to some linguistic (lexical, syntactic, semantic, or pragmatic) manipulation, or (ii) the selectivity of the two networks for linguistic vs. non-linguistic conditions. For example, Experiment 2 compared responses to one-liner jokes vs. closely matched non-joke controls (Kline et al., submitted); Experiment 16 compared responses to spoken linguistic materials vs. speech-accompanying gestures (Jouravlev et al., 2019); and Experiment 24 contrasted sentences that contained a temporary syntactic ambiguity vs. control unambiguous sentences (following the design of Snijders et al., 2009). Data from some of these experiments have been published or are reported in papers under review (see **Table 1**); other experiments are parts of ongoing projects and have not yet been reported anywhere (we make all the data reported here available on OSF: https://osf.io/pdtk9/). For the purposes of this study, in each experiment, we (i) 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, (ii) 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), six involved word-level comprehension, and the remaining six 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). (Note that "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 1,080 trials per condition (each between 3 and 8 s in duration). The materials for all experiments are available from the authors upon 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 MIT. T1-weighted structural images were collected in 176 axial slices with 1 mm isotropic voxels (repetition time (TR) = 2,530 ms; echo time (TE) = 3.48 ms). Functional, blood oxygenation level-dependent (BOLD) data were acquired using an EPI sequence with a 90^{o} flip angle and using GRAPPA with an acceleration factor of 2; the following parameters were used: thirty-one 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 were excluded to allow for steady state magnetization.

Preprocessing. Data preprocessing was carried out with SPM5 (using default parameters, unless specified otherwise) and supporting, custom MATLAB scripts. (Note that preprocessing and basic modeling have not changed much in the later versions of SPM, as confirmed by direct comparisons on several datasets performed in our lab; we chose to use the older version here because some of the datasets were collected and analyzed many years ago, and we wanted to have all the data analyzed through the same pipeline, for better comparability.) Preprocessing of anatomical data included normalization into a common space (Montreal Neurological Institute (MNI) template), resampling into 2 mm isotropic voxels. Preprocessing of functional data included motion correction (realignment to the mean image using 2nd-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 general linear model (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 Hemodynamic Response Function (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.

Definition of the MD functional regions of interest (fROIs)

For each critical experiment dataset of each participant, we defined a set of multiple demand (MD) functional ROIs using group-constrained, subject-specific localization (Fedorenko et al., 2010). In particular, as described above, for 26 of the 30 experiments (507/939 critical experiment datasets), we used the spatial working memory (WM) MD localizer. Each individual map for the *hard* > *easy* spatial WM contrast was intersected with a set of twenty binary masks. These masks (**Figure 2**; available for download from OSF: 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 in Fedorenko et al. (2010), and corresponded to relatively large areas within which most participants showed activity for the target contrast. These masks covered the fronto-parietal MD network (including what some treat as a separate, cingulo-opercular, sub-network; e.g., Power et al., 2011), and were highly overlapping with a set of anatomical masks used in Fedorenko et al. (2013). For the remaining 4 experiments (432/939 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 twenty masks (see Fedorenko et al., 2013, for evidence that this contrast yields similar activations to more typical executive function tasks). 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 (Assem et al., 2017; Mineroff et al., 2018). 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 (**Figure 2**; available for download from OSF: 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 in Fedorenko et al. (2010) for a smaller set of participants. These masks covered the fronto-temporal language network in the left hemisphere (**Table SI-3b**; available for download at OSF: https://osf.io/pdtk9/).

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 to the easy one, or a greater response to the nonwords condition compared to the sentences condition), we used an across-runs cross-validation procedure (e.g., Nieto-Castañon & 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 273 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 (**Table SI-3a**; available for download at OSF: https://osf.io/pdtk9/). 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. Statistical analyses were performed on these extracted PSC values. 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.

Critical analyses

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 (S) and/or Word comprehension (W)) 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 (**Table SI-2**; available for download from OSF: https://osf.io/pdtk9/).

To characterize the role of the MD network in language comprehension, we ran several linear mixed-effects 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 to the right hemisphere. Third, we compared the MD network's responses to sentences vs. lists of unconnected words. One robust signature of the language network is a stronger response to sentences compared to word lists (e.g., Snijders et al., 2009; Fedorenko et al., 2010; Pallier et al., 2010; Fedorenko et al., 2016), presumably because processing sentences requires additional computations compared to processing individual word meanings. 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 to the ones where participants passively read or listened to stimuli.

**** Is the MD network engaged during language comprehension?***

Effect size ~ condition + (1+condition|ID) + (1+condition|ROI) + (1+condition|experiment)

We fit a linear mixed effects regression model, predicting the level of BOLD response in the MD fROIs across the thirty experiments. The model included a fixed effect for condition (sentences vs. 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?***

Effect size ~ hemisphere + (1+hemisphere|ID) + (1+hemisphere|ROI) + (1+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×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 & Fedorenko, 2016).

**** Does the MD network respond differentially to sentences vs. lists of unconnected words?***

Effect size ~ condition + (1+condition|ID) + (1+condition|ROI) + (1+condition|experiment)

We fit a linear mixed effect regression model, predicting the level of BOLD response in the MD fROIs across the thirty experiments. The model included a fixed effect for condition (sentences vs. word lists), and random intercepts and slopes for condition by participant, fROI, and experiment. The mean difference between the fixed effects for sentences vs. 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; Pallier et al., 2010; Fedorenko et al., 2016).

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

Effect size ~ task + (1|ID) + (1+task|ROI) + (1|experiment)

We fit a linear mixed effect regression model, predicting the level of BOLD response in the MD fROIs across the thirty 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 active vs. 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 & 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; Assem et al., 2017; Mineroff et al., 2018), each of the MD fROIs showed a highly robust *hard* > *easy* effect (all $ts_{(271)}>18.7$; $ps<10^{-72}$, FDR-corrected for the twenty ROIs; Cohen *ds* > 0.65, based on a conservative independent-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 ($ts_{(365)}>12$; $ps<10^{-68}$, FDR-corrected for the twenty ROIs; Cohen *ds* > 0.41, based on a conservative independent-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), we found that – across experiments – language comprehension tasks elicited an above baseline response in the MD network (**Figure 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 (**Figure 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 to 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 lists of words than sentences (**Figure 4**) ($b = 0.15$, $SE = 0.06$, $z = 2.55$, $p = 0.011$).

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 ($b = 0.56$, $SE = 0.14$, $z = 4$, $p < 10^{-4}$). In fact, some passive reading/listening experiments elicited a response at or below the fixation baseline in the MD network (**Figure 5)**. In contrast, in the language fROIs, the task did not affect the responses ($b = -.18$, $SE = 0.14$, $z = -1.27$, $p = 0.203$), with robust responses elicited by both experiments with an explicit task and passive reading/listening paradigms (passive reading/listening: $b = 0.89$, $SE = 0.16$, $t_{(22.38)} = 5.5$, $p < 10^{-4}$; explicit task: $b = 0.71$, $SE = 0.15$, $t_{(18.85)} = 4.6$, $p < 10^{-3}$).

