nature portfolio

Peer Review File

A language network in the individualized functional connectomes of 1,199 human brains doing arbitrary tasks

Corresponding Author: Dr Cory Shain

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Nature Communications.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have adequately address all my previous comments. I appreciate them clarifying the scope and purpose of this work, which is now conveyed in the manuscript. I think this paper will make a strong contribution to the field.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
The authors have addressed my concerns. We disagree on the role of different nodes of the language network but I don't need to impose my point of view.

(Remarks on code availability)
no comments

Reviewer #3

(Remarks to the Author)
I remain unclear about two key aspects of this study. I apologize to the authors for not making this clearer in the original review:

1. A key, central argument in this paper is that the “language” network can be pulled out of an undifferentiated (“arbitrary”) collection of task fMRI data: but how “undifferentiated” is the fMRI dataset? The authors have not provided an analysis of the tasks included in the big dataset, nor an analysis of whether the functional connectivity results vary across them. This might be important because, if most of the tasks run by the Fedorenko group over this long period of time most commonly involved language/comprehension, then the main result might not be so surprising. If one could pull the “language” network out of tasks that focus participants on the phonological structure of words or entirely non-language tasks then it might be more novel. The authors have provided a link to a csv OSF file which contains an abbreviated shorthand label for each study task. Most of these labels are uninterpretable and no categorisation or more detailed analysis is offered. The abbreviated labels suggest that most (but not all) relate to comprehension and language. An obvious thing to check would be to divide these studies into those that involve language vs those that do not. And see if these individual task connectivity maps for the language network can be extracted from the non-language tasks and, if so, whether this extracted functional-connectivity map is as accurate as that extracted for the language-tasks.
An alternative approach would be to use the collected works from another fMRI group whose research is not focussed on language (e.g., Kanwisher’s explorations of face processing) and see if the language network could be recovered from it.
2. How “accurate” is the extracted network: the utility of pulling out this network from other task data depends on how accurate the method is. It is unclear if a value of z(r)~=0.35 is sufficiently good?

(Remarks on code availability)

We have provided point-by-point responses (in blue) to comments from R3 (in red) below:

1. A key, central argument in this paper is that the “language” network can be pulled out of an undifferentiated (“arbitrary”) collection of task fMRI data: but how “undifferentiated” is the fMRI dataset? The authors have not provided an analysis of the tasks included in the big dataset, nor an analysis of whether the functional connectivity results vary across them. This might be important because, if most of the tasks run by the Fedorenko group over this long period of time most commonly involved language/comprehension, then the main result might not be so surprising. If one could pull the “language” network out of tasks that focus participants on the phonological structure of words or entirely non-language tasks then it might be more novel.

We directly investigated this concern in section "LangFC Emerges Across Task States" and Fig 3A ("Nonlinguistic runs"), where we found that the topography and selectivity of LangFC is qualitatively conserved even when parcellating it solely from nonlinguistic tasks (see Materials & Methods for details of nonlinguistic run selection).

The authors have provided a link to a csv OSF file which contains an abbreviated shorthand label for each study task. Most of these labels are uninterpretable and no categorisation or more detailed analysis is offered. The abbreviated labels suggest that most (but not all) relate to comprehension and language.

We apologize for the confusion. We included two tables: a table of task names completed by each participant in each session (which uses the abbreviated labels), and a table of brief descriptions of each task by name. Unfortunately, we had accidentally swapped the titles of these tables, which likely led to difficulty finding the relevant information. This error has now been corrected in our OSF repository for verification by the reviewer. The list of task descriptions is called “langlocFC_task_descriptions.csv” and the list of task names by session is called “langlocFC_task_names_by_session.csv”. We will add both tables to our future SDC repository as well.

An obvious thing to check would be to divide these studies into those that involve language vs those that do not. And see if these individual task connectivity maps for the language network can be extracted from the non-language tasks and, if so, whether this extracted functional-connectivity map is as accurate as that extracted for the language-tasks.

Per our comment above, this is what we indeed did. The resulting network is quantitatively less similar to the language localizer task maps than it is when estimated from the full dataset, but qualitatively similar in topography and function. Because the non-linguistic parcellation is by definition based on less data, we cannot determine whether the quantitative degradation is due to task contents, data quantity, or both. However, critically, even data from purely non-linguistic tasks allow for the recovery of the language network.

