Non-literal language processing is jointly supported by the language and Theory of Mind networks: Evidence from a novel meta-analytic fMRI approach

Miriam Hauptman[1,2,3], Idan Blank[†1,2,4,5], and Evelina Fedorenko[†1,2,6]

[1]Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139, USA
[2]McGovern Institute for Brain Research, MIT, Cambridge, MA 02139, USA
[3]Department of Psychological & Brain Sciences, Johns Hopkins University, Baltimore, MD 21218, USA
[4]Department of Psychology, UCLA, Los Angeles, CA 90095, USA
[5]Department of Linguistics, UCLA, Los Angeles, CA 90095, USA
[6]Program in Speech and Hearing in Bioscience and Technology, Harvard University, Boston, MA 02114, USA

† Co-senior authors

Corresponding Authors
Miriam Hauptman and Ev Fedorenko
mhauptm1@jhu.edu and evelina9@mit.edu; 43 Vassar Street, Room 46-3037, Cambridge, MA, 02139

Acknowledgements
We would like to thank Alex Paunov, Josef Affourtit, and Ben Lipkin for help with some of the analyses, and Ariel Goldberg and Jayden Ziegler for comments on the undergraduate thesis that resulted from this work. This work was supported by the NIH R01 award DC016607 and funds from MIT's Simons Center for the Social Brain (via SFARI); EF was further supported by the NIH R01 award DC016950, and funds from the McGovern Institute for Brain Research and the Brain and Cognitive Sciences department.

Abstract

Going beyond the literal meaning of utterances is key to communicative success. However, the mechanisms that support non-literal inferences remain debated. Using a novel meta-analytic approach, we evaluate the contribution of linguistic, social-cognitive, and executive mechanisms to non-literal interpretation. We identified 74 fMRI experiments (n=1,430 participants) from 2001-2021 that contrasted non-literal language comprehension with a literal control condition, spanning ten phenomena (e.g., metaphor, irony, indirect speech). Applying the activation likelihood estimation approach to the 825 activation peaks yielded six left-lateralized clusters. We then evaluated the locations of both the individual-study peaks and the clusters against probabilistic functional atlases (cf. macroanatomy, as is typically done) for three candidate brain networks—the language-selective network (Fedorenko et al., 2011), which supports language processing, the Theory of Mind (ToM) network (Saxe & Kanwisher, 2003), which supports social inferences, and the domain-general Multiple-Demand (MD) network (Duncan, 2010), which supports executive control. These atlases were created by overlaying individual activation maps of participants who performed robust and extensively validated 'localizer' tasks that target each network in question (n=806 for language; n=198 for ToM; n=691 for MD). We found that both the individual-study peaks and the ALE clusters fell primarily within the language network and the ToM network. These results suggest that non-literal processing is supported by both i) mechanisms that process literal linguistic meaning, and ii) mechanisms that support general social inference. They thus undermine a strong divide between literal and non-literal aspects of language and challenge the claim that non-literal processing requires additional executive resources.

Introduction

Communicative success often requires going beyond the literal meaning of utterances (e.g., Grice, 1975; Sperber & Wilson, 1986). Neurotypical adults engage in such 'non-literal' interpretation—from understanding metaphors, to interpreting sarcasm, to using prosodic cues—seemingly effortlessly. However, the cognitive and neural mechanisms that enable comprehension of non-literal language remain debated.

Early patient investigations linked difficulties in non-literal interpretation with damage to the *right hemisphere (RH)* (e.g., Winner & Gardner, 1977; Myers & Linebaugh, 1981; Delis et al., 1983; Brownell et al., 1983; 1986; Bryan, 1988; Joanette et al., 1989; Weylman et al., 1989; Burgess & Charello, 1996; Myers, 1998). More recently, a growing number of empirical studies have begun to challenge the idea of RH dominance in non-literal processing: brain imaging experiments (e.g., Rapp et al., 2004; Lee & Dapretto, 2006; Bosco et al., 2017; see also Oliveri et al., 2004), meta-analyses of such studies (e.g., Bohrn et al., 2012; Rapp et al., 2012; Reyes-Aguilar et al., 2018), and patient investigations (Ianni et al., 2014; Cardillo et al., 2018; Klooster et al., 2020; see Giora et al., 2000; Zaidel et al., 2002; Klepousniotou & Baum, 2005 for evidence of bilateral involvement) have instead implicated the *left hemisphere (LH).*

Furthermore, both the left and the right hemisphere each contain multiple distinct functional networks, including the language-selective network (e.g., Fedorenko et al., 2011) and its right homotope, the Theory of Mind network (e.g., Saxe & Kanwisher, 2003), and the domain-general Multiple Demand network (e.g., Duncan, 2010). These networks are associated with distinct cognitive operations, all of which have been argued to contribute to non-literal language comprehension, including in individuals with communication disorders: linguistic processing (e.g., Papagno & Genoni, 2004; Beaty & Silvia, 2013; Whyte & Nelson, 2015), social

inference (e.g., Sperber & Wilson, 1986; Happé, 1993; Winner et al., 1998), and executive control (e.g., McDonald & Pearce, 1998; Champagne-Lavau & Stip, 2010).

In recent years, the field of cognitive neuroscience has begun to move away from traditional group analyses, in which brain activations are averaged voxel-wise across individuals and the activation peaks are interpreted via reverse inference from anatomy to function (Poldrack, 2006, 2011; Fedorenko, 2021). Given evidence that functional areas vary in their precise locations in individual brains (e.g., Fischl et al., 2008; Frost & Goebel, 2011; Tahmasebi et al., 2012), the use of individual-subject-level analyses, or 'precision fMRI', has become increasingly widespread. In this type of analysis, the relevant functional areas and networks are identified in individual participants using 'localizer' tasks that target particular functional areas/networks (e.g., Kanwisher et al., 1997; Saxe et al., 2006; Fedorenko et al., 2010, 2013; Shashidara et al., 2019; Fedorenko, 2021; Gratton & Braga, 2021). Although the shift away from traditional group analyses promises to address questions that could not be answered when the field was dominated by this approach (e.g., Fedorenko et al., 2012; Deen et al., 2015; Braga & Buckner, 2017), ideally, we would not simply abandon many hundreds of past group-level fMRI studies. However, an analysis method that integrates findings from both individual-subject functional localization and traditional group-averaging experiments has not yet emerged.

Here, we present a novel meta-analytic approach for interpreting past fMRI studies and apply it to non-literal language processing. In particular, we leverage *probabilistic functional atlases* (e.g., Dworetsky et al., 2021; Thirion et al., 2021; Lipkin et al., in press), which are created from large numbers of individuals (n>100 for all atlases) who have performed extensively validated functional localizers that selectively target particular functional networks. Given that these atlases capture inter-individual variability in the locations of the relevant

networks, we can estimate for any given location in the common space the probability that it belongs to each candidate network. In other words, the probabilistic atlases provide a layer of information beyond anatomy. For example, a particular location in the common space may have a non-zero probability of belonging to two adjacent networks, and by comparing the relative probabilities associated with each network, we may infer which network is more likely to include this location.

This new approach differs from other commonly used meta-analytic approaches, particularly the Activation Likelihood Estimation (ALE; Turkeltaub et al., 2002; Eickhoff et al., 2009) approach. The interpretation of ALE estimates still relies on anatomical landmarks, or on comparisons between the results of multiple meta-analyses performed on group-level data. As such, inter-individual variability in functional architecture is not taken into account. This means that the functional resolution of ALE—its ability to discriminate between nearby functional networks—is low (the logic described in Nieto-Castañón & Fedorenko, 2012 for simple group analyses also applies to ALE meta-analyses). The current approach also differs from NeuroSynth (Yarkoni et al., 2011), as NeuroSynth involves comparing brain areas that are associated with subjectively selected keywords. Because probabilistic atlases are constructed using tasks that have been selectively and robustly linked to particular cognitive processes, relating peak locations to these atlases affords a higher degree of interpretability and a straightforward way to link the results to those from studies that rely on functional localization.

In the present study, we examined data from past fMRI studies (74 studies, 825 activation peaks) that contrasted neural responses to non-literal vs. literal conditions across diverse phenomena with respect to three candidate networks: the language-selective network (Fedorenko et al., 2011), which supports language processing, the Theory of Mind (ToM) network (Saxe &

Kanwisher, 2003), which supports social inferences, and the domain-general Multiple Demand (MD) network (Duncan, 2010), which supports executive control. In addition to incorporating probabilistic functional atlases in a novel manner, our meta-analysis improves upon past meta-analyses of non-literal language (e.g., Bohrn et al., 2012; Rapp et al., 2012; Reyes-Aguilar et al., 2018) in that we a) include a larger number of studies, b) focus on targeted, more interpretable contrasts (i.e., non-literal > literal; cf. non-literal > fixation), and c) examine a larger number of non-literal phenomena. To foreshadow our results, we find support for the role of the language-selective and ToM networks, but not the MD network, in non-literal interpretation. Further, in line with past meta-analyses (Bohrn et al., 2012; Rapp et al., 2012; Reyes-Aguilar et al., 2018), we do not find support for an RH bias: the majority of peaks fall in the LH.

Materials and Methods

Article selection criteria

The literature search was conducted in accordance with PRISMA guidelines (Moher et al., 2009). Relevant studies were identified in *NeuroSynth*, *Google Scholar*, *APA PsycInfo*, and *PubMed* databases using Boolean searches of the following keywords: “anaphora,” “anthropomorphism,” “comedy,” “discourse comprehension,” “figurative language,” “figure of speech,” “hyperbole,” “humor”, “idioms,” “indirect request,” “indirect speech,” “ironic,” “irony,” “jokes,” “lying,” “metaphor,” “metonymy,” “narrative,” “non-literal language,” “oxymoron,” “paradox,” “personification,” “platitude,” “pragmatics,” “prosody,” “proverbs,” “pun,” “sarcasm,” “sarcastic,” “saying,” “speech act,” “synecdoche,” “text coherence,” “text comprehension,” “understatement”, “fMRI,” “brain,” and “neuroimaging.” Note that we defined “non-literal language” as encompassing any linguistic input whose meaning cannot be reduced to

the literal meanings of its parts. As a result, this definition both includes paradigmatic phenomena like metaphors and idioms and extends to phenomena such as establishing text coherence/constructing broader thematic meaning for a narrative text (e.g., Johnson-Laird, 1983; Mason & Just, 2009), as well as prosody.

We additionally examined the reference lists of past neuroimaging meta-analyses on non-literal language processing to minimize the possibility of missing relevant studies (Ferstl et al., 2008; Bohrn et al., 2012; Rapp et al., 2012; Vartanian, 2012; Vrticka et al., 2013; Lisofsky et al., 2014; Yang, 2014; Yang & Shu, 2016; Reyes-Aguilar et al., 2018; Farkas et al., 2021). (Note that we had originally planned to use the NeuroSynth database of extracted peaks (Yarkoni et al., 2011); however, we discovered that NeuroSynth automatically extracts peaks from SPM-style tables in published papers, and such tables frequently include peaks for contrasts in both directions: e.g., in our case, non-literal>literal and literal>non-literal. As a result, we opted to manually extract peaks from only non-literal>literal contrasts.)

260 articles were selected for full-text screening based on the content of their abstracts. The selection criteria included: (1) fMRI (not PET or MEG/EEG) was used; (2) participants were neurotypical, and not aging, adults; (3) participants were native speakers of the language in which the experiment was conducted; (4) a standard whole-brain random effects analysis (Holmes & Friston, 1998) was performed; (5) activation peaks were reported in Talairach (Talairach & Tournoux, 1988) or Montreal Neurological Institute (MNI) (Evans et al., 1993) coordinate systems; and (6) contrasts targeted non-literal language comprehension, in the listening or reading modality, versus a (literal) linguistic baseline (cf. coarser-grain contrasts like non-literal language processing vs. fixation). The 74 studies that satisfied these criteria were published between 2001 and 2021 and targeted ten linguistic phenomena (**Table 1**). Across these

74 studies, 825 activation peaks (from 102 contrasts, between 1 and 4 contrasts per study) were extracted for analysis. Participants included 1,430 individuals (between 8 and 39 per study; *M* = 19.2) aged 18 to 55 (*M* = 24.8 years), 55% female (see **SI Table 1** for further details). **Table 1** summarizes the distribution of studies and peaks across the ten phenomena.

Linguistic phenomenon	Number of studies	Number of peaks
Humor	8	121
Idiom	6	54
Indirect Speech	10	160
Irony	10	68
Metaphor	19	201
Metonymy	2	10
Prosody	4	64
Proverb	1	9
Sarcasm	2	10
Text Coherence	13	128
TOTAL	**74 unique studies**	**825**

Table 1. Distribution of studies and peaks across linguistic phenomena. (One of the 74 studies included separate contrasts for non-sarcastic irony and sarcasm and is therefore counted twice in the table.) Studies on prosody that focused on the role of emotionally charged prosodic cues were excluded.