Discussion (1,500/1,500 words)

Across 30 fMRI language comprehension experiments (481 participants, 678 sessions), we examined how the regions of the domain-general Multiple Demand (MD) network (Duncan, 2010, 2013)—which are 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 fronto-temporal language-selective network (Mahowald & Fedorenko, 2016). However, in sharp contrast to the language-selective network, which responds more strongly when participants process structured and meaningful stimuli (sentences) compared to lists of unconnected words, the MD network exhibited the opposite preference. And most importantly, MD responses strongly depended on the presence of an explicit task, with passive reading/listening tasks—which elicit strong responses in the language areas—failing to elicit an above-baseline response in the MD network.

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 working memory, updating focal attention, inhibiting irrelevant information, selecting an option among alternatives, and predictive processing (e.g., Johnson-Laird & Nicholas, 1983; Abney & Johnson, 1991; King & Just, 1991; Resnick, 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 & Levy, 2013; van Schijndel et al., 2013; Rasmussen & 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 & Cohen, 2001; Duncan & Owen, 2000; Duncan, 2010). Indeed, prior neuroimaging studies have attributed core linguistic computations, like the ones above, to (parts of) the MD network (Thompson-Schill et al., 2002; Novick et al., 2005; Hirshorn & Thompson-Schill, 2006; Moser et al., 2007; Grindrod et al., 2008; January et al., 2009; Strijkers et al., 2019). However, alternatively, computations like 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, like a memory 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 many passive comprehension tasks does not differ from the fixation baseline. These findings suggest that the MD network's engagement reflects artificial task demands rather than language comprehension per se, and that all the core linguistic computations take place outside of MD areas, presumably in the language-selective areas (see Blank & Fedorenko, 2017, for converging evidence, 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; also Shain et al., 2019). 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.

1. Transient MD responses?

If the MD network is only active *transiently*, when a particular linguistic phenomenon is encountered (e.g., a low-frequency word/construction or a difficult temporary ambiguity), perhaps we are not picking up these responses because our neural measure (the BOLD signal) is diffused in time and includes many time-points that do not contain the relevant phenomena. This possibility is unlikely for two reasons. Conceptually, an architecture where the same computation (related to lexical retrieval or parsing) relies on the core language network up to a certain difficulty threshold and on a different network after that threshold seems implausible and unwieldy. And empirically, the linguistic materials in the current study, including in the passive-reading/listening tasks, included phenomena that cause comprehension difficulty (e.g., difficult temporary ambiguity in Experiment 24, infrequent words in Experiment 25, and non-local dependencies in Experiments 5-6). If the MD level of response during language processing when these kinds of phenomena are absent is ~0, then we should still see above-baseline MD responses for these tasks because the presence of linguistic complexity in some conditions should lead to increased activity. We don't see this pattern.

2. Noisy language comprehension?

The stimuli in this 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 (Levy et al., 2008; Gibson et al., 2013; Traxler, 2014). Prior fMRI studies of acoustically (Adank, 2012; Hervais-Adelman et al., 2012; Wild et al., 2012; Scott & McGettigan, 2013; Vaden et al., 2013; Peelle, 2018) and linguistically (e.g., containing syntactic errors; Kuperberg et al., 2003; Nieuwland et al., 2012) 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 (Pliatsikas & Luk, 2016) because the representations of linguistic input are plausibly noisier in L2 speakers (Futrell & Gibson, 2017). However, in non-linguistic domains, the MD network responds more during *any* more cognitively demanding condition, not only conditions with noisy input (e.g., Duncan & Owen, 2000; Crittenden & Duncan, 2012; Fedorenko et al., 2013; Hugdahl et al., 2015). As a result, the nature of the MD network's contribution to processing noisy input remains unclear.

3. 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 (Alm and Nilsson, 2001; Roelofs and Piai, 2011; Strijkers et al., 2011, cf. Ivanova & Ferreira, 2017), neuroimaging (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 (Ziegler et al., 1997; Nestor et al., 2003; Coelho et al., 2012; Endo et al., 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 (Hartshorne & 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-Elorietta & Pylkkanen, 2017).

4. Recovery in aphasia?

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 (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 (Freedman et al., 2001; Cromer et al., 2010; Jackson et al., 2016; Kumano et al., 2016). Recent behavioral (Martin & Allen, 2008; Corbett et al., 2009; El Hachioui et al., 2014; Bonini & Radanovic, 2015; Villard & Kiran, 2016; Simic et al., 2017; Wall et al., 2017) and neuroimaging (Brownsett et al., 2014; Geranmayeh et al., 2016, 2017; Sims et al., 2016; Meier et al., 2016) studies have begun to suggest a possible role for the MD network in recovery from aphasia (see Hartwigsen, 2018, for a review). Related evidence comes from increases in the MD network's activity during language processing in aging (e.g., Wingfield & Grossman, 2006). However, whether or not the MD engagement is functionally important (cf. simply reflecting greater processing demands) remains to be discovered.

To conclude, 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 possibility that the MD network (i) plays a role in processing noisy linguistic input, (ii) supports core linguistic computations during language production, or (iii) helps compensate for language loss after brain damage or in healthy aging.

Author contributions

EF designed research. All authors performed research and analyzed data. IB created the figures. ED and EF wrote the paper with input from other authors.

Tables

Table 1 Design, materials, and procedure details for Experiments 1-30

	Experiment 1	**Experiment 2**	**Experiment 3**
Number 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	450 ms	whole-sentence	whole-sentence
Number of Words per Trial	12	6-23	5-9
Trial Length	6 s	8 s	4 s
Number of Trials per Block	3	N/A	4
Number of Blocks/Events per Condition per Run	8	52	20
Range of Experimental Runs	1-2	2-3	3-4
Associated Publications/Manuscripts	N/A	Kline et al., submitted https://psyarxiv.com/h2nyx/	N/A
Version of the MD Localizer Used	Nonwords>sentences contrast of the language localizer	Spatial WM localizer	Spatial WM localizer

Table 1 (continued)

	Experiment 4	**Experiment 5**	**Experiment 6**
Number 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	whole-sentence	350 ms	variable (audit presentation)
Number of Words per Trial	8-14	10-11	9
Trial Length	6 s	6 s	6 s

Number of Trials per Block	N/A	N/A	N/A
Number 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., submitted https://www.biorxiv.org/content/10.1101/477851v1.article-info	N/A	Blank et al., 2016
Version of the MD Localizer Used	Spatial WM localizer	Nonwords>sentences contrast of the language localizer	Spatial WM localizer

Table 1 (continued)

	Experiment 7	**Experiment 8**	**Experiment 9**
Number 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	350 ms	300 ms	whole-sentence
Number of Words per Trial	10	24	6-9
Trial Length	6 s	7.2 s	1.5 s
Number of Trials per Block	N/A	N/A	10
Number 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., submitted https://www.biorxiv.org/content/10.1101/	N/A	Ivanova et al., submitted https://www.biorxiv.org/content/10.1101/696484v1

	477851v1.article-info		
Version of the MD Localizer Used	Spatial WM localizer	Spatial WM localizer	Spatial WM localizer