An alternative approach would be to use the collected works from another fMRI group whose research is not focussed on language (e.g., Kanwisher’s explorations of face processing) and see if the language network could be recovered from it.

Some of our nonlinguistic task data included Kanwisher-style localizer tasks for category-selective visual areas. We did consider additional datasets from other labs, but unfortunately, those datasets do not include a language localizer task in each individual, which is what is needed for our goal of validating the FC-based approach.

2. How “accurate” is the extracted network: the utility of pulling out this network from other task data depends on how accurate the method is. It is unclear if a value of z(r)~=0.35 is sufficiently good?

Unless we have misunderstood, this appears to be a restatement of a comment from the last round, to which we provided a detailed response, which we are reproducing for convenience below. If you have any specific concerns with these arguments, we would be happy to elaborate further.

R3-14 (pasted from the response letter during the first round of review; we pasted our earlier response R3-17 below, as well)

5. Fig 2 – what counts as a good value? Lang and Lana should be good because they are used to aligned the parcels. But the values are not high? How does this compare to – same analysis applied to active task data (most likely to give maximal values) and rs-fMRI.

“What counts as a good value?” is an insightful question that is difficult to answer *a priori*. As noted here and in R3-17, even the z(r) between LanA and LANG is relatively low in absolute terms (see R3-17 for discussion, and R3-4 for discussion of the question about comparison to parcellating active task data). There are two key factors that might put downward pressure on the correlations between LangFC and the localizer contrasts (e.g., sentences vs. nonwords). *First*, fMRI is a noisy modality and we are estimating from relatively little data by parcellating sessions independently, in contrast with recent related work that has used many sessions per participant (e.g., Braga et al., 2020; Du et al., 2024). The correlations will inevitably be influenced by various types of measurement error. *Second*, perfect correspondence to some task contrast is not necessarily even desirable, since task contrasts, like connectivity, are useful but imperfect tools for studying functional organization. Task contrasts may have systematic artifacts. For example, the sentences vs. nonwords contrast tends to capture temporoparietal areas that have been argued on other grounds not to belong to the language network in most people (Shain et al., 2023). And the intact vs. degraded speech contrast tends to pull in some non-language-selective auditory areas (Salvo et al., 2025). LangFC may lack these artifacts, which would lower its correlation to task maps despite being a success of the method.

For this reason, we have tried to provide other landmarks as to what these numbers mean in practice. *First*, the overlap maps in Fig 2D and SI A help visualize what a particular level of spatial correlation looks like. *Second*, the *t*-value analyses shed partially complementary light on the selectivity properties of connectivity-based networks, since this statistic primarily rewards strong task responses within the network, irrespective of whether task responses also fall outside of the network. The fact that LangFC has a mean z(r) of around 0.35 to the auditory language localizer contrast while also showing a mean *t*-value of around 3 means that 0.35 is a sufficient level of correlation to produce strong average task responses within the network. *Third*, by quantifying z(r) between many different types of statistical maps, we have given a useful baseline for judging *relative* similarity. In particular, it is clear from Fig 3A and 3B that LangFC's z(r) to language contrasts is much higher than its z(r) to any other type of contrast we considered, and that other networks show different tuning.

R3-17

7. The two reference maps are only related at Z(r)=0.37 – this seems surprisingly low.

We have re-checked this number and are confident that it is correct. As to why the number is low despite visual similarity, we see several plausible contributors. The first is that the visual impression of similarity is influenced by areas falling in the general vicinity of each other, whereas correlation cares about exact overlap: even if a region (e.g., in temporoparietal cortex) has a hotspot in both maps, any relative shift in location between the two maps will substantially lower the spatial correlation. The second is that each map contains areas that are lacking in the other. For example, LanA covers a greater extent of the left temporal lobe and includes some unique areas of medial occipital and frontal cortex, whereas LANG includes an anterior frontal area not found in LanA. The third is that the maps differ in kind: LanA is a continuous probabilistic atlas, and LANG is a binary atlas. Thus the correlation will be reduced even for overlapping areas because the degrees of probability in LanA will not be captured in LANG. Finally, as shown, LANG is more bilateral than LanA, so their correlation will be reduced by mismatch between the two atlases in their degree of hemispheric asymmetry.

Nonetheless, we share your impression that the two maps are nonetheless quite similar, which connects to your question in R3-14 about what counts as a good z(r). In this case, a good z(r) seems to be 0.37, which is approximately on par with the mean z(r) between LangFC and language tasks reported in Fig 3C.