The probabilistic functional atlases for the three brain networks of interest

The individual activation peaks and the clusters derived from these peaks via the standard activation likelihood estimation (ALE) analysis (e.g., Turkeltaub et al., 2002; Eickhoff et al., 2009), as described below (section Activation likelihood estimation (ALE) analysis), were evaluated with respect to probabilistic functional atlases for three candidate networks of interest: the language network, the Theory of Mind (ToM) network, and the Multiple Demand (MD) network. For each network, an activation overlap map was created by overlaying a large (n>100) number of individual binarized activation maps for the 'localizer' task targeting each network, as described below (this is the first step in the Group-Constrained Subject-Specific analytic approach, as described in Fedorenko et al., 2010 and Julian et al., 2012). To account for inter-

individual variability in the overall level of activation, we selected in each individual the top 10% most localizer-responsive voxels across the brain (fixed-statistical-threshold approaches yield near-identical results; Lipkin et al., 2022). Specifically, we sorted the *t*-values for the relevant contrast and took 10% of voxels per participant with the highest values. In the resulting activation overlap map, the value in each voxel represents the number of participants for whom that voxel belongs to the top 10% most localizer-responsive voxels. These values—turned into proportions by dividing each value by the total number of participants that contribute to the overlap map—can then be used to estimate the probability that a given voxel belongs to the target network. Consider the extreme cases: if a voxel does not belong to the top 10% most localizer-responsive voxels in *any* participant, that voxel is extremely unlikely to belong to the network of interest, and if a voxel belongs to the top 10% most localizer-responsive voxels in *every* participant, that voxel is extremely likely to belong to the network of interest. In practice, for brain networks that support high-level cognitive functions and fall within the association cortex, network probability is unlikely to ever be 1 because of high inter-individual variability in the precise locations of these networks (e.g., Frost & Goebel, 2011; Tahmasebi et al., 2012), as discussed above. Our degree of confidence in assigning a voxel to a network will therefore be constrained by the maximum inter-subject overlap value for that network.

The task used to localize the ***language network*** is described in detail in Fedorenko et al. (2010) and targets brain regions that support high-level language processing, including phonological, lexical-semantic, and combinatorial (semantic and syntactic) processes (e.g., Fedorenko et al., 2010, 2012, 2016, 2020; Bautista & Wilson, 2016; Blank et al., 2016; Regev et al., 2021). It also identifies right-hemisphere homotopes of the left-hemisphere language regions (e.g., Mahowald & Fedorenko, 2016), which have been proposed to play a role in non-literal

language comprehension / pragmatic reasoning (e.g., Joanette et al.,1990; Kuperberg et al., 2000; Mashal et al., 2005; Coulson & Williams, 2005; Diaz & Hogstrom, 2011; Eviatar & Just, 2006). Briefly, we used a reading task that contrasted sentences (the critical condition) and lists of unconnected, pronounceable nonwords (the control condition; **Figure 1**) in a standard blocked design with a counterbalanced condition order across runs. By design, this localizer contrast subtracts out lower-level perceptual (speech or reading-related) and articulatory motor processes (see Fedorenko & Thompson-Schill, 2014 for discussion) and has been shown to generalize across materials, tasks, visual/auditory presentation modality, and languages (e.g., Fedorenko et al., 2010; Fedorenko, 2014; Scott et al., 2017; Ivanova et al., 2020; Chen et al., 2021; Malik-Moraleda, Ayyash et al., 2021). Further, this network emerges robustly from task-free naturalistic data (e.g., Braga et al., 2020; see also Blank et al., 2014, Paunov et al., 2018). Participants read the stimuli one word/nonword at a time in a blocked design, with condition order counterbalanced across runs. Each sentence/nonword sequence was followed by a button-press task to maintain alertness. A version of this localizer is available from https://evlab.mit.edu/funcloc/download-paradigm, and the details of the procedure and timing are described in **Figure 1** and **Table 2**. The probabilistic functional atlas used in the current study (Language Atlas (LanA); Lipkin et al., 2022) was constructed using data from 806 participants, and the voxel with the highest network probability had a value of 0.82 (i.e., belonged to the top 10% of most language-responsive voxels in 82% of participants).

The task used to localize the ***ToM network*** is described in detail in Saxe and Kanwisher (2003) and targets brain regions that support reasoning about others' mental states. Briefly, we used a task based on the classic false belief paradigm (Wimmer & Perner, 1983) that contrasted verbal vignettes about false beliefs (e.g., a protagonist has a false belief about an object's

location; the critical condition) versus linguistically matched vignettes about false physical states (physical representations depicting outdated scenes, e.g., a photograph showing an object that has since been removed; the control condition; **Figure 1**). By design, this localizer focuses on ToM reasoning to the exclusion of affective or non-propositional aspects of mentalizing and has been shown to generalize across materials (verbal and non-verbal), tasks, and visual/auditory presentation modality (e.g., Gallagher et al., 2000; Saxe & Kanwisher, 2003; Saxe et al., 2006; Jacoby et al., 2016). Furthermore, this network emerges robustly from task-free naturalistic data (e.g., Braga & Buckner, 2017; DiNicola et al., 2020; see also Paunov et al., 2018). Participants read the vignettes one at a time in a long-event-related design, with condition order counterbalanced across runs. Each vignette was followed by a true/false comprehension question. A version of this localizer is available from http://saxelab.mit.edu/use-our-efficient-false-belief-localizer, and the details of the procedure and timing are described in **Figure 1** and **Table 2**. The probabilistic functional atlas used in the current study (unpublished data from the Fedorenko lab) was constructed using data from 198 participants, and the voxel with the highest network probability had a value of 0.88.

The task used to localize the ***MD network*** is described in detail in Fedorenko et al. (2013) (see also Blank et al., 2014) and targets brain regions that are sensitive to general executive demands. Briefly, we used a spatial working memory task that contrasted a harder and an easier condition. On each trial, participants saw a 3 x 4 grid and kept track of eight (the critical, harder, condition) or four (the control, easier, condition; **Figure 1**) locations that were sequentially flashed two at a time or one at a time, respectively. Participants indicated their memory for these locations in a two-alternative, forced-choice paradigm via a button press. Feedback was provided after every trial. There is ample evidence that demanding tasks of many kinds activate this

network (e.g., Duncan & Owen 2000; Fedorenko et al. 2013; Hugdahl et al. 2015; Shashidara et al., 2019; and Assem et al., 2020a). Furthermore, this network emerges robustly from task-free naturalistic data (e.g., Assem et al., 2020a; Braga et al., 2020; see also Blank et al., 2014, Paunov et al., 2019). Hard and easy conditions were presented in a blocked design, with condition order counterbalanced across runs. This localizer is available from the authors upon request and the details of the procedure and timing are described in **Figure 1** and **Table 2**. The probabilistic functional atlas used in the current study (MDAtlas; Lipkin et al., in prep.) was constructed using data from 691 participants, and the voxel with the highest network probability had a value of 0.75. (It is worth noting that the three networks of interest differ somewhat in the range of their non-zero network probability values: language = 0.001-0.82, ToM = 0.005-0.88, and MD = 0.001-0.75. We decided against normalizing these values for each map (so that the highest network probability would be set to 1, or to the highest value observed across any of the three networks) because doing so would obscure meaningful differences in the likelihood that a given voxel belongs to each network.)

Critically, these three networks are robustly spatially and functionally dissociable within individuals in both naturalistic (e.g., Blank et al., 2014; Paunov et al., 2018; Braga et al., 2020) and task-based fMRI paradigms despite their close proximity to each other within the association cortex. The ***language network*** selectively supports linguistic processing, showing little or no response to diverse executive function tasks (e.g., Fedorenko et al., 2011; Monti et al., 2012; see Fedorenko & Blank, 2020 for a review) and mentalizing tasks (Deen et al., 2015, Paunov, 2018; Paunov et al., 2021; Shain, Paunov, Chen et al., 2022). The data from the patient literature mirrors this selectivity: damage to the language network does not appear to lead to difficulties in executive or social processing (see Fedorenko and Varley, 2016 for a review). The ***ToM network***

selectively supports social cognition, showing little or no response to linguistic input without mental state content (Deen et al., 2015; Paunov et al., 2021; Deen & Freiwald, 2021; Shain, Paunov, Chen et al., 2022) or to executive demands (Saxe et al., 2006; Scholz et al., 2009; Willems et al., 2010). Finally, the ***MD network*** supports diverse executive demands (e.g., Duncan, 2010, 2013; Assem et al., 2020a; Smith et al., 2021) and is linked to fluid reasoning ability (e.g., Woolgar et al., 2010; Assem et al., 2020b), but plays a limited role in language comprehension once task demands are controlled for (e.g., Diachek, Blank, Siegelman et al., 2020; Shain, Blank et al., 2020; Wehbe et al., 2021; see Fedorenko & Shain, 2021 for a review), and in social cognition (e.g., Willems et al., 2010). Because of the selective relationship between each of the three networks and a particular set of cognitive processes, activity in these networks can be used as an index of the engagement of the relevant processes (see e.g., Mather et al., 2013, for discussion), circumventing the need for precarious reverse inference from anatomical locations to function (e.g., Poldrack et al., 2006, 2011; Fedorenko, 2021). This is the approach that is adopted in studies that rely on functional localization (e.g., Brett et al., 2002; Saxe et al., 2006; Fedorenko, 2021). Here, we extend this general logic to the meta-analysis of group-level activation peaks by leveraging information about the landscape of each network of interest based on probabilistic functional atlases for the relevant localizer tasks.

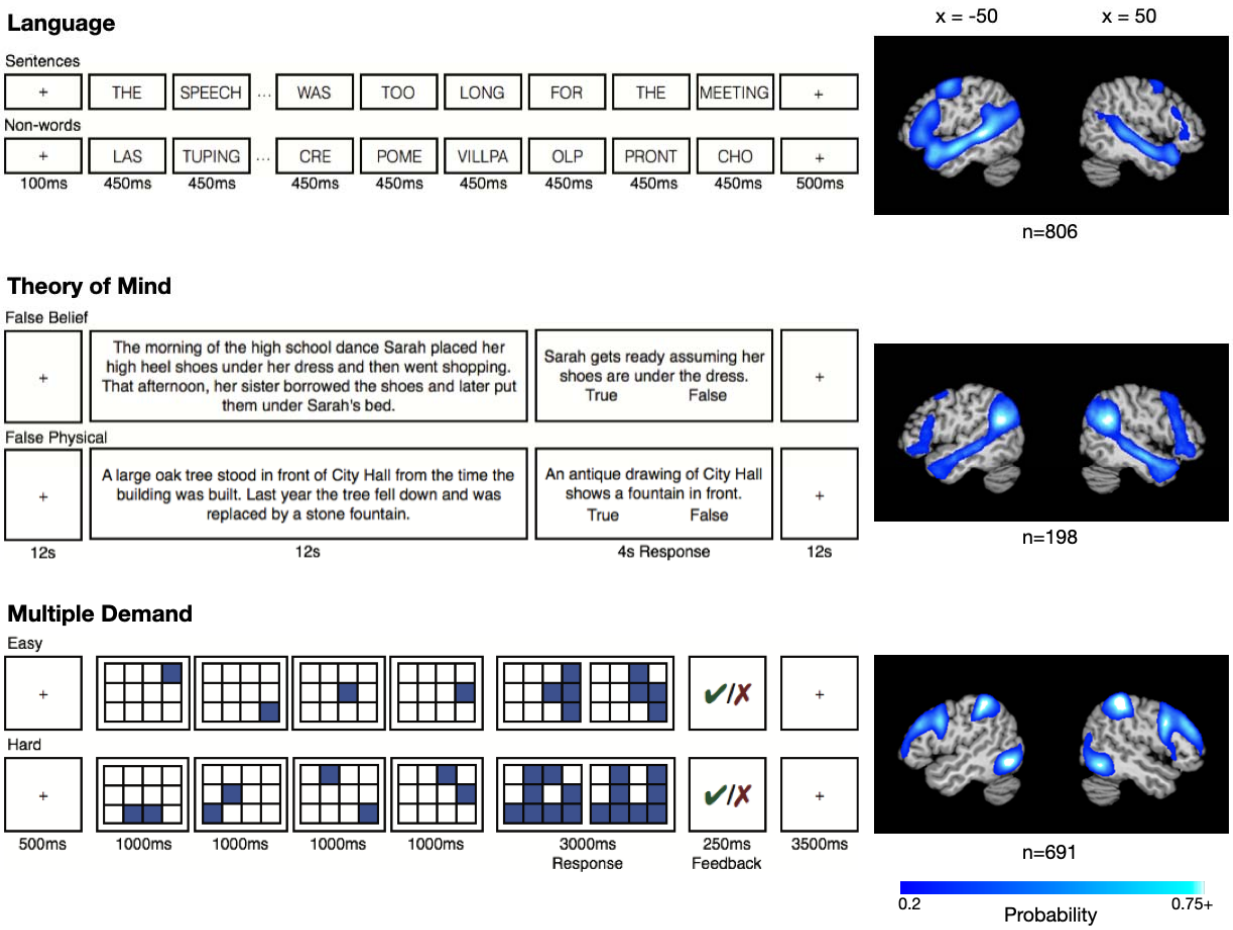


Fig. 1. The three functional localizer paradigms (language, ToM, MD) and the resulting probabilistic functional atlases. The maps illustrate, for each voxel, the proportion of participants for whom that voxel belongs to the top 10% of localizer-responsive voxels.