Table 1 (continued)

	Experiment 10	**Experiment 11**	**Experiment 12**
Number 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	whole-sentence	whole-sentence	whole-sentence
Number of Words per Trial	8-11	4-11	11-16
Trial Length	6 s	3 s	3 s
Number of Trials per Block	N/A	N/A	6
Number of Blocks/Events per Condition per Run	60	90	2
Range of Experimental Runs	2-5	8-12	5
Associated Publications/Manuscripts	N/A	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

Table 1 (continued)

	Experiment 13	**Experiment 14**	**Experiment 15**
Number of Subjects	21	12	79
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	400 ms	variable (audit presentation)	variable(audit presentation)
Number of Words per Trial	6	variable	variable
Trial Length	4 s	18 s	18 s
Number of Trials per Block	N/A	1	1
Number of Blocks/Events per Condition per Run	10	8	4
Range of Experimental Runs	2-4	1-2	3
Associated Publications/Manuscripts	N/A	Scott et al., 2017	Ayyash et al., in prep.
Version of the MD Localizer Used	Spatial WM localizer	Spatial WM localizer	Spatial WM localizer

Table 1 (continued)

	Experiment 16	**Experiment 17**	**Experiment 18**
Number 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 with a warning about a subsequent memory task
Critical Conditions	S	S	S
fMRI Design	Blocked	Event-related	Blocked
Time per Word	variable (audit presentation)	whole-sentence	whole-sentence
Number of Words per Trial	variable	variable	variable
Trial Length	7 s	8 s	24 s
Number of Trials per Block	3	N/A	1
Number 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	N/A	Jacoby&Fedorenko, 2018

Version of the MD Localizer Used	Spatial WM localizer	Spatial WM localizer	Spatial WM localizer

Table 1 (continued)

	Experiment 19	Experiment 20	Experiment 21
Number 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	450 ms	300 ms	400 ms
Number of Words per Trial	12	12	9
Trial Length	7 s	3.6 s	5 s
Number of Trials per Block	N/A	N/A	4
Number 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., submitted https://www.biorxiv.org/content/10.1101/436204v1	Mollica et al., in prep.	N/A
Version of the MD Localizer Used	Spatial WM localizer	Spatial WM localizer	Spatial WM localizer

Table 1 (continued)

	Experiment 22	Experiment 23	Experiment 24
Number 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	whole-sentence	450 ms	300 ms

Number of Words per Trial	5-7	12	6-12
Trial Length	4 s	7 s	3.5-6.5 s
Number of Trials per Block	4	N/A	4
Number 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	N/A	Mollica et al., submitted https://www.biorxiv.org/content/10.1101/436204v1	N/A
Version of the MD Localizer Used	Spatial WM localizer	Spatial WM localizer	Spatial WM localizer

Table 1 (continued)

	Experiment 25	**Experiment 26**	**Experiment 27**
Number 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	see Trial Length	see Trial Length	see Trial Length
Number of Words per Trial	1	1	1
Trial Length	800 ms	1750 ms	1750 ms
Number of Trials per Block	25	8	8
Number of Blocks/Events per Condition per Run	6	18	18
Range of Experimental Runs	1-2	3-4	3-4
Associated Publications/Manuscripts	N/A	N/A	N/A
Version of the MD Localizer Used	Spatial WM localizer	Spatial WM localizer	Spatial WM localizer

Table 1 (continued)

	Experiment 28	**Experiment 29**	**Experiment 30**
Number 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	see Trial Length	see Trial Length	see Trial Length
Number of Words per Trial	4	3-5	1
Trial Length	5 s	4 s	2 s
Number of Trials per Block	4	N/A	3
Number 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	N/A	N/A
Version of the MD Localizer Used	Nonwords>sentences contrast of the language localizer	Spatial WM localizer	Spatial WM localizer

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

TypeName	Localizer 1	Localizer 2	Localizer 3	Localizer 4	Localizer 5
Number of Unique Scanning Sessions	40	3	4	24	484
IPS	189	198	198	198	179
Conditions	Sentences (S), Nonwords (N)	Sentences (S), Wordlist (W), Nonwords (N)	Sentences (S), Wordlist (W), Nonwords (N)	Sentences (S), Wordlist (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 - morphological complexity manipulation	12 words/nonwords; words - morphological complexity manipulation, sentences - content manipulation	12 words/ nonwords	12 words/ nonwords
Expt block duration	18s	18s	18s	18s	18s
Trials per block	3	3	3	3	3
Trial duration	6s	6s	6s	6s	6s
Trial structure	300ms trial-initial fixation; 12 words/nonwords presented for 350 ms each; 1000 ms probe; 500 ms trial-final fixation	300ms trial-initial fixation; 12 words/nonwords presented for 350 ms each; 1000 ms probe; 500 ms trial-final fixation	300ms trial-initial fixation; 12 words/nonwords presented for 350 ms each; 1000 ms probe; 500 ms trial-final fixation	300ms trial-initial fixation; 12 words/nonwords presented for 350 ms each; 1000 ms probe; 500 ms trial-final fixation	100ms trial-initial fixation; 12 words/nonwords presented for 450ms each; 400ms hand icon; 100ms trial-final fixation
Expt blocks per run	16	18	18	18	16

Expt blocks per cond per run	8	6	6	6	8
Fix block duration	18s	18s	18s	18s	14s
Fix blocks per run	5	4	4	4	5
Run duration (in s)	378	396	396	396	358

Figures

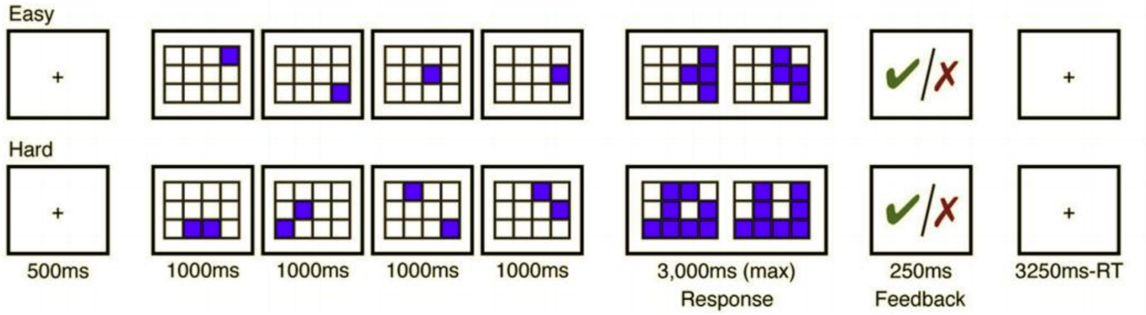

Figure 1. Procedure and timing for the spatial working memory task used to localize the multiple demand (MD) fROIs.

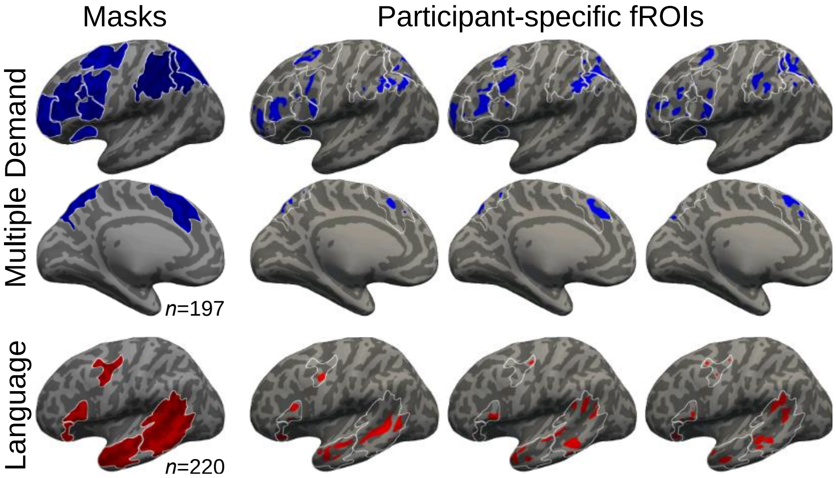

Figure 2. Masks and subject-specific functional regions of interest (fROIs). Data are shown for the multiple-demand (MD) network (top, middle; blue) and language network (bottom; red), and are approximate projections from functional volumes onto the cortical surface of an inflated average brain in common 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 example of 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. Note that, 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 denote the borders of the masks.

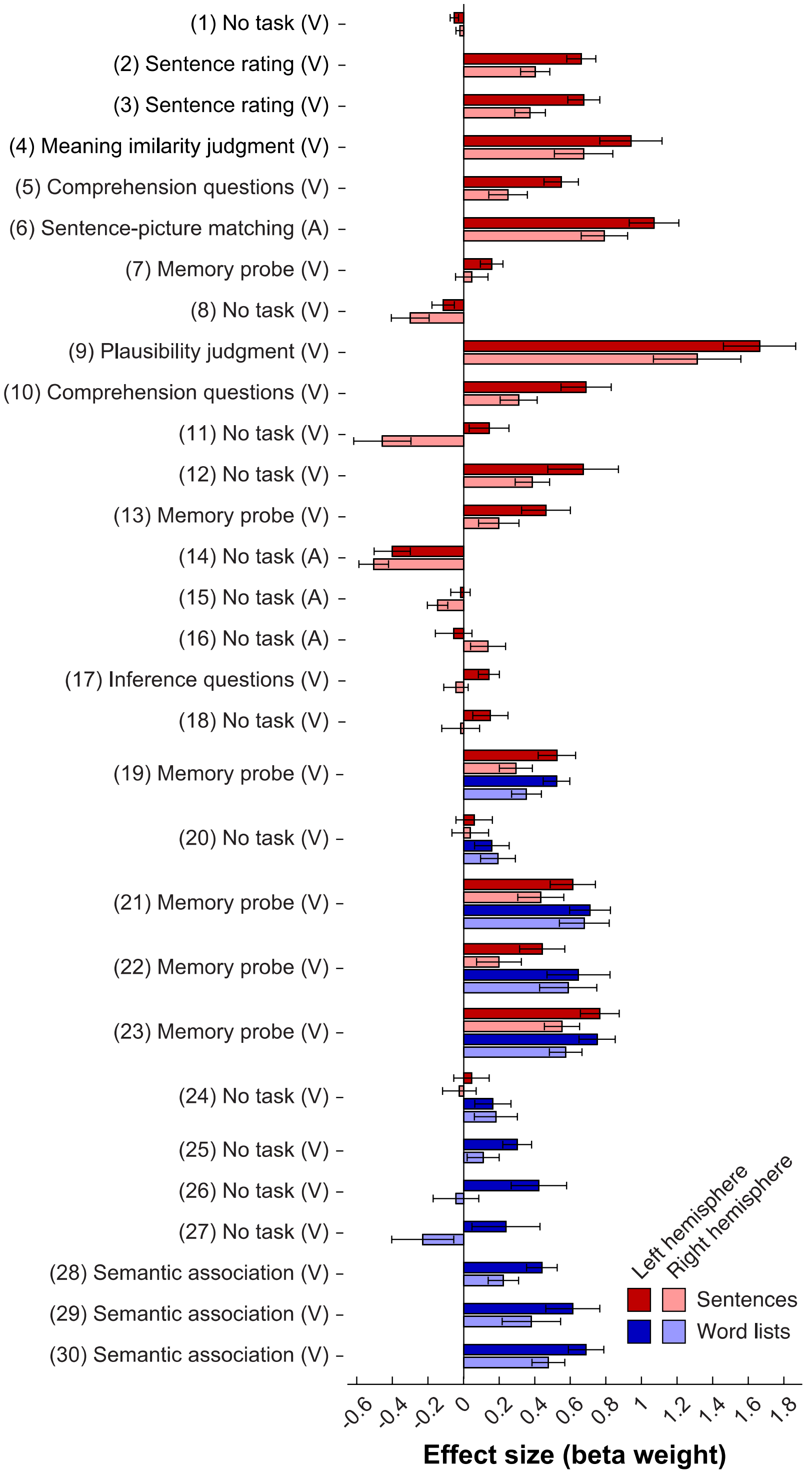

Figure 3. Laterality of responses to sentences and word lists in the MD network across experiments. Responses (beta weights for the corresponding regressors in the GLM) are shown averaged across fROIs in the left (darker colors) and right (brighter colors) hemispheres, separately for sentences (red shades) and word lists (blue shades) conditions. Responses are measured as beta weights for the corresponding condition regressors in the GLM. Bars show the average response across subjects, and errors bars

show standard errors of the mean across subjects. Most experiments include either sentences or word lists, except for experiments 19-24.