	Language	ToM	MD
Design	Blocked	Long-event-related	Blocked
Length of trial	4.8-18 s	14 s	8-9 s
Trials per block	1-5	N/A	2-4
Blocks/events per run	12-18	10	10-12
Blocks/events per condition per run	4-16	5	5-6
Length of run	336-504 s	272 s	288-448 s
Runs	2-8	1-2	2-4
Versions of the design	10	1	3

Table 2. Details on the design and timing of the language, ToM and MD localizer tasks.

Extracting network probabilities for the individual-study peaks

Prior to analysis, activation peaks that were reported in Talairach space were converted to the MNI space using icbm2tal transform SPM conversion in GingerALE 3.0.2 (Eickhoff et al., 2009, 2012). The MNI coordinates of the 825 activation peaks included in the dataset were tested against the probabilistic functional atlases (created as described above, section The probabilistic functional atlases for the three brain networks of interest) for each of the three functional networks of interest. For each coordinate, three network probability values were extracted (one from each functional atlas). (Note that because of the variability in the precise locations of functional areas and the proximity of the three networks to each other in parts of the association cortex, many voxels have non-zero values for more than a single network (despite the fact that there is little to no overlap among the networks in individual participants)—this is precisely the argument against traditional group-averaging analyses in fMRI; e.g., Fedorenko & Blank, 2020; DiNicola & Buckner, 2021; Fedorenko, 2021; Gordon & Nelson, 2021; Gratton & Braga, 2021; Smith et al., 2021).

Activation likelihood estimation (ALE) analysis

In addition to examining the individual-study activation peaks, we used the GingerALE software (Eickhoff et al., 2009, 2012) to perform a traditional fMRI meta-analysis via the activation likelihood estimation (ALE) method (Turkeltaub et al., 2002). ALE identifies regions that show consistent activation across experiments (Eickhoff et al., 2009, 2012). The clusters that ALE yields should be less noisy than the individual-study activation peaks, especially when multiple-comparison correction is not appropriately applied and participant sample sizes are small in individual studies (e.g., Genovese et al., 2002; Eklund et al., 2016; Chen et al., 2018).

The 74 studies in our dataset yielded 825 activation peaks (**Table 1;** see **SI Table 1** for a complete list of studies), but 39 peaks (fewer than 5%) fell outside the MNI template used by GingerALE and were therefore excluded, leaving 786 peaks. For each study, a map was created in which each voxel in the MNI space received a modeled activation score. Modeled activation scores reflect the likelihood that significant activation for one particular experiment was observed at a given voxel. The 74 modeled activation maps were then unified, generating an ALE value for each voxel.

Significance was assessed by comparing the observed ALE values to a null distribution that was generated by repeatedly calculating ALE values using randomly placed activation peaks (1,000 permutations). A cluster-forming threshold of $p<0.001$ uncorrected identified contiguous volumes of significant voxels ("clusters"), and clusters that survived a cluster-level family-wise error (FWE) of $p<0.05$ were considered significant. Cluster-level FWE correction has been argued to be the most appropriate correction for ALE, as it minimizes false positives while remaining more sensitive to true effects in comparison to other correction methods (Eickhoff et al., 2016).

Applying ALE to the 74 experiments yielded six significant clusters that were located in the left frontal and temporal cortex, and in the left amygdala (**SI Figure 1**). The clusters varied in size from 176 to 1,695 voxels.

Critical analyses

We asked two key research questions: the first, critical question examines the locations of the activation peaks with respect to the *three brain networks of interest*: the language network, the ToM network, and the MD network, and the second question is motivated by the prior claim about the privileged role of the *right hemisphere* in non-literal language processing (e.g., Winner & Gardner, 1977). Note that in all the analyses, we collapse the data from across the ten phenomena. We adopted this approach because i) none of the phenomena have a sufficient number of studies to enable meaningful phenomenon-level examination, and ii) there is currently a lack of consensus about the ways to carve up the space of non-literal phenomena (see Discussion). However, in an exploratory analysis, we examined the contribution of different phenomena to the peaks/clusters that load on the functional networks.

Q1: How are the activation peaks from prior fMRI studies of non-literal language comprehension distributed across the language, ToM, and MD networks?

Analysis of individual-study peaks

We evaluated the locations of the activation peaks with respect to our three brain networks of interest (language, ToM, MD). Almost all the peaks—820/825 (99%)—were observed in voxels with a non-zero network probability in at least one functional atlas. To test whether the network probabilities associated with the activation peaks differed between the

networks, we developed the following statistical procedure (note that for all analyses, we excluded the 13 peaks that fell on the cortical midline, i.e., had an x coordinate value of 0, given that we wanted to examine each network in each hemisphere separately, which left 812 peaks for analysis):

1) For each of the 812 peaks, we calculated the difference in network probabilities between each pair of networks (language vs. ToM, language vs. MD, and ToM vs. MD), and recorded the median (across peaks) of these difference values for each of the three contrasts in each hemisphere separately. (We used median instead of mean values to better capture the central tendencies given the skewness of the distributions.)
2) We then created 3D maps from which random peak sets could be selected (against which the location of true observed peaks could be evaluated, as described in Step 4 below). Specifically, we created a map for each hemisphere whereby we removed the voxels whose locations were associated with a network probability of 0 in all three functional atlases (for the language, ToM, and MD networks). This restriction was imposed so as to a) constrain the locations of the baseline voxel sets to the parts of the brain where true observed peaks were found (as noted above, almost all the peaks were observed in voxels that had a non-zero network probability in at least one network); and thus, b) to construct a more conservative test, making it more difficult to detect between-network differences.
3) In each hemisphere, the same number of peaks as in our dataset (n=490 in the LH, and n=322 in the RH) were then randomly sampled from the maps that were created in Step 2, and the median difference (across peaks) for each contrast (language vs. ToM, language vs. MD, ToM vs. MD) was computed, as in Step 1. This procedure was repeated 10,000 times, yielding an empirical null distribution of median network probability differences.
4) Finally, we compared the true median network probability differences against the distribution of network probability differences obtained in Step 3 to yield a significance value for each of the three contrasts within each hemisphere.

Analysis of activation likelihood estimation (ALE) clusters

We evaluated the locations of 1) the center peaks of the six significant ALE clusters, and 2) all voxels contained within each cluster with respect to our three brain networks of interest (language, ToM, MD). Importantly, whereas individual study peaks might be skewed by a single study with a particularly large number of activation peaks, ALE estimates incorporate the sample size of each study and are therefore less susceptible to this source of bias. To test whether the network probabilities differed between the networks, we evaluated the network probabilities of the voxels in each cluster and, for each cluster individually, subsequently performed the statistical procedure described above (in Analysis of individual-study peaks).

Q2: Do the activation peaks from prior fMRI studies of non-literal language comprehension exhibit a right hemisphere bias?

Analysis of individual-study peaks

To test whether the two hemispheres differed in the number of the activation peaks, we ran two logistic mixed effect regression models using the "lme4" package in R. The first tested the full set of 812 activation peaks (excluding the 13 peaks that fell on the cortical midline), and the second focused on the subset of the 812 peaks (n=636) that exhibited a network probability of 0.10 or greater in at least one network. The second model was included to ensure that the results are not driven by a subset of peaks that do not load strongly on any of the three networks of interest. For both models, the following formula was used, which included a random intercept for experiment (n=74):

$$\textit{Location of a peak in LH or RH} \sim 1 + (1|\textit{Experiment})$$

Analysis of activation likelihood estimation (ALE) clusters

We examined the locations of the six significant ALE clusters with respect to hemisphere.

Results

Non-literal language comprehension draws primarily on the language and ToM networks

Across the 812 non-midline peaks, the highest network probability was observed for the ToM network functional atlas (maximum=0.86, mean=0.19, SD=0.18, median=0.12; 802 total peaks with nonzero network probabilities), followed by the language atlas (maximum=0.80, mean=0.17, SD=0.17, median=0.10; 807 nonzero peaks), and the MD atlas (maximum=0.65, mean=0.11, SD=0.14, median=0.04; 806 nonzero peaks). In both hemispheres, the median non-zero network probabilities across the activation peaks numerically exceeded the median non-zero probability values across the functional atlases as a whole: language atlas (LH: peaks=0.12, atlas=0.04; RH: peaks=0.07, atlas=0.04), ToM (LH: peaks=0.13, atlas=0.04; RH: peaks=0.13, atlas=0.04), and MD (LH: peaks=0.03, atlas=0.02; RH: peaks=0.04, atlas=0.02). See **Figure 2** for a depiction of the network probability distributions of the individual-study peaks for each network.

We then examined between-network differences in network probabilities. Among the 490 LH peaks, the magnitude of the median difference in network probabilities was highest between the language and MD networks (language vs. MD = 0.07, language vs. ToM = 0.02, ToM vs. MD = 0.04). In the RH (322 peaks), the magnitude of the median difference in network probabilities was highest between the ToM and MD networks (ToM vs. MD = 0.04, language vs. ToM = 0.02, language vs. MD = 0.02). Our permutation analysis (see Methods) revealed that the

LH activation peaks were located more centrally in the language network than in either the ToM or the MD network (both ps<0.0001) and more centrally in the ToM network than the MD network (p<0.0001). The RH activation peaks were located more centrally in the ToM network than either the language or the MD network (both ps<0.0001) and more centrally in the language network than the MD network (p<0.0001) (see **Figure 2**). Importantly, this analysis controls for network lateralization (language=LH, ToM=RH) by generating an empirical null distribution of probability differences in each hemisphere separately. Higher network probabilities observed in the language network among the LH voxels and in the ToM network among the RH voxels therefore exceed what would be expected given the (already-skewed) distribution of network probabilities in each hemisphere.

With respect to the ALE clusters (see **Table 3** and **SI Figure 1**), the center peaks of four clusters, including the largest cluster, exhibited the highest network probabilities in the language atlas. Two of these were located in the left temporal lobe (primarily within the superior and middle temporal gyri), one within the left inferior frontal gyrus, and one within the left amygdala (although note that the network probability for the amygdala peak is overall relatively low compared to the cortical peaks). The center peaks of two additional clusters exhibited the highest network probabilities in the ToM atlas. One of these clusters was located in the left medial frontal gyrus and one in the left superior and middle temporal gyri.

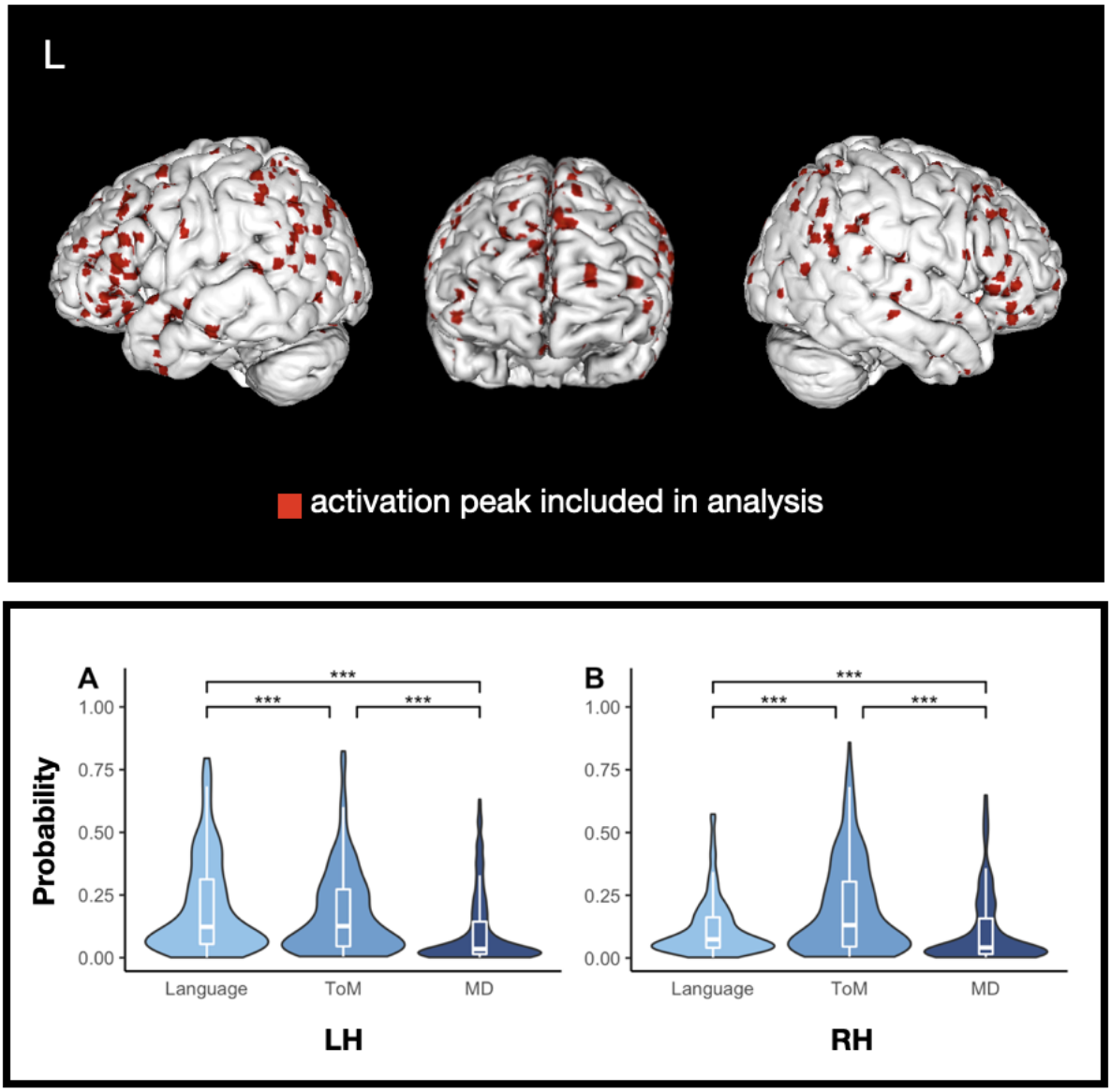


Fig. 2. Individual peaks. The top panel displays individual activation peaks plotted on the smoothed (width = 2; kernel size = 3) cortical surface of a high-resolution structural MRI scan in MNI space. The bottom panel displays the distribution of nonzero network probabilities associated with all individual peaks, by hemisphere. Significance of median differences between networks was assessed via a permutation test; all ps < .0001.

Analysis of the network probabilities associated with the full set of voxels comprising each ALE cluster yielded similar results (**Table 3**). The voxels in the four clusters whose center peaks had the highest network probabilities in the language atlas also exhibited the greatest

median network probabilities in the language atlas, whereas the voxels in the two clusters whose center peaks had the highest network probabilities in the ToM atlas also exhibited the greatest *median network probabilities* in the ToM atlas. Results from the permutation analysis supported this apparent distinction between the four "language" clusters and the two "ToM" clusters: voxels comprising the language clusters had significantly higher network probabilities in the language network than in the ToM and MD networks, and voxels comprising the ToM clusters had significantly higher network probabilities in the ToM network than in the language and MD networks (all ps<0.0001). These results are displayed in **Figure 3**.

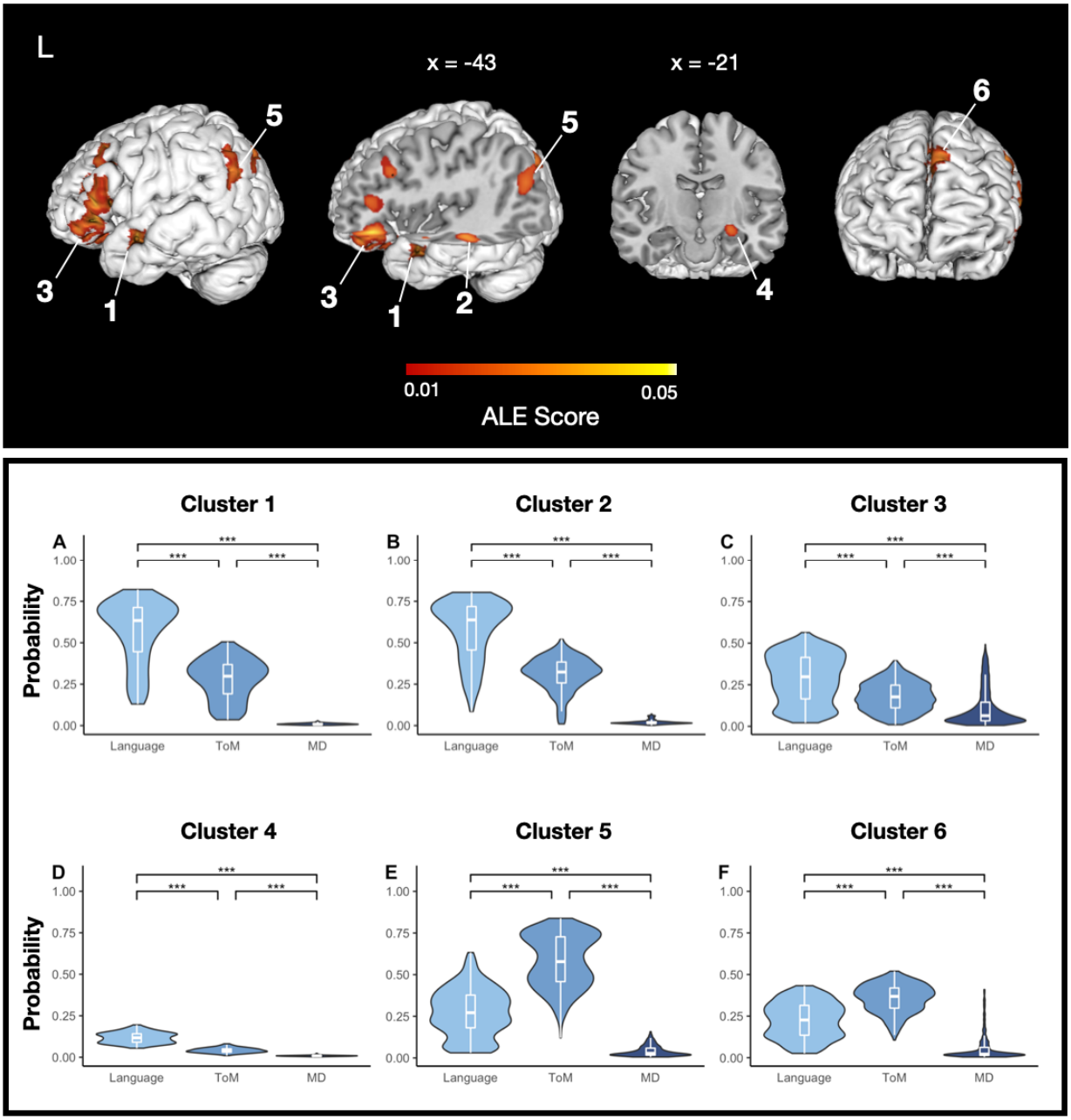


Fig. 3. Activation likelihood estimation (ALE) results. The top panel displays ALE scores associated with each of the 6 significant ALE clusters, plotted on the smoothed (width = 2; kernel size = 3) cortical surface of a high-resolution structural MRI scan in MNI space (cluster-forming threshold = $p<0.001$ uncorrected, cluster-level family-wise error (FWE) = $p<0.05$, 1000 permutations). The bottom panel displays the distribution of nonzero network probabilities associated with all voxels contained within each cluster. Significance of median differences between networks was assessed via a permutation test; all ps < .0001.

Finally, in an exploratory analysis, we evaluated the phenomena that contributed to the six significant ALE clusters (see **SI Table 2**). No clear differentiation in terms of phenomena that contribute to the language vs. the ToM clusters was apparent.

Cluster	Size (in voxels)	Hemisphere	Broad Anatomical Area	x	y	z	Language	ToM	MD
1	283	LH	STG	-54	-1	-16	**0.743 / 0.634**	0.348 / 0.298	0.001 / 0.009
2	357	LH	MTG	-55	-33	-3	**0.738 / 0.638**	0.394 / 0.323	0.016 / 0.016
3	1,695	LH	IFG	-47	26	3	**0.424 / 0.298**	0.273 / 0.177	0.030 / 0.064
4	176	LH	Amygdala	-23	-10	-17	**0.136 / 0.117**	0.030 / 0.040	0.004 / 0.009
5	528	LH	STG	-49	-63	29	0.197 / 0.272	**0.611 / 0.578**	0.016 / 0.028
6	454	LH	MedFG	-3	51	31	0.355 / 0.227	**0.449 / 0.369**	0.012 / 0.022

Table 3. Network probabilities for the six significant ALE clusters. The number of voxels in each cluster is listed under "Size (in voxels)". The "Broad Anatomical Area" column lists the macroanatomical region that overlaps the most with each cluster (as determined by the ALE). In the "Language," "ToM," and "MD" columns, the first value represents the network probability associated with that cluster's center peak and the second value represents the median network probability of all voxels contained in the cluster. In bold, we highlighted for each cluster the network that has the highest network probabilities.

No evidence of a RH bias for non-literal language comprehension

Of the full set of 812 peaks (excluding the 13 that fell on the cortical midline), 322 peaks (~40%) were located in the RH, and 490 (~60%) were located in the LH (**Figure 4**). This difference was highly reliable ($b = -0.44$, $SE = 0.1$, $z = -4.42$, $p<0.001$). Similarly, of the subset of the 636 peaks with network probabilities of at least 0.10 in at least one network of interest, 245 peaks (~39%) were located in the RH, and 391 (~61%) were located in the LH (**Figure 4**). This difference was highly reliable ($b = -0.50$, $SE = 0.11$, $z = -4.5$, $p<0.001$). The LH bias was also mirrored in the ALE analysis: all six clusters fell in the LH. These findings provide additional support for the notion that the LH is an important contributor—seemingly more important than the RH—to non-literal language processing (Bohrn et al., 2012; Rapp et al., 2012; Reyes-Aguilar et al., 2018).

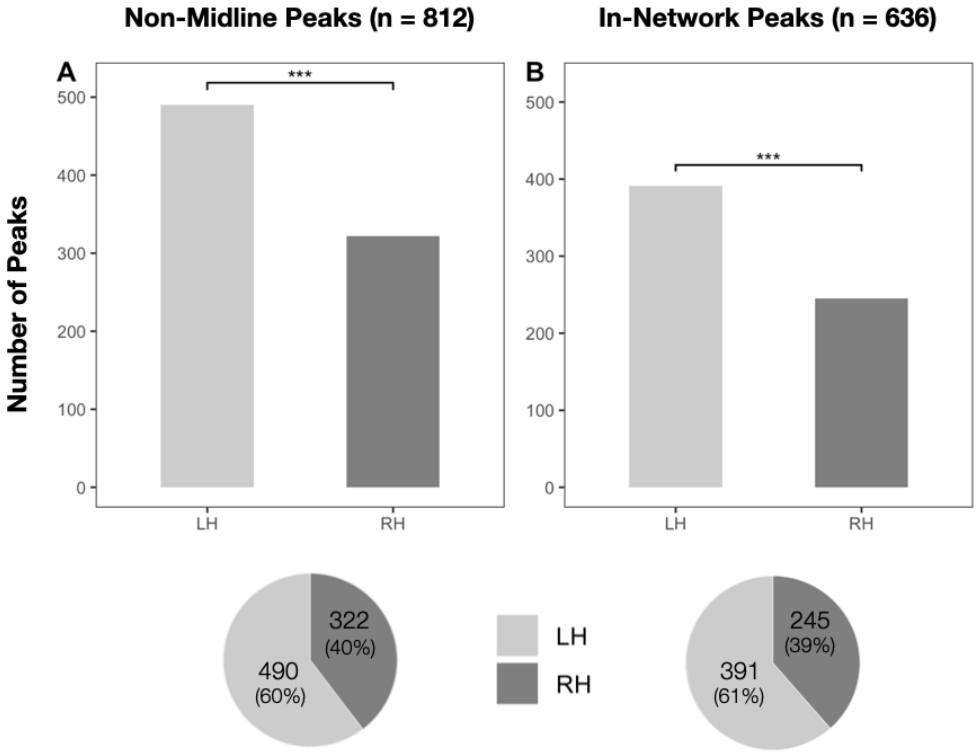


Fig. 4. Peaks by hemisphere. In the top panel, A displays the number of non-midline peaks in the left vs. right hemisphere. B displays the left vs. right breakdown of non-midline peaks with an network probability of at least 0.10 in at least one network. Significance was assessed via logistic mixed effects regression; all ps < .001. Pie charts displaying the proportion of activation peaks in the left vs. right hemisphere are shown in the bottom panel.

Discussion

To illuminate the cognitive and neural bases of non-literal language comprehension, we performed a meta-analysis of group-level activation peaks from past fMRI studies. Specifically, we developed a novel approach that leverages 'probabilistic functional atlases' for three brain networks that have been implicated in non-literal comprehension (the language network, the Theory of Mind network, and the Multiple Demand network). The atlases are built using large numbers of individual activation maps for extensively validated 'localizer' tasks (e.g., Saxe et al., 2006; Fedorenko, 2021) and provide estimates of the probability that a location in the

common brain space belongs to a particular network. Because each of these networks has been rigorously characterized and selectively linked to particular cognitive processes in past work, activation peak locations therein can be interpreted as evidence for the engagement of the relevant process(es) (Mather et al., 2013). This approach is therefore superior to the traditional meta-analytic approach where activation peaks or ALE clusters (Turkeltaub et al., 2002) are interpreted solely based on their anatomical locations or in relation to the results of additional meta-analyses.