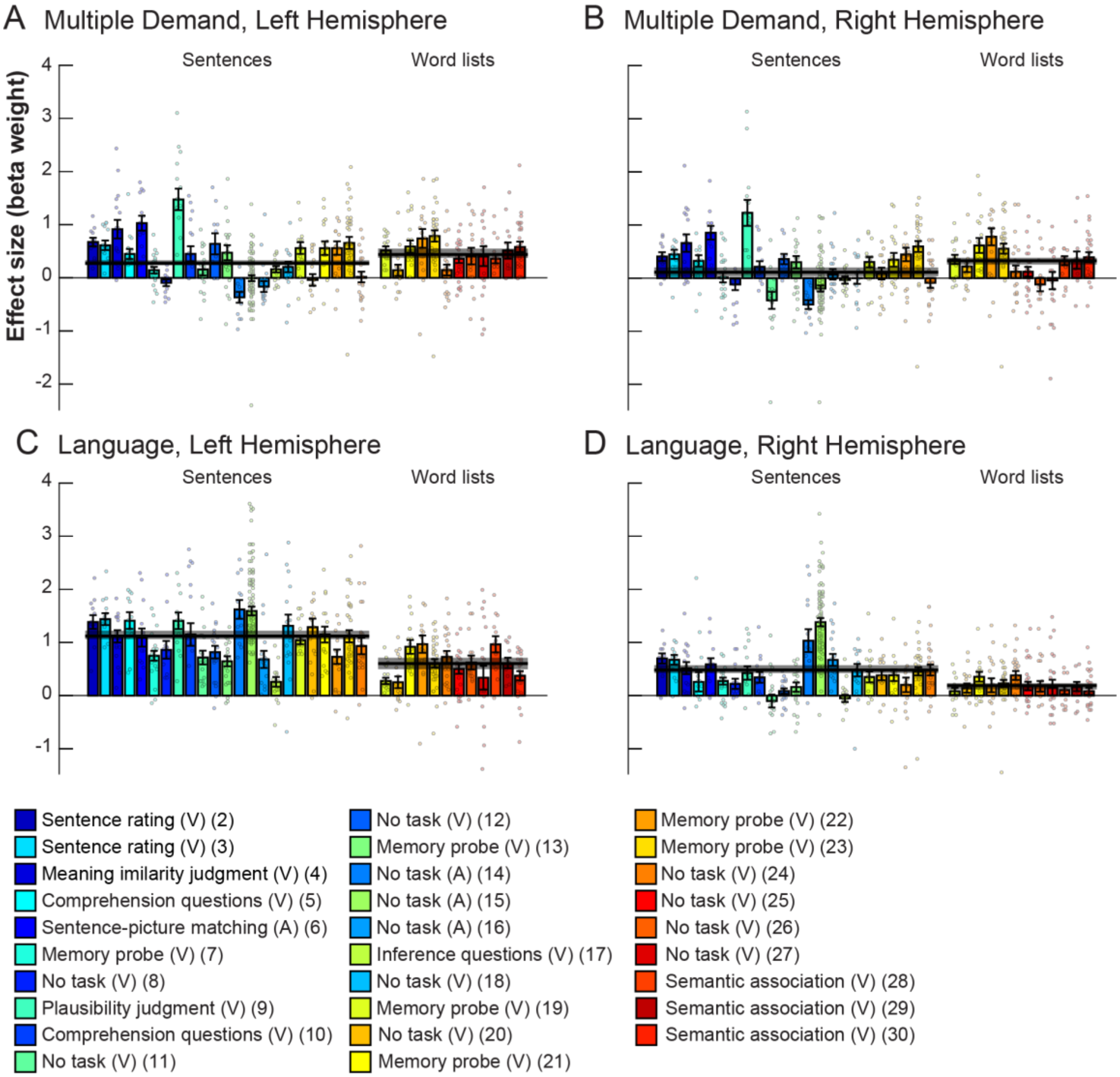

Figure 4. Responses of each network and hemisphere to the sentences and word lists conditions across experiments. Responses (beta weights for the corresponding regressors in the GLM) are shown averaged across fROIs in the MD (top) and language (bottom) networks, separately for the left hemisphere (left) and right hemisphere (right). 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 Figure 3). Dots show data for individual subjects, bars show the average response across subjects, and errors bars show standard errors of the mean across subjects. Thick, horizontal black lines are averaged across experiments, and gray rectangles are the corresponding 95% confidence intervals. Note that 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 east of comparison.

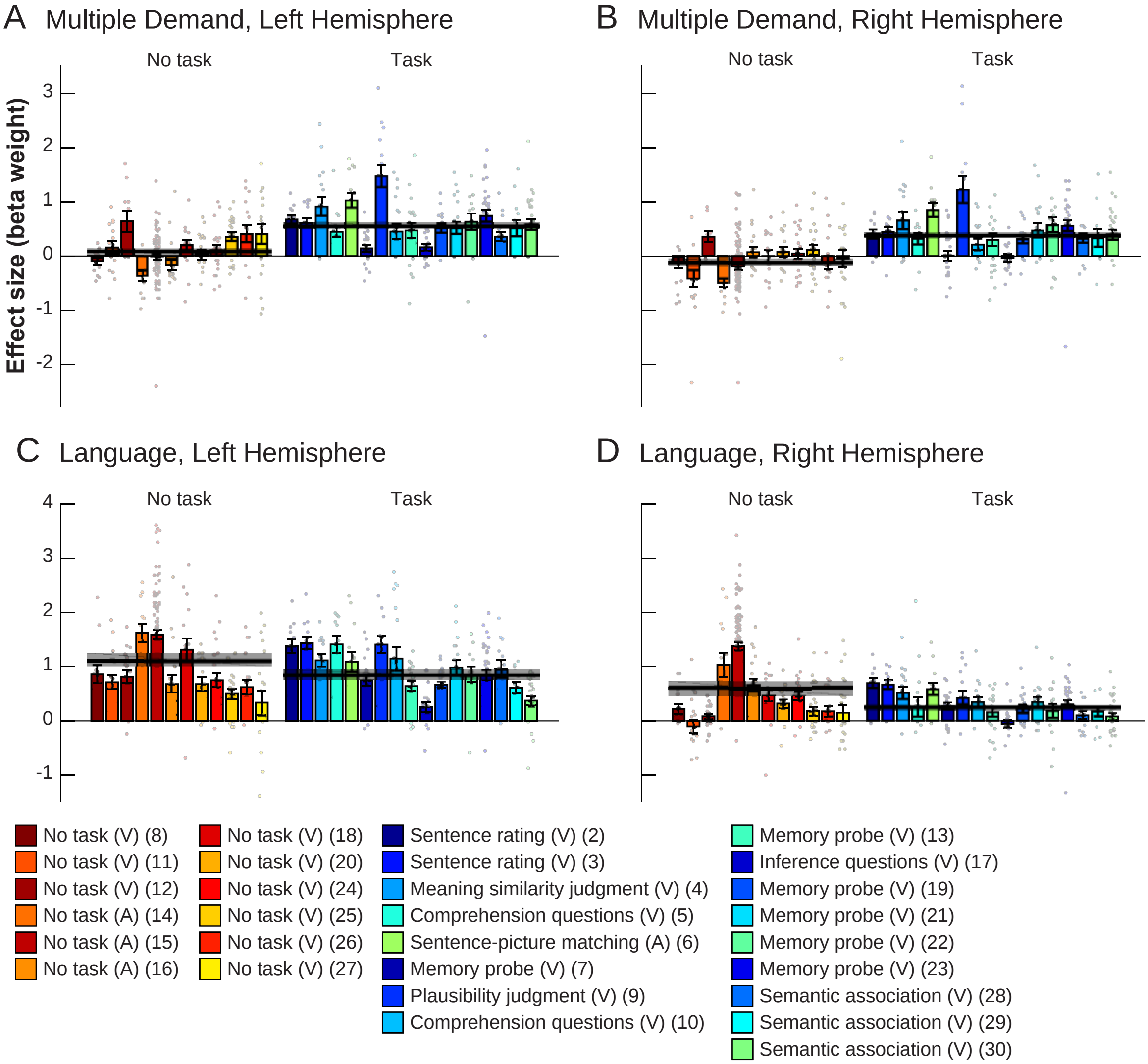

Figure 5. Responses of each network and hemisphere to passive and task-based paradigms across experiments. Same conventions as in Figure 4, but bars are now grouped by whether the experimental paradigm was passive comprehension or an active task. For experiments that contained both sentences and word lists, responses are averaged across these two conditions.