The three networks that we examined include i) the language-selective network, which supports literal comprehension, including lexical and combinatorial operations (Fedorenko et al., 2020), ii) the Theory of Mind (ToM) network, which supports social inference, including mentalizing (Saxe & Kanwisher, 2003), and iii) the Multiple Demand (MD) network, which supports executive control (Duncan, 2010). We asked two research questions. The first, critical question concerned the distribution of peaks across the networks. The individual peaks and ALE clusters tended to fall in the language and ToM networks, but not the MD network. The second question, motivated by past patient investigations that have linked non-literal comprehension impairments to RH damage (e.g., Winner & Gardner, 1977), asked whether the peaks were more likely to fall in the RH. In line with past meta-analyses of non-literal language (Bohrn et al., 2012; Rapp et al., 2012; Reyes-Aguilar et al., 2018), we found that the peaks/clusters fell primarily in the LH. Below, we discuss several issues that these results inform, and highlight some outstanding questions and some methodological implications.

The role of the ToM and language networks in non-literal comprehension

In line with past studies that have reported activations in putative ToM areas for non-literal phenomena (e.g., Spotorno et al., 2012; van Ackeren et al., 2012; Feng et al., 2017), some individual study activation peaks and two ALE clusters had a high probability of falling within the ToM network. What is the role of mentalizing—a capacity supported by the ToM network—in language comprehension?

Because linguistic inputs often underspecify intended meaning (Wittgenstein, 1953; Sperber & Wilson, 1986), language comprehension routinely requires inferences about communicative intent. The computation of such inferences has historically been a focus of the field of pragmatics (Grice, 1957, 1975). Some early proposals drew a sharp boundary between literal and non-literal/pragmatic processing (Grice, 1975; Searle, 1979). However, defining the scope of pragmatic inference has proven challenging, and many have questioned the divide between literal and inferred meaning, or between semantics and pragmatics (Jackendoff, 2002). But if no such boundary exists, does language understanding *always* recruit the ToM network, in addition to the language network, to enable inferences about communicative intent?

The notion of continuous ToM engagement during language comprehension does not seem *a priori* plausible (many phenomena requiring context-based inferences—lexical disambiguation or pronoun resolution—are so common that it would seem inefficient and unnecessary to constantly engage in full-blown mentalizing) and does not find empirical support (Deen et al., 2015; Paunov et al., 2021; Shain, Paunov, Chen et al., 2022). Yet, in some cases, the ToM network does appear to contribute to language comprehension (e.g., Spotorno et al., 2012). Delineating the precise conditions under which comprehension requires ToM resources remains an important goal for future work (Paunov et al., 2021). Brain-imaging investigations of diverse

non-literal phenomena using approaches with functionally localized ToM and language networks may offer some clarity. This approach could be complemented by experiments testing linguistic abilities in individuals with impaired ToM reasoning (e.g., Siegal et al., 1996; Happé et al., 1999) or in statistical language models (Devlin et al., 2019). The latter can reveal which non-literal phenomena can be handled by language models, and which might require additional machinery that approximates mental state inference (e.g., Hu et al., 2021).

Many individual study activation peaks and four ALE clusters had a high probability of falling within the language network. Of course, understanding both literal and non-literal utterances requires language processing, so both conditions should elicit a strong response in the language network. But why do non-literal conditions often elicit a stronger response? Differences in linguistic complexity might provide one explanation. The brain's language areas are sensitive to comprehension difficulty (Wehbe et al., 2021). Although studies that compare literal and non-literal conditions commonly match stimuli on some linguistic variables, this matching is often limited to word-level features, which do not account for potential differences in context-based processing difficulty (e.g., surprisal; Smith & Levy, 2011). Further, despite the robust dissociation between the language and ToM networks, Paunov et al. (2019) found that the two networks show reliable correlation during language processing. This inter-network interaction may lead increased ToM demands to additionally manifest in the language system via inter-network connections. Finally, the language network may actually support some pragmatic computations, although to argue for such effects, linguistic confounds and inter-network information leakage would need to be eliminated as possible explanations.

No evidence for the role of the MD network in non-literal comprehension

In contrast to the ToM and language networks, the Multiple Demand network does not appear to support non-literal comprehension: individual-study activation peaks and ALE clusters were least likely to fall in this network. Selecting the non-literal interpretation of an utterance may tax working memory (because multiple interpretations may be activated) or require inhibitory control (e.g., Gernsbacher & Robertson, 1999; Channon & Watts, 2003). However, recent work has shown that any such operation related to linguistic processing appears to be implemented within the language-selective system (Shain, Blank et al., 2020; see Fedorenko & Shain, 2021 for a review). Indeed, as discussed above, greater cognitive demands associated with non-literal interpretation may well explain the responses in the *language* system.

We suspect that neural activity observed during non-literal processing in putative executive control areas in past studies (e.g., AbdulSabur et al., 2014; Chan & Lavallee, 2015; Bosco et al., 2017) is due either to i) reliance on anatomical landmarks, which do not warrant functional interpretation in the association cortex (Poldrack, 2006; Fedorenko, 2021), or ii) extraneous task demands (see Diachek, Blank, Siegelman et al., 2020 for evidence that sentence comprehension only engages the MD network when accompanied by a secondary task, like a sentence judgment). Behavioral investigations that find correlations between executive abilities and non-literal interpretation abilities (e.g., Akbar et al., 2013; Caillies et al., 2014; Rints et al., 2015) are also likely affected by methodological issues, from small sample sizes to confounded experimental paradigms (Matthews et al., 2018). Indeed, recent large-scale investigations (Cardillo et al., 2020; Fairchild & Papafragou, 2021) argue against the role of executive functions in pragmatic ability.

Possible reasons for the inconsistencies regarding the hemispheric bias

One remaining puzzle concerns the difference between patient work, which has implicated the RH in non-literal language processing (e.g., Winner & Gardner, 1977; Myers & Linebaugh, 1981; Delis et al., 1983; Brownell et al., 1983; 1986; Van Lancker & Kempler, 1987; Brownell, 1988; Stemmer et al., 1994; Giora et al., 2000; Ferstl et al., 2005) and our results, which demonstrate a LH bias. As noted above, left-lateralized activations have also been reported in past studies (e.g., Rapp et al., 2004; Lee & Dapretto, 2006; Hillert & Burac□as, 2009; Piñango et al., 2015; Bosco et al., 2017) and meta-analyses (Bohrn et al., 2012; Rapp et al., 2012; Reyes-Aguilar et al., 2018; Farkas et al., 2021). One possibility is that no RH bias exists for non-literal language processing. Both hemispheres contribute, perhaps with the LH contributing more strongly, as per the standard LH language bias (e.g., Geschwind, 1970). Because lexical and grammatical impairments resulting from LH damage are more salient and devastating, non-literal comprehension difficulties go unnoticed. In contrast, RH damage, which does not strongly affect basic language processing, may make apparent more subtle linguistic impairments, leading to the apparent RH bias for non-literal comprehension. In line with this possibility, several studies have reported that patients with LH damage show similar, or even greater, deficits in non-literal comprehension compared to RH-damaged patients (e.g., Tompkins, 1990; Giora et al., 2000; Zaidel et al., 2002; Gagnon et al., 2003; Klepousniotou & Baum, 2005; Ianni et al., 2014; Cardillo et al., 2018; Klooster et al., 2020).

Alternatively, the RH may indeed contribute more strongly than the LH to non-literal language comprehension, and past fMRI studies may have failed to detect the RH effects due the generally low sensitivity of the group-averaging analytic approach (Nieto-Castañón & Fedorenko, 2012). This possibility seems unlikely: if the RH were more active than the LH

during non-literal comprehension, then RH activations should be easier to detect in a group analysis. This is because stronger responses would be more spatially extensive in individual brains and thus more likely to lead to overlap at the group level. (Further, using our probabilistic functional atlases, we did not find evidence for the possibility that functional areas in the RH are more variable in their locations across individuals and therefore less likely to emerge in a group analysis.) Future fMRI studies where areas of interest are identified via functional localizers, as well as intracranial stimulation studies, which enable spatially precise perturbations of neural activity, can provide important insight into the hemisphere-bias question.

The need to carve non-literal processing at its joints

We did not find any systematic patterns with regard to the phenomena that 'load' onto the language vs. the ToM network (**SI Table 2**). However, many phenomena had small numbers of data points. In general, the field would benefit from clear, formal hypotheses about differences in the cognitive processes that contribute to each non-literal phenomenon. Such hypotheses can then be tested using behavioral, brain imaging, and computational modeling approaches.

We would like to conclude with a methodological point. fMRI has been a critical tool in human cognitive neuroscience. However, two disparate approaches are currently in use: i) the traditional approach where brains are averaged voxel-wise in the common space (Friston et al., 1999), and ii) the subject-specific functional localization approach where regions of interest are identified using functional 'localizers' and critical effects are examined therein (Saxe et al., 2006). Until recently, comparing findings across these approaches—and thus establishing a cumulative research enterprise—has proven challenging. Probabilistic functional atlases, like the ones we

used in the current study (see also Dworetsky et al., 2021; Thirion et al., 2021), can bridge these two approaches by providing a common framework for functional areas/networks.

Supplementary Information

Authors	Year	Participants	Female	Age Range	Contrast	Contrast #	Peaks	Phenomenon
Adamczyk et al.	2017	20	9	n/a	Funny story punchline > Neutral story punchline	1	5	Humor
Bekinschtein et al.	2011	12	6	18-35	Jokes > Non-jokes	1	34	Humor
Campbell et al.	2015	24	14	n/a	Intended to be funny > Not funny linguistic cartoons	1	20	Humor
Chan & Lavallee	2015	27	16	21-29	Jokes > Non-jokes	1	6	Humor
Dai et al.	2017	27	12	21-30	Incongruity resolution jokes > Non-jokes	1	5	Humor
Dai et al.	*2017*	*27*	*12*	*21-30*	*Absurd jokes > Non-jokes*	*2*	*3*	*Humor*
Shibata et al.	2014	20	15	20-34	Humorous story punchline > Non-humorous story punchline	1	17	Humor
Tian et al.	2017	24	11	19-25	Humorous story > Non-humorous story	1	12	Humor
Wu et al.	2020	27	14	n/a	Humorous reppraisal > Ordinary reappraisal of image	1	19	Humor
Bendersky et al.	2021	21	10	19-48	Idiomatic expressions > Literal sentences	1	6	Idiom
Citron et al.	2019	23	12	19-36	Idiomatic sentences > Literal sentences	1	15	Idiom
Hillert & Buracas	2009	20	13	18-23	Literally non-interpretable idiomatic sentences > Literal sentences	1	5	Idiom
Hillert & Buracas	*2009*	*20*	*13*	*18-23*	*Literally interpretable idiomatic sentences > Literal sentences*	*2*	*6*	*Idiom*
Lauro et al.	2008	22	13	19-40	Idiomatic sentences > Literal sentences	1	10	Idiom
Yang et al.	2016	19	10	20-25	Transparent + opaque idioms > Non-idiomatic phrases	1	2	Idiom
Yang et al.	*2016*	*19*	*10*	*20-25*	*Opaque idioms > Non-idiomatic phrases*	*2*	*2*	*Idiom*
Zempleni et al.	2007	15	7	n/a	Idiomatic sentences > Literal sentences	1	8	Idiom
Bašnáková et al.	2015	20	16	n/a	Indirect responses > Direct responses	1	42	Indirect Speech
Bašnáková et al.	2014	28	23	n/a	Pooled indirect replies > Direct replies	1	12	Indirect Speech
Bašnáková et al.	*2014*	*28*	*23*	*n/a*	*Indirect informative replies > Direct replies*	*2*	*5*	*Indirect Speech*
Bašnáková et al.	*2014*	*28*	*23*	*n/a*	*Indirect face-saving replies > Direct replies*	*3*	*12*	*Indirect Speech*
Egorova et al.	2016	18	10	18-41	Request > Naming	1	11	Indirect Speech
Feng et al.	2017	19	8	n/a	Pooled indirect replies > Direct replies	1	7	Indirect Speech
Jang et al.	2013	23	9	n/a	Highly implicit answer > Explicit answer	1	11	Indirect Speech
Jang et al.	*2013*	*23*	*9*	*n/a*	*Highly implicit answer > Moderately implicit answer*	*2*	*1*	*Indirect Speech*
Jang et al.	*2013*	*23*	*9*	*n/a*	*Moderately implicit answer > Explicit answer*	*3*	*4*	*Indirect Speech*
Licea-Haquet et al.	2021	31	15	19-28	Speech acts > Control sentences	1	8	Indirect Speech
Prado et al.	2015	20	12	n/a	Disjunctive stories > Control stories	1	6	Indirect Speech
Shibata et al.	2011	15	4	n/a	Indirect reply > Literal reply	1	10	Indirect Speech
van Ackeren et al.	2012	13	13	18-24	Indirect reply > Pooled control conditions	1	26	Indirect Speech
van Ackeren et al.	2016	22	22	n/a	Indirect speech > Direct speech	1	5	Indirect Speech
Akimoto et al.	2014	25	14	18-23	Ironic statement > Literal statement	1	2	Irony
Bosco et al.	2017	18	6	19-27	Ironic sentence > Literal sentence	1	6	Irony