References

Abney, S. P., & Johnson, M. (1991). Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research*, *20*(3), 233-250.

Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: two activation likelihood estimation (ALE) meta-analyses. *Brain and language*, *122*(1), 42-54.

Alario, F. X., Chainay, H., Lehericy, S., & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Research*, *1076*, 129-143. doi:10.1016/j.brainres.2005.11.104

Alm, H., & Nilsson, L. (2001). The use of car phones and changes in driver behaviour. *International Journal of Vehicle Design*, *26*(1), 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. *Journal of cognitive neuroscience*, *23*(2), 349-361.

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

Ayyash, D., Gallée, J., Mineroff, Z., Moraleda, S. M., Jouravlev, O., & Fedorenko, E. (in prep.). The universal language network: A cross-linguistic investigation spanning 41 languages and 10 language families.

Bates, D., Machler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Blanco-Elorrieta, E., & Pylkkänen, L. (2017). Bilingual language switching in the laboratory versus in the wild: The spatiotemporal dynamics of adaptive language control. *Journal of Neuroscience, 37*(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. *Journal of Neuroscience, 37*(41), 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. *Journal of Neurophysiology*, *112*(5), 1105-1118. doi:10.1152/jn.00884.2013

Bock, K. (1995). Sentence Production: From Mind to Mouth. In J. L. Miller & P. D. Eimas (Eds.), *Speech, Language, and Communication. Handbook of perception and cognition.* (pp. 181–216). San Diego, US: Academic Press.

Bonini, M. V., & Radanovic, M. (2015). Cognitive deficits in post-stroke aphasia. *Arquivos de neuro-psiquiatria, 73*(10), 840-847.

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). *Region of interest analysis using an SPM toolbox.* Paper presented at the 8th international conference on functional mapping of the human brain.

Brownsett, S. (2014). Using fMRI and Behavioural Measures to Investigate Rehabilitation in Post-Stroke Aphasic Deficits.

Buchweitz, A., Mason, R. A., Tomitch, L. M. B., & Just, M. A. (2009). Brain activation for reading and listening comprehension: An fMRI study of modality effects and individual differences in language comprehension. *Psychol Neurosci, 2*(2), 111-123.

Campbell, K. L., & Tyler, L. K. (2018). Language-related domain-specific and domain-general systems in the human brain. *Current opinion in behavioral sciences, 21*, 132-137.

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

Coelho, C., Lê, K., Mozeiko, J., Krueger, F., & Grafman, J. (2012). Discourse production following injury to the dorsolateral prefrontal cortex. *Neuropsychologia*, *50*(14), 3564-3572.

Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *Neuroimage, 37*(1), 343-360. doi:10.1016/j.neuroimage.2007.03.071

Corbett, F., Jefferies, E., Ehsan, S., & Lambon Ralph, M. A. (2009). Different impairments of semantic cognition in semantic dementia and semantic aphasia: evidence from the non-verbal domain. *Brain*, *132*(9), 2593-2608.

Cramer, S. C. (2008). Repairing the human brain after stroke: I. Mechanisms of spontaneous recovery. *Annals of neurology*, *63*(3), 272-287.

Crittenden, B. M., & Duncan, J. (2012). Task difficulty manipulation reveals multiple demand activity but no frontal lobe hierarchy. *Cerebral Cortex, 24*(2), 532-540.

Cromer, J. A., Roy, J. E., & Miller, E. K. (2010). Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron*, *66*(5), 796-807.

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

Duncan, J. (2013). The Structure of Cognition: Attentional Episodes in Mind and Brain. *Neuron, 80*(1), 35-50. doi:10.1016/j.neuron.2013.09.015

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences, 23*(10), 475-483.

Eickhoff, S. B., Heim, S., Zilles, K., & Amunts, K. (2009). A systems perspective on the effective connectivity of overt speech production. *Movement Disorders, 24*, S187-S187.

El Hachioui, H., Visch-Brink, E. G., Lingsma, H. F., van de Sandt-Koenderman, M. W., Dippel, D. W., Koudstaal, P. J., & Middelkoop, H. A. (2014). Nonlinguistic cognitive impairment in poststroke aphasia: a prospective study. *Neurorehabilitation and Neural Repair, 28*(3), 273-281.

Endo, R. (2013). Realities, rewards, and risks of heritage-language education: Perspectives from Japanese immigrant parents in a Midwestern community. *Bilingual Research Journal, 36*(3), 278-294.

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in Psychology, 5*, 335. doi:10.3389/fpsyg.2014.00335

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

Fedorenko, E., & Blank, I. A. (submitted). Broca's area is not a natural kind.

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

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of memory and language, 54*(4), 541-553.

Fedorenko, E., Gibson, E., & Rohde, D. (2007). The nature of working memory in linguistic, arithmetic and spatial integration processes. *Journal of Memory and Language, 56*(2), 246-269.

Fedorenko, E., Hsieh, P. J., Nieto-Castanon, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New Method for fMRI Investigations of Language: Defining ROIs Functionally in Individual Subjects. *Journal of Neurophysiology, 104*(2), 1177-1194. doi:10.1152/jn.00032.2010

Fedorenko, E., & Kanwisher, N. (2009). Neuroimaging of Language: Why Hasn't a Clearer Picture Emerged? *Language and Linguistics Compass, 3*(4), 839-865.

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

Fedorenko, E., & Varley, R. (2016). Language and thought are not the same thing: Evidence from neuroimaging and neurological patients. *Annals of the New York Academy of Sciences, 1369*(1), 132-153. doi:10.1111/nyas.13046

Feydy, A., Carlier, R., Roby-Brami, A., Bussel, B., Cazalis, F., Pierot, L., & Maier, M. (2002). Longitudinal study of motor recovery after stroke: recruitment and focusing of brain activation. *Stroke, 33*(6), 1610-1617.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science, 291*(5502), 312-316.

Futrell, R., & Gibson, E. (2017). L2 processing as noisy channel language comprehension. *Bilingualism: Language and Cognition, 20*(4), 683-684.

Geranmayeh, F., Brownsett, S. L., Leech, R., Beckmann, C. F., Woodhead, Z., & Wise, R. J. (2012). The contribution of the inferior parietal cortex to spoken language production. *Brain and language, 121*(1), 47-57.

Geranmayeh, F., Brownsett, S. L., & Wise, R. J. (2014). Task-induced brain activity in aphasic stroke patients: what is driving recovery? *Brain, 137*(10), 2632-2648.

Geranmayeh, F., Chau, T. W., Wise, R. J., Leech, R., & Hampshire, A. (2017). Domain-general subregions of the medial prefrontal cortex contribute to recovery of language after stroke. *Brain, 140*(7), 1947-1958.

Geranmayeh, F., Leech, R., & Wise, R. J. (2016). Network dysfunction predicts speech production after left hemisphere stroke. *Neurology, 86*(14), 1296-1305.