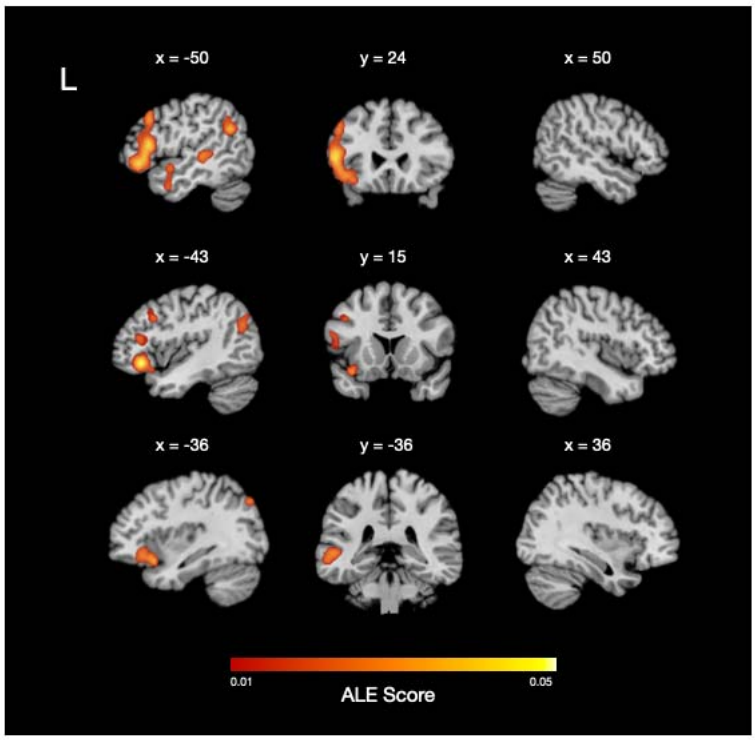
Filik et al.	2019	17	13	19-27	Non-sarcastic ironic sentence > Literal sentence	1	3	Irony
Herold et al.	2018	12	7	26-55	Non-sarcastic ironic sentence > Literal sentence	1	7	Irony
Obert et al.	2016	21	10	20-27	Ironic statement > Control statement	1	8	Irony
Rapp et al.	2010	15	15	21-53	Ironic sentences > Literal sentences	1	6	Irony
Shibata et al.	2010	13	3	20-29	Ironic sentences > Literal target sentences	1	4	Irony
Spotorno et al.	2012	20	12	n/a	Ironic sentences > Literal sentences	1	11	Irony
Varga et al.	2013	24	14	23-55	Ironic statement > Control statement	1	13	Irony
Varga et al.	*2013*	*24*	*14*	*23-55*	*Ironic statement with linguistic help > Control statement*	*2*	*5*	*Irony*
Wakusawa et al.	2007	38	17	18-38	Ironic sentence > Control sentence	1	2	Irony
Wang et al.	2006	12	6	23-33	Ironic sentence > Control sentence	1	4	Irony
den Ouden et al.	2016	16	15	21-34	Natural > Pooled flat prosody conditions	1	3	Linguistic Prosody
den Ouden et al.	*2016*	*16*	*15*	*21-34*	*Natural > Flat prosody in garden path sentences without cue*	*2*	*5*	*Linguistic Prosody*
Hellbernd & Sammler	2018	22	11	n/a	Clear prosody > Ambiguous prosody	1	24	Linguistic Prosody
Kristensen et al.	2013	24	18	18-24	Pitch accent > No pitch accent	1	22	Linguistic Prosody
Perrone-Bertolotti et al.	2013	24	12	19-34	Narrow contrastive focus > Broad/neutral focus	1	10	Linguistic Prosody
Ahrens et al.	2007	8	0	20-22	Conventional metaphor > Literal sentences	1	1	Metaphor
Bambini et al.	2011	9	4	n/a	Metaphors > Non-metaphors	1	10	Metaphor
Chen et al.	2008	14	4	n/a	Predicate metaphor > Motive literal sentence	1	6	Metaphor
Chen et al.	*2008*	*14*	*4*	*n/a*	*Predicate metaphor > Non-motive literal sentence*	*2*	*1*	*Metaphor*
Citron & Goldberg	2014	26	19	n/a	Conventional metaphors > Literal sentences	1	19	Metaphor
Citron et al.	2016	24	15	21-35	Metaphorical stories > Literal stories	1	69	Metaphor
Diaz & Hogstrom	2011	16	8	21-31	Metaphor > literal sentence	1	6	Metaphor
Lee & Dapretto	2006	12	6	23-35	Metaphorical word triad > Literal word triad	1	3	Metaphor
Mashal et al.	2007	15	7	21-31	Novel metaphor > Literal word pair	1	5	Metaphor
Mashal et al.	*2007*	*15*	*7*	*21-31*	*Conventional metaphor > Literal word pair*	*2*	*3*	*Metaphor*
Mashal et al.	2009	15	6	22-28	Novel metaphor > Literal sentences	1	1	Metaphor
Nagels et al.	2013	16	0	n/a	Simile > Control sentence	1	2	Metaphor
Obert et al.	2014	19	10	19-27	Metaphorical sentence > Literal sentence	1	3	Metaphor
Pomp et al.	2018	18	6	19-39	Olfactory metaphors > Literal paraphrases	1	2	Metaphor
Samur et al.	2015	20	16	18-27	Metaphorical sentences > Literal sentences	1	3	Metaphor
Schmidt & Seger	2009	11	6	19-42	Easy familiar metaphors > Literal sentences	1	11	Metaphor
Schmidt & Seger	*2009*	*11*	*6*	*19-42*	*Pooled metaphors > Literal sentences*	*2*	*6*	*Metaphor*
Schuil et al.	2013	20	10	18-25	Metaphor > Literal sentences	1	1	Metaphor
Shibata et al.	2012	24	5	21-29	Metaphor > Literal sentences	1	5	Metaphor
Shibata et al.	*2012*	*24*	*5*	*21-29*	*Simile > Literal sentences*	*2*	*9*	*Metaphor*

Shibata et al.	*2012*	*24*	*5*	*21-29*	*Metaphor + Simile > Literal Sentences*	*3*	*6*	*Metaphor*
Shibata et al.	2007	13	10	21-47	Metaphor > Literal sentences	1	5	Metaphor
Stringaris et al.	2007	11	0	n/a	Metaphor > Literal sentence	1	9	Metaphor
Yang et al.	2009	18	10	19-44	Novel metaphor > Literal sentence with informative content	1	7	Metaphor
Yang et al.	*2009*	*18*	*10*	*19-44*	*Novel metaphor > Literal sentence with redundant content*	*2*	*5*	*Metaphor*
Yang et al.	*2009*	*18*	*10*	*19-44*	*Conventional metaphor > Literal sentence with redundant content*	*3*	*3*	*Metaphor*
Piñango et al.	2016	16	7	18-30	Systematic metonymy > Literal sentences	1	1	Metonymy
Piñango et al.	*2016*	*16*	*7*	*18-30*	*Circumstantial metonymy > Literal sentences*	*2*	*1*	*Metonymy*
Rapp et al.	2011	14	12	n/a	Metonymy > Literal sentences	1	8	Metonymy
Yi et al.	2017	15	8	27-33	Proverbs > Literal sentences	1	3	Proverb
Yi et al.	*2017*	*15*	*8*	*27-33*	*Transparent proverbs > Literal sentences*	*2*	*3*	*Proverb*
Yi et al.	*2017*	*15*	*8*	*27-33*	*Opaque proverbs > Literal sentences*	*3*	*3*	*Proverb*
Filik et al.	2019	17	13	19-27	Sarcastic sentence > Literal sentence	1	6	Sarcasm
Matsui et al.	2016	21	10	19-29	Sarcastic sentence > Literal sentence	1	1	Sarcasm
AbdulSabur et al.	2014	18	11	20-32	Narrative comprehension > Nursery rhyme comprehension	1	12	Text Coherence
Chow et al.	2008	15	11	n/a	Predictive reading > Normal reading	1	3	Text Coherence
Chow et al.	*2008*	*15*	*11*	*n/a*	*Predictive reading > Normal reading at lowered threshold*	*2*	*5*	*Text Coherence*
Ferstl & von Cramon	2001	12	4	19-31	Coherent stories > Incoherent stories	1	2	Text Coherence
Ferstl & von Cramon	2002	9	5	22-27	Coherent stories > Incoherent stories	1	12	Text Coherence
Friese et al.	2008	13	7	n/a	Sentence containing inference > Paraphrase sentence	1	1	Text Coherence
Heidlmayr et al.	2020	25	17	18-35	Highly informative > Intermediately informative title	1	2	Text Coherence
Heidlmayr et al.	*2020*	*25*	*17*	*18-36*	*Highly informative > No title*	*2*	*7*	*Text Coherence*
Kuperberg et al.	2006	15	12	n/a	Intermediately related sentences > Highly related sentences	1	8	Text Coherence
Lin et al.	2018	39	18	n/a	Sentences with high integration demands > Low demands	1	13	Text Coherence
Martín-Loeches et al.	2008	23	11	19-39	Titled paragraph > Untitled paragraph	1	13	Text Coherence
Mason & Just	2011	14	7	n/a	Intentional inference-inducing sentence > Control sentence #2	1	17	Text Coherence
Mason & Just	*2011*	*14*	*7*	*n/a*	*Intentional inference-inducing sentence > Control sentence #3*	*2*	*5*	*Text Coherence*
Mason & Just	*2011*	*14*	*7*	*n/a*	*Physical > Control sentence #2*	*3*	*1*	*Text Coherence*
Mason & Just	*2011*	*14*	*7*	*n/a*	*All inference-inducing sentences > Control sentence #2*	*4*	*7*	*Text Coherence*
Virtue et al.	2006	17	12	18-28	Verb cued inference > Explicit sentences	1	1	Text Coherence
Virtue et al.	*2006*	*17*	*12*	*18-28*	*Coherence break cued inference > Explicit sentences*	*2*	*2*	*Text Coherence*
Virtue et al.	2008	19	10	19-34	Predictable inference condition > Explicit condition	1	6	Text Coherence
Virtue et al.	*2008*	*19*	*10*	*19-34*	*Unpredictable inference condition > Explicit condition*	*2*	*3*	*Text Coherence*
Yarkoni et al.	2008	29	17	18-32	Story > Scrambled (out of order) sentences within a story	1	8	Text Coherence

SI Table 1. All individual contrasts included in the meta-analysis. Multiple contrasts taken from the same publication are displayed in italics.

	1	**2**	**3**	**4**	**5**	**6**
Humor	3		8	3	5	
Idiom	1		5	2	1	1
Indirect Speech	3	4	20		7	11
Irony	5	4	8		2	3
Metaphor	1		19	2	6	6
Metonymy		1	1			1
Prosody		2	1	1	2	
Proverb		2	1			
Sarcasm			1			
Text Coherence	2	4	17	2	5	1

SI Table 2. Individual peaks contributing to the six significant ALE clusters by phenomenon. Clusters (numbers the same as in Table 3) are listed by column. Empty space indicates that zero peaks from a given phenomenon contributed to the cluster assigned to that column.



SI Figure 1. Main ALE results. ALE results for the 825 total peaks in our dataset, where peaks from each publication (n=74) are considered independent. ALE scores associated with each of the 6 significant ALE clusters are overlaid on a high-resolution structural MRI scan in MNI

space (cluster-forming threshold = $p<0.001$ uncorrected, cluster-level family-wise error (FWE) = $p<0.05$, 1000 permutations).

References

AbdulSabur, N. Y., Xu, Y., Liu, S., Chow, H. M., Baxter, M., Carson, J., & Braun, A. R. (2014). Neural correlates and network connectivity underlying narrative production and comprehension: A combined fMRI and PET study. *Cortex*, *57*, 107–127. https://doi.org/10.1016/j.cortex.2014.01.017

Adamczyk, P., Wyczesany, M., Domagalik, A., Daren, A., Cepuch, K., Błądziński, P., Cechnicki, A., & Marek, T. (2017). Neural circuit of verbal humor comprehension in schizophrenia—An fMRI study. *NeuroImage: Clinical*, *15*, 525–540. https://doi.org/10.1016/j.nicl.2017.06.005

Ahrens, K., Liu, H.-L., Lee, C.-Y., Gong, S.-P., Fang, S.-Y., & Hsu, Y.-Y. (2007). Functional MRI of conventional and anomalous metaphors in Mandarin Chinese. *Brain and Language*, *100*(2), 163–171. https://doi.org/10.1016/j.bandl.2005.10.004

Akbar, M., Loomis, R., & Paul, R. (2013). The interplay of language on executive functions in children with ASD. *Research in Autism Spectrum Disorders*, *7*(3), 494–501. https://doi.org/10.1016/j.rasd.2012.09.001

Akimoto, Y., Sugiura, M., Yomogida, Y., Miyauchi, C. M., Miyazawa, S., & Kawashima, R. (2014). Irony comprehension: Social conceptual knowledge and emotional response. *Human Brain Mapping*, *35*(4), 1167–1178. https://doi.org/10.1002/hbm.22242

Assem, M., Blank, I. A., Mineroff, Z., Ademoğlu, A., & Fedorenko, E. (2020). Activity in the fronto-parietal multiple-demand network is robustly associated with individual differences in working memory and fluid intelligence. *Cortex*, *131*, 1–16. https://doi.org/10.1016/j.cortex.2020.06.013

Assem, M., Glasser, M. F., Van Essen, D. C., & Duncan, J. (2020). A Domain-General Cognitive Core Defined in Multimodally Parcellated Human Cortex. *Cerebral Cortex (New York, NY)*, *30*(8), 4361–4380. https://doi.org/10.1093/cercor/bhaa023

Ayyash, D., Malik-Moraleda, S., Gallée, J., Affourtit, J., Hoffman, M., Mineroff, Z., Jouravlev, O., & Fedorenko, E. (2021). *The universal language network: A cross-linguistic investigation spanning 45 languages and 11 language families* (p. 2021.07.28.454040). https://doi.org/10.1101/2021.07.28.454040

Bambini, V., Gentili, C., Ricciardi, E., Bertinetto, P. M., & Pietrini, P. (2011). Decomposing metaphor processing at the cognitive and neural level through functional magnetic resonance imaging. *Brain Research Bulletin*, *86*(3), 203–216. https://doi.org/10.1016/j.brainresbull.2011.07.015

Bašnáková, J., van Berkum, J., Weber, K., & Hagoort, P. (2015). A job interview in the MRI scanner: How does indirectness affect addressees and overhearers? *Neuropsychologia*, *76*, 79–91. https://doi.org/10.1016/j.neuropsychologia.2015.03.030

Bašnáková, J., Weber, K., Petersson, K. M., van Berkum, J., & Hagoort, P. (2014). Beyond the Language Given: The Neural Correlates of Inferring Speaker Meaning. *Cerebral Cortex*, *24*(10), 2572–2578. https://doi.org/10.1093/cercor/bht112

Bautista, A., & Wilson, S. M. (2016). Neural responses to grammatically and lexically degraded speech. *Language, Cognition and Neuroscience*, *31*(4), 567–574. https://doi.org/10.1080/23273798.2015.1123281

Beaty, R. E., & Silvia, P. J. (2013). Metaphorically speaking: Cognitive abilities and the production of figurative language. *Memory & Cognition*, *41*(2), 255–267. https://doi.org/10.3758/s13421-012-0258-5

Bekinschtein, T. A., Davis, M. H., Rodd, J. M., & Owen, A. M. (2011). Why Clowns Taste Funny: The Relationship between Humor and Semantic Ambiguity. *Journal of Neuroscience*, *31*(26), 9665–9671. https://doi.org/10.1523/JNEUROSCI.5058-10.2011

Bendersky, M., Lomlomdjian, C., Abusamra, V., Elizalde Acevedo, B., Kochen, S., & Alba-Ferrara, L. (2021). Functional anatomy of idiomatic expressions. *Brain Topography*, *34*(4), 489–503. https://doi.org/10.1007/s10548-021-00843-3

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

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. https://doi.org/10.1152/jn.00884.2013

Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012). Looking at the brains behind figurative language—A quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia*, *50*(11), 2669–2683. https://doi.org/10.1016/j.neuropsychologia.2012.07.021

Bosco, F. M., Parola, A., Valentini, M. C., & Morese, R. (2017). Neural correlates underlying the comprehension of deceitful and ironic communicative intentions. *Cortex*, *94*, 73–86. https://doi.org/10.1016/j.cortex.2017.06.010

Braga, R. M., & Buckner, R. L. (2017). Parallel Interdigitated Distributed Networks within the Individual Estimated by Intrinsic Functional Connectivity. *Neuron*, *95*(2), 457-471.e5. https://doi.org/10.1016/j.neuron.2017.06.038

Braga, R. M., DiNicola, L. M., Becker, H. C., & Buckner, R. L. (2020). Situating the left-lateralized language network in the broader organization of multiple specialized large-scale distributed networks. *Journal of Neurophysiology*, *124*(5), 1415–1448. https://doi.org/10.1152/jn.00753.2019

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience*, *3*(3), 243–249. https://doi.org/10.1038/nrn756

Brownell, H. H. (1988). The neuropsychology of narrative comprehension. *Aphasiology*, *2*(3–4), 247–250. https://doi.org/10.1080/02687038808248918

Brownell, H. H., Michel, D., Powelson, J., & Gardner, H. (1983). Surprise but not coherence: Sensitivity to verbal humor in right-hemisphere patients. *Brain and Language*, *18*(1), 20–27. https://doi.org/10.1016/0093-934X(83)90002-0

Brownell, H. H., Potter, H. H., Bihrle, A. M., & Gardner, H. (1986). Inference deficits in right brain-damaged patients. *Brain and Language*, *27*(2), 310–321. https://doi.org/10.1016/0093-934X(86)90022-2

Bryan, K. L. (1988). Assessment of language disorders after right hemisphere damage. *International Journal of Language & Communication Disorders*, *23*(2), 111–125. https://doi.org/10.3109/13682828809019881

Burgess, C., & Chiarello, C. (1996). Neurocognitive mechanisms underlying metaphor comprehension and other figurative language. *Metaphor and Symbolic Activity*, *11*(1), 67–84. https://doi.org/10.1207/s15327868ms1101_4

Caillies, S., Bertot, V., Motte, J., Raynaud, C., & Abely, M. (2014). Social cognition in ADHD: Irony understanding and recursive theory of mind. *Research in Developmental Disabilities*, *35*(11), 3191–3198. https://doi.org/10.1016/j.ridd.2014.08.002

Campbell, D. W., Wallace, M. G., Modirrousta, M., Polimeni, J. O., McKeen, N. A., & Reiss, J. P. (2015). The neural basis of humour comprehension and humour appreciation: The roles of the temporoparietal junction and superior frontal gyrus. *Neuropsychologia*, *79*, 10–20. https://doi.org/10.1016/j.neuropsychologia.2015.10.013

Cardillo, E. R., McQuire, M., & Chatterjee, A. (2018). Selective Metaphor Impairments After Left, Not Right, Hemisphere Injury. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02308

Cardillo, R., Mammarella, I. C., Demurie, E., Giofrè, D., & Roeyers, H. (2021). Pragmatic Language in Children and Adolescents With Autism Spectrum Disorder: Do Theory of Mind and Executive Functions Have a Mediating Role? *Autism Research*, *14*(5), 932–945. https://doi.org/10.1002/aur.2423

Champagne-Lavau, M., & Stip, E. (2010). Pragmatic and executive dysfunction in schizophrenia. *Journal of Neurolinguistics*, *23*(3), 285–296. https://doi.org/10.1016/j.jneuroling.2009.08.009

Chan, Y.-C., & Lavallee, J. P. (2015a). Temporo-parietal and fronto-parietal lobe contributions to theory of mind and executive control: An fMRI study of verbal jokes. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01285

Chan, Y.-C., & Lavallee, J. P. (2015b). Temporo-parietal and fronto-parietal lobe contributions to theory of mind and executive control: An fMRI study of verbal jokes. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/article/10.3389/fpsyg.2015.01285

Channon, S., & Watts, M. (2003). Pragmatic language interpretation after closed head injury: Relationship to executive functioning. *Cognitive Neuropsychiatry*, *8*(4), 243–260. https://doi.org/10.1080/135468000344000002

Chen, E., Widick, P., & Chatterjee, A. (2008). Functional–anatomical organization of predicate metaphor processing. *Brain and Language*, *107*(3), 194–202. https://doi.org/10.1016/j.bandl.2008.06.007

Chen, X., Affourtit, J., Ryskin, R., Regev, T. I., Norman-Haignere, S., Jouravlev, O., Malik-Moraleda, S., Kean, H., Varley, R., & Fedorenko, E. (2021). *The human language system does not support music processing* (p. 2021.06.01.446439). https://doi.org/10.1101/2021.06.01.446439

Chen, X., Lu, B., & Yan, C.-G. (2018). Reproducibility of R-fMRI metrics on the impact of different strategies for multiple comparison correction and sample sizes. *Human Brain Mapping*, *39*(1), 300–318. https://doi.org/10.1002/hbm.23843

Chow, H. M., Kaup, B., Raabe, M., & Greenlee, M. W. (2008). Evidence of fronto-temporal interactions for strategic inference processes during language comprehension. *NeuroImage*, *40*(2), 940–954. https://doi.org/10.1016/j.neuroimage.2007.11.044

Citron, F. M. M., & Goldberg, A. E. (2014). Metaphorical Sentences Are More Emotionally Engaging than Their Literal Counterparts. *Journal of Cognitive Neuroscience*, *26*(11), 2585–2595. https://doi.org/10.1162/jocn_a_00654

Citron, F. M. M., Güsten, J., Michaelis, N., & Goldberg, A. E. (2016). Conventional metaphors in longer passages evoke affective brain response. *NeuroImage*, *139*, 218–230. https://doi.org/10.1016/j.neuroimage.2016.06.020

Coulson, S., & Williams, R. F. (2005). Hemispheric asymmetries and joke comprehension. *Neuropsychologia*, *43*(1), 128–141. https://doi.org/10.1016/j.neuropsychologia.2004.03.015

Dai, R. H., Chen, H.-C., Chan, Y. C., Wu, C.-L., Li, P., Cho, S. L., & Hu, J.-F. (2017). To Resolve or Not To Resolve, that Is the Question: The Dual-Path Model of Incongruity Resolution and

Absurd Verbal Humor by fMRI. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00498

Deen, B., & Freiwald, W. A. (2021). *Parallel systems for social and spatial reasoning within the cortical apex* (p. 2021.09.23.461550). https://doi.org/10.1101/2021.09.23.461550

Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional Organization of Social Perception and Cognition in the Superior Temporal Sulcus. *Cerebral Cortex*, *25*(11), 4596–4609. https://doi.org/10.1093/cercor/bhv111

Delis, D. C., Wapner, W., Gardner, H., & Moses, J. A. (1983). The Contribution of the Right Hemisphere to the Organization of Paragraphs. *Cortex*, *19*(1), 43–50. https://doi.org/10.1016/S0010-9452(83)80049-5

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv:1810.04805 [Cs]*. http://arxiv.org/abs/1810.04805

Diachek, E., Blank, I., Siegelman, M., Affourtit, J., & Fedorenko, E. (2020). The Domain-General Multiple Demand (MD) Network Does Not Support Core Aspects of Language Comprehension: A Large-Scale fMRI Investigation. *The Journal of Neuroscience*, *40*(23), 4536–4550. https://doi.org/10.1523/JNEUROSCI.2036-19.2020

Diaz, M. T., & Hogstrom, L. J. (2011a). The Influence of Context on Hemispheric Recruitment during Metaphor Processing. *Journal of Cognitive Neuroscience*, *23*(11), 3586–3597. https://doi.org/10.1162/jocn_a_00053

Diaz, M. T., & Hogstrom, L. J. (2011b). The Influence of Context on Hemispheric Recruitment during Metaphor Processing. *Journal of Cognitive Neuroscience*, *23*(11), 3586–3597. https://doi.org/10.1162/jocn_a_00053

DiNicola, L. M., Braga, R. M., & Buckner, R. L. (2020). Parallel distributed networks dissociate episodic and social functions within the individual. *Journal of Neurophysiology*, *123*(3), 1144–1179. https://doi.org/10.1152/jn.00529.2019

DiNicola, L. M., & Buckner, R. L. (2021). Precision estimates of parallel distributed association networks: Evidence for domain specialization and implications for evolution and development. *Current Opinion in Behavioral Sciences*, *40*, 120–129. https://doi.org/10.1016/j.cobeha.2021.03.029

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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1016/S0166-2236(00)01633-7

Dworetsky, A., Seitzman, B. A., Adeyemo, B., Neta, M., Coalson, R. S., Petersen, S. E., & Gratton, C. (2021). Probabilistic mapping of human functional brain networks identifies regions of high group consensus. *NeuroImage*, *237*, 118164. https://doi.org/10.1016/j.neuroimage.2021.118164

Egorova, N., Shtyrov, Y., & Pulvermüller, F. (2016). Brain basis of communicative actions in language. *NeuroImage*, *125*, 857–867. https://doi.org/10.1016/j.neuroimage.2015.10.055

Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *NeuroImage*, *59*(3), 2349–2361. https://doi.org/10.1016/j.neuroimage.2011.09.017

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, *30*(9), 2907–2926. https://doi.org/10.1002/hbm.20718

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, *113*(28), 7900–7905. https://doi.org/10.1073/pnas.1602413113

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., & Peters, T. M. (1993). 3D statistical neuroanatomical models from 305 MRI volumes. *1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, 1813–1817 vol.3. https://doi.org/10.1109/NSSMIC.1993.373602

Eviatar, Z., & Just, M. A. (2006). Brain correlates of discourse processing: An fMRI investigation of irony and conventional metaphor comprehension. *Neuropsychologia*, *44*(12), 2348–2359. https://doi.org/10.1016/j.neuropsychologia.2006.05.007