Gernsbacher, M. A. (1993). Less skilled readers have less efficient suppression mechanisms. In: SAGE Publications Sage CA: Los Angeles, CA.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition, 68*(1), 1-76. doi:10.1016/s0010-0277(98)00034-1

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National*

Academy of Sciences of the United States of America, *110*(20), 8051-8056. doi:10.1073/pnas.1216438110

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological science, 13*(5), 425-430.

Grande, M., Meffert, E., Schoenberger, E., Jung, S., Frauenrath, T., Huber, W., & 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*(3), 702-714.

Grindrod, C. M., Bilenko, N. Y., Myers, E. B., & Blumstein, S. E. (2008). The role of the left inferior frontal gyrus in implicit semantic competition and selection: An event-related fMRI study. *Brain Research, 1229*, 167-178.

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological science, 26*(4), 433-443.

Hartwigsen, G. (2018). Flexible redistribution in cognitive networks. *Trends in cognitive sciences, 22*(8), 687-698.

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: memory as an integral component of information processing. *Trends in Cognitive Sciences*, *19*(6), 304-313. doi:10.1016/j.tics.2015.04.006

Hasson, U., Egidi, G., Marelli, M., & Willems, R. M. (2018). Grounding the neurobiology of language in first principles: The necessity of non-language-centric explanations for language comprehension. *Cognition, 180*, 135-157.

Heim, S., Amunts, K., Hensel, T., Grande, M., Huber, W., Binkofski, F., & Eickhoff, S. B. (2012). The role of human parietal area 7A as a link between sequencing in hand actions and in overt speech production. *Frontiers in psychology, 3*, 534.

Hervais-Adelman, A. G., Carlyon, R. P., Johnsrude, I. S., & Davis, M. H. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. *Language and Cognitive Processes*, *27*(7-8), 1145-1166.

Hirshorn, E. A., & Thompson-Schill, S. L. (2006). Role of the left inferior frontal gyrus in covert word retrieval: neural correlates of switching during verbal fluency. *Neuropsychologia*, *44*(12), 2547-2557.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *50*(3), 346-363.

Hugdahl, K., Raichle, M. E., Mitra, A., & Specht, K. (2015). On the existence of a generalized non-specific task-dependent network. *Frontiers in Human Neuroscience, 9*. doi:10.3389/fnhum.2015.00430

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1-2), 101-144.

Ivanova, I., Ferreira, V. S., & Gollan, T. H. (2017). Form overrides meaning when bilinguals monitor for errors. *Journal of memory and language, 94*, 75-102.

Jackson, J., Rich, A. N., Williams, M. A., & Woolgar, A. (2017). Feature-selective attention in frontoparietal cortex: multivoxel codes adjust to prioritize task-relevant information. *Journal of cognitive neuroscience, 29*(2), 310-321.

Jacoby, N., & Fedorenko, E. (2018). Discourse-level comprehension engages medial frontal Theory of Mind brain regions even for expository texts. *Language, Cognition and Neuroscience*, 1-17.

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

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*: Harvard University Press.

Jouravlev, O., Zheng, D., Balewski, Z., Goldin-Meadow, S., & Fedorenko, E. (2019). Speech-accompanying gestures do not engage high-level language processing brain regions. *Neuropsychologia*.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of memory and language*, *30*(5), 580-602.

Kleim, J. A. (2011). Neural plasticity and neurorehabilitation: teaching the new brain old tricks. *Journal of communication disorders*, *44*(5), 521-528.

Kline, M., Gallee, J., Balewski, Z., & Fedorenko, E. (submitted). Understanding jokes relies on the theory of mind system.

Kong, X.-Z., Tzourio-Mazoyer, N., Joliot, M., Fedorenko, E., Liu, J., Fisher, S. E., & Francks, C. (2018). Gene Expression Correlates of the Cortical Network Underlying Sentence Processing. *bioRxiv*, 439984.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience, 12*(5), 535-540. doi:10.1038/nn.2303

Kumano, H., Suda, Y., & Uka, T. (2016). Context-dependent accumulation of sensory evidence in the parietal cortex underlies flexible task switching. *Journal of Neuroscience*, *36*(48), 12192-12202.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience, 15*(2), 272-293. doi:10.1162/089892903321208204

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13).

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, Hawaii, USA.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10*(10), 447-454. doi:10.1016/j.tics.2006.08.007

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior research methods, 49*(4), 1494-1502.

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

Martin, R. C., & Allen, C. M. (2008). *A disorder of executive function and its role in language processing.* Paper presented at the Seminars in speech and language.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7-8), 953-978.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of psycholinguistic research, 29*(2), 111-123.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(3), 817.

McMillan, C. T., Coleman, D., Clark, R., Liang, T.-W., Gross, R. G., & Grossman, M. (2013). Converging evidence for the processing costs associated with ambiguous quantifier comprehension. *Frontiers in psychology, 4*, 153.

Meier, E. L., Kapse, K. J., & Kiran, S. (2016). The relationship between frontotemporal effective connectivity during picture naming, behavior, and preserved cortical tissue in chronic aphasia. *Frontiers in human neuroscience, 10*, 109.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience, 24*(1), 167-202.

Mineroff, Z. A., 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*.

Monti, M. M., Parsons, L. M., & Osherson, D. N. (2012). Thought Beyond Language: Neural Dissociation of Algebra and Natural Language. *Psychological Science, 23*(8), 914-922. doi:10.1177/0956797612437427

Moser, D. C., Fridriksson, J., & Healy, E. W. (2007). Sentence comprehension and general working memory. *Clinical linguistics & phonetics, 21*(2), 147-156.

Nestor, P. J., Graham, N. L., Fryer, T. D., Williams, G. B., Patterson, K., & Hodges, J. R. (2003). Progressive non-fluent aphasia is associated with hypometabolism centred on the left anterior insula. *Brain, 126*(11), 2406-2418.

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

Nieuwland, M. S., Martin, A. E., & Carreiras, M. (2012). Brain regions that process case: evidence from Basque. *Human brain mapping, 33*(11), 2509-2520.

Noesselt, T., Shah, N. J., & Jäncke, L. (2003). Top-down and bottom-up modulation of language related areas–an fMRI study. *BMC neuroscience, 4*(1), 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*(1), 361-378.

Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive neuropsychology, 26*(6), 527-567.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive Affective & Behavioral Neuroscience, 5*(3), 263-281. doi:10.3758/cabn.5.3.263

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Pallier, C., Devauchelle, A. D., & Dehaene, S. (2010). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences of the United States of America, 108*(6), 2522-2527. doi:10.1073/pnas.1018711108

Paunov, A., Blank, I., & Fedorenko, E. (2019). A tight relationship between the language and theory of mind networks especially during language comprehension.

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing, 39*(2), 204.

Peelle, J. E., Johnsrude, I., & Davis, M. H. (2010). Hierarchical processing for speech in human auditory cortex and beyond. *Frontiers in human neuroscience, 4*, 51.

Petersen, S. E., & Sporns, O. (2015). Brain networks and cognitive architectures. *Neuron, 88*(1), 207-219.

Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., ... & Fedorenko, E. (2018). Toward a universal decoder of linguistic meaning from brain activation. *Nature communications*, *9*(1), 963.

Pliatsikas, C., & Luk, G. (2016). Executive control in bilinguals: A concise review on fMRI studies. *Bilingualism: Language and Cognition*, *19*(4), 699-705.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., & Schlaggar, B. L. (2011). Functional network organization of the human brain. *Neuron, 72*(4), 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.

Rasmussen, N. E., & Schuler, W. (2018). Left-corner parsing with distributed associative memory produces surprisal and locality effects. *Cognitive science, 42*, 1009-1042.

Resnik, P. (1992). *Left-corner parsing and psychological plausibility.* Paper presented at the Proceedings of the 14th conference on Computational linguistics-Volume 1.

Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex, 15*(8), 1261-1269.

Rodd, J. M., Johnsrude, I. S., & Davis, M. H. (2010). The role of domain-general frontal systems in language comprehension: Evidence from dual-task interference and semantic ambiguity. *Brain and language, 115*(3), 182-188.

Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: A review. *Frontiers in psychology*, *2*, 307.

Roskies, A. L., Fiez, J. A., Balota, D. A., Raichle, M. E., & Petersen, S. E. (2001). Task-dependent modulation of regions in the left inferior frontal cortex during semantic processing. *Journal of cognitive neuroscience*, *13*(6), 829-843.

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

Schuler, W., AbdelRahman, S., Miller, T., & Schwartz, L. (2010). Broad-coverage parsing using human-like memory constraints. *Computational Linguistics, 36*(1), 1-30.

Scott, S. K., & McGettigan, C. (2013). The neural processing of masked speech. *Hearing research, 303*, 58-66.

Scott, T. L., Gallee, J., & Fedorenko, E. (2017). A new fun and robust version of an fMRI localizer for the frontotemporal language system. *Cognitive Neuroscience*, 1-10. doi:10.1080/17588928.2016.1201466

Shain, C., Blank, I. A., van Schijndel, M., Schuler, W., & Fedorenko, E. (2019). fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *bioRxiv*, 717512.

Shuster, L. I., & Lemieux, S. K. (2005). An fMRI investigation of covertly and overtly produced mono-and multisyllabic words. *Brain and language*, *93*(1), 20-31.

Siegelman, M., Blank, I. A., Mineroff, Z., & Fedorenko, E. (2019). An attempt to conceptually replicate the dissociation between syntax and semantics during sentence comprehension. *Neuroscience*.

Siegelman, M., Kean, H., & Fedorenko, E. (in prep.). The insensitivity of the high-level language areas to task manipulations.

Simic, T., Rochon, E., Greco, E., & Martino, R. (2017). Baseline executive control ability and its relationship to language therapy improvements in post-stroke aphasia: a systematic review. *Neuropsychological rehabilitation*, *29*(3), 395-439.

Sims, J. A., 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.

Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J. A., Petersson, K. M., & Hagoort, P. (2009). Retrieval and Unification of Syntactic Structure in Sentence Comprehension: an fMRI Study Using Word-Category Ambiguity. *Cerebral Cortex, 19*(7), 1493-1503. doi:10.1093/cercor/bhn187

Strijkers, K., Chanoine, V., Munding, D., Dubarry, A.-S., Trébuchon, A., Badier, J.-M., & Alario, F.-X. (2019). Grammatical class modulates the (left) inferior frontal gyrus within 100 milliseconds when syntactic context is predictive. *Scientific reports*, *9*(1), 4830.

Strijkers, K., Costa, A., & Thierry, G. (2009). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral cortex*, *20*(4), 912-928.

Strijkers, K., Holcomb, P. J., & Costa, A. (2011). Conscious intention to speak proactively facilitates lexical access during overt object naming. *Journal of memory and language*, *65*(4), 345-362.

Thompson-Schill, S. L., Jonides, J., Marshuetz, C., Smith, E. E., D'Esposito, M., Kan, I. P., & Swick, D. (2002). Effects of frontal lobe damage on interference effects in working memory. *Cognitive, Affective, & Behavioral Neuroscience*, *2*(2), 109-120.

Traxler, M. J. (2014). Trends in syntactic parsing: anticipation, Bayesian estimation, and good-enough parsing. *Trends in Cognitive Sciences, 18*(11), 605-611. doi:10.1016/j.tics.2014.08.001

Troiani, V., Fernández-Seara, M. A., Wang, Z., Detre, J. A., Ash, S., & Grossman, M. (2008). Narrative speech production: an fMRI study using continuous arterial spin labeling. *Neuroimage*, *40*(2), 932-939.

Vaden, K. I., Kuchinsky, S. E., Cute, S. L., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2013). The cingulo-opercular network provides word-recognition benefit. *Journal of Neuroscience*, *33*(48), 18979-18986.

Vagharchakian, L., Dehaene-Lambertz, G., Pallier, C., & Dehaene, S. (2012). A temporal bottleneck in the language comprehension network. *Journal of Neuroscience, 32*(26), 9089-9102. doi:10.1523/jneurosci.5685-11.2012

Van Schijndel, M., Exley, A., & Schuler, W. (2013). A model of language processing as hierarchic sequential prediction. *Topics in Cognitive Science*, *5*(3), 522-540.

Vergauwe, E., Barrouillet, P., & Camos, V. (2010). Do mental processes share a domain-general resource? *Psychological science*, *21*(3), 384-390.

Villard, S., & Kiran, S. (2017). To what extent does attention underlie language in aphasia? *Aphasiology*, *31*(10), 1226-1245.

Wall, K. J., Cumming, T. B., & Copland, D. A. (2017). Determining the association between language and cognitive tests in poststroke aphasia. *Frontiers in neurology, 8*, 149.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology Section A, 49*(1), 51-79.

Wehbe, L., Blank, I. A., Shain, C., Futrell, R., Levy, R., von der Malsburg, T., & Fedorenko, E. (submitted). Neural activity in the fronto-temporal language system is predicted by incremental language comprehension difficulty.

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention. *Journal of Neuroscience, 32*(40), 14010-14021.

Willems, R. M., Van der Haegen, L., Fisher, S. E., & Francks, C. (2014). On the other hand: including left-handers in cognitive neuroscience and neurogenetics. *Nature Reviews Neuroscience, 15*(3), 193-201. doi:10.1038/nrn3679

Wilson, S. M., Isenberg, A. L., & Hickok, G. (2009). Neural correlates of word production stages delineated by parametric modulation of psycholinguistic variables. *Human brain mapping, 30*(11), 3596-3608.

Wingfield, A., & Grossman, M. (2006). Language and the aging brain: patterns of neural compensation revealed by functional brain imaging. *Journal of Neurophysiology, 96*(6), 2830-2839.

Woolgar, A., Duncan, J., Manes, F., & Fedorenko, E. (2018). Fluid intelligence is supported by the multiple-demand system not the language system. *Nature Human Behaviour, 2*(3), 200.

Wright, P., Randall, B., Marslen-Wilson, W. D., & Tyler, L. K. (2011). Dissociating linguistic and task-related activity in the left inferior frontal gyrus. *Journal of Cognitive Neuroscience, 23*(2), 404-413.

Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language, 37*(4), 533-554.