Fairchild, S., & Papafragou, A. (2021). The Role of Executive Function and Theory of Mind in Pragmatic Computations. *Cognitive Science*, *45*(2), e12938. https://doi.org/10.1111/cogs.12938

Farkas, A. H., Trotti, R. L., Edge, E. A., Huang, L.-Y., Kasowski, A., Thomas, O. F., Chlan, E., Granros, M. P., Patel, K. K., & Sabatinelli, D. (2021). Humor and emotion: Quantitative meta analyses of functional neuroimaging studies. *Cortex*, *139*, 60–72. https://doi.org/10.1016/j.cortex.2021.02.023

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/article/10.3389/fpsyg.2014.00335

Fedorenko, E. (2021). The early origins and the growing popularity of the individual-subject analytic approach in human neuroscience. *Current Opinion in Behavioral Sciences*, *40*, 105–112. https://doi.org/10.1016/j.cobeha.2021.02.023

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

Fedorenko, E., & Blank, I. A. (2020). Broca's Area Is Not a Natural Kind. *Trends in Cognitive Sciences*, *24*(4), 270–284. https://doi.org/10.1016/j.tics.2020.01.001

Fedorenko, E., Blank, I. A., Siegelman, M., & Mineroff, Z. (2020). Lack of selectivity for syntax relative to word meanings throughout the language network. *Cognition*, *203*, 104348. https://doi.org/10.1016/j.cognition.2020.104348

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-Selective and Domain-General Regions Lie Side by Side within Broca's Area. *Current Biology*, *22*(21), 2059–2062. https://doi.org/10.1016/j.cub.2012.09.011

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

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

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

Fedorenko, E., & Shain, C. (2021). Similarity of Computations Across Domains Does Not Imply Shared Implementation: The Case of Language Comprehension. *Current Directions in Psychological Science*, *30*(6), 526–534. https://doi.org/10.1177/09637214211046955

Fedorenko, E., & Thompson-Schill, S. L. (2014). Reworking the language network. *Trends in Cognitive Sciences*, *18*(3), 120–126. https://doi.org/10.1016/j.tics.2013.12.00

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

Feng, W., Wu, Y., Jan, C., Yu, H., Jiang, X., & Zhou, X. (2017). Effects of contextual relevance on pragmatic inference during conversation: An fMRI study. *Brain and Language*, *171*, 52–61. https://doi.org/10.1016/j.bandl.2017.04.005

Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, *29*(5), 581–593. https://doi.org/10.1002/hbm.20422

Ferstl, E. C., & von Cramon, D. Y. (2001). The role of coherence and cohesion in text comprehension: An event-related fMRI study. *Cognitive Brain Research*, *11*(3), 325–340. https://doi.org/10.1016/S0926-6410(01)00007-6

Ferstl, E. C., Walther, K., Guthke, T., & von Cramon, D. Y. (2005). Assessment of Story Comprehension Deficits After Brain Damage. *Journal of Clinical and Experimental Neuropsychology*, *27*(3), 367–384. https://doi.org/10.1080/13803390490515784

Filik, R., Țurcan, A., Ralph-Nearman, C., & Pitiot, A. (2019). What is the difference between irony and sarcasm? An fMRI study. *Cortex*, *115*, 112–122. https://doi.org/10.1016/j.cortex.2019.01.025

Friese, U., Rutschmann, R., Raabe, M., & Schmalhofer, F. (2008). Neural Indicators of Inference Processes in Text Comprehension: An Event-related Functional Magnetic Resonance Imaging Study. *Journal of Cognitive Neuroscience*, *20*(11), 2110–2124. https://doi.org/10.1162/jocn.2008.20141

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

Gagnon, L., Goulet, P., Giroux, F., & Joanette, Y. (2003). Processing of metaphoric and non-metaphoric alternative meanings of words after right- and left-hemispheric lesion. *Brain and Language*, *87*(2), 217–226. https://doi.org/10.1016/S0093-934X(03)00057-9

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia*, *38*(1), 11–21. https://doi.org/10.1016/S0028-3932(99)00053-6

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage*, *15*(4), 870–878. https://doi.org/10.1006/nimg.2001.1037

Gernsbacher, M. A., & Robertson, R. R. W. (1999). The role of suppression in figurative language comprehension. *Journal of Pragmatics*, *31*(12), 1619–1630. https://doi.org/10.1016/S0378-2166(99)00007-7

Geschwind, N. (1970). The Organization of Language and the Brain. *Science*. https://doi.org/10.1126/science.170.3961.940

Giora, R., Zaidel, E., Soroker, N., Batori, G., & Kasher, A. (2000). Differential effect of right- and left-hemisphere damage on understanding sarcasm and metaphor. *Metaphor and Symbol*, *15*(1–2), 63–83. https://doi.org/10.1207/S15327868MS151&2_5

Gordon, E. M., & Nelson, S. M. (2021). Three types of individual variation in brain networks revealed by single-subject functional connectivity analyses. *Current Opinion in Behavioral Sciences*, *40*, 79–86. https://doi.org/10.1016/j.cobeha.2021.02.014

Gratton, C., & Braga, R. M. (2021). Editorial overview: Deep imaging of the individual brain: past, practice, and promise. *Current Opinion in Behavioral Sciences*, *40*, iii–vi. https://doi.org/10.1016/j.cobeha.2021.06.011

Grice, H. P. (1975). Logic and conversation. *Syntax and Semantics*, 3, 41–58.

Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, *48*(2), 101–119. https://doi.org/10.1016/0010-0277(93)90026-R

Happé, F., Brownell, H., & Winner, E. (1999). Acquired `theory of mind' impairments following stroke. *Cognition*, *70*(3), 211–240. https://doi.org/10.1016/S0010-0277(99)00005-0

Heidlmayr, K., Weber, K., Takashima, A., & Hagoort, P. (2020). No title, no theme: The joined neural space between speakers and listeners during production and comprehension of multi-sentence discourse. *Cortex*, *130*, 111–126. https://doi.org/10.1016/j.cortex.2020.04.035

Hellbernd, N., & Sammler, D. (2018). Neural bases of social communicative intentions in speech. *Social Cognitive and Affective Neuroscience*, *13*(6), 604–615. https://doi.org/10.1093/scan/nsy034

Herold, R., Varga, E., Hajnal, A., Hamvas, E., Berecz, H., Tóth, B., & Tényi, T. (2018). Altered Neural Activity during Irony Comprehension in Unaffected First-Degree Relatives of

Schizophrenia Patients—An fMRI Study. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.02309

Hillert, D. G., & Buračas, G. T. (2009). The neural substrates of spoken idiom comprehension. *Language and Cognitive Processes*, *24*(9), 1370–1391. https://doi.org/10.1080/01690960903057006

Hu, J., Levy, R., & Zaslavsky, N. (2021). Scalable pragmatic communication via self-supervision. *Proceedings of the 2021 ICML Workshop on Self-Supervised Learning for Reasoning and Perception. arXiv:2108.05799.*

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*. https://doi.org/10.3389/fnhum.2015.00430

Ianni, G. R., Cardillo, E. R., McQuire, M., & Chatterjee, A. (2014). Flying under the radar: Figurative language impairments in focal lesion patients. *Frontiers in Human Neuroscience*, *8*. https://www.frontiersin.org/article/10.3389/fnhum.2014.00871

Ivanova, A. A., Srikant, S., Sueoka, Y., Kean, H. H., Dhamala, R., O'Reilly, U.-M., Bers, M. U., & Fedorenko, E. (2020). Comprehension of computer code relies primarily on domain-general executive brain regions. *ELife*, *9*, e58906. https://doi.org/10.7554/eLife.58906

Jackendoff, R. (2002). Semantic and conceptual foundations. In *Foundations of language: Brain, meaning, grammar, evolution* (pp. 267-293). Oxford University Press.

Jacoby, N., Bruneau, E., Koster-Hale, J., & Saxe, R. (2016). Localizing Pain Matrix and Theory of Mind networks with both verbal and non-verbal stimuli. *NeuroImage*, *126*, 39–48. https://doi.org/10.1016/j.neuroimage.2015.11.025

Jang, G., Yoon, S., Lee, S.-E., Park, H., Kim, J., Ko, J. H., & Park, H.-J. (2013). Everyday conversation requires cognitive inference: Neural bases of comprehending implicated meanings in conversations. *NeuroImage*, *81*, 61–72. https://doi.org/10.1016/j.neuroimage.2013.05.027

Joanette, Y., Goulet, P., Hannequin, D., & Boeglin, J. (1990). Right hemisphere and verbal communication. Springer-Verlag Publishing. https://doi.org/10.1007/978-1-4612-4460-8

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

Klepousniotou, E., & Baum, S. R. (2005). Unilateral brain damage effects on processing homonymous and polysemous words. *Brain and Language*, *93*(3), 308–326. https://doi.org/10.1016/j.bandl.2004.10.011

Klooster, N., McQuire, M., Grossman, M., McMillan, C., Chatterjee, A., & Cardillo, E. (2020). The Neural Basis of Metaphor Comprehension: Evidence from Left Hemisphere Degeneration. *Neurobiology of Language*, *1*(4), 474–491. https://doi.org/10.1162/nol_a_00022

Kristensen, L. B., Wang, L., Petersson, K. M., & Hagoort, P. (2013). The Interface Between Language and Attention: Prosodic Focus Marking Recruits a General Attention Network in Spoken Language Comprehension. *Cerebral Cortex*, *23*(8), 1836–1848. https://doi.org/10.1093/cercor/bhs164

Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., & Holcomb, P. J. (2006). Making sense of discourse: An fMRI study of causal inferencing across sentences. *NeuroImage*, *33*(1), 343–361. https://doi.org/10.1016/j.neuroimage.2006.06.001

Kuperberg, G. R., McGuire, P. K., Bullmore, E. T., Brammer, M. J., Rabe-Hesketh, S., Wright, I. C., Lythgoe, D. J., Williams, S. C. R., & David, A. S. (2000). Common and Distinct Neural

Substrates for Pragmatic, Semantic, and Syntactic Processing of Spoken Sentences: An fMRI Study. *Journal of Cognitive Neuroscience*, *12*(2), 321–341. https://doi.org/10.1162/089892900562138

Lauro, L. J. R., Tettamanti, M., Cappa, S. F., & Papagno, C. (2008). Idiom Comprehension: A Prefrontal Task? *Cerebral Cortex*, *18*(1), 162–170. https://doi.org/10.1093/cercor/bhm042

Lee, S. S., & Dapretto, M. (2006). Metaphorical vs. literal word meanings: FMRI evidence against a selective role of the right hemisphere. *NeuroImage*, *29*(2), 536–544. https://doi.org/10.1016/j.neuroimage.2005.08.003

Licea-Haquet, G. L., Velásquez-Upegui, E. P., Holtgraves, T., & Giordano, M. (2019). Speech act recognition in Spanish speakers. *Journal of Pragmatics*, *141*, 44–56. https://doi.org/10.1016/j.pragma.2018.12.013

Lin, N., Yang, X., Li, J., Wang, S., Hua, H., Ma, Y., & Li, X. (2018). Neural correlates of three cognitive processes involved in theory of mind and discourse comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(2), 273–283. https://doi.org/10.3758/s13415-018-0568-6

Lipkin, B., Tuckute, G., Affourtit, J., Small, H., Mineroff, Z., Kean, H., Jouravlev, O., Rakocevic, L., Pritchett, B., Siegelman, M., Hoeflin, C., Pongos, A., Blank, I. A., Struhl, M. K., Ivanova, A., Shannon, S., Sathe, A., Hoffmann, M., Nieto-Castañón, A., & Fedorenko, E. (2022). *LanA (Language Atlas): A probabilistic atlas for the language network based on fMRI data from >800 individuals*. bioRxiv. https://doi.org/10.1101/2022.03.06.483177

Lipkin, B., Tuckute, G., Affourtit, J., Small, H., Mineroff, Z., Nieto-Castañón, A., and Fedorenko, E. (in preparation). A probabilistic atlas for the Multiple Demand (MD) network based on data from 691 individuals performing a spatial working memory localizer task.

Lisofsky, N., Kazzer, P., Heekeren, H. R., & Prehn, K. (2014). Investigating socio-cognitive processes in deception: A quantitative meta-analysis of neuroimaging studies. *Neuropsychologia*, *61*, 113–122. https://doi.org/10.1016/j.neuropsychologia.2014.06.001

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. https://doi.org/10.1016/j.neuroimage.2016.05.073

Martín-Loeches, M., Casado, P., Hernández-Tamames, J. A., & Álvarez-Linera, J. (2008). Brain activation in discourse comprehension: A 3t fMRI study. *NeuroImage*, *41*(2), 614–622. https://doi.org/10.1016/j.neuroimage.2008.02.047

Mashal, N., Faust, M., & Hendler, T. (2005a). The role of the right hemisphere in processing nonsalient metaphorical meanings: Application of Principal Components Analysis to fMRI data. *Neuropsychologia*, *43*(14), 2084–2100. https://doi.org/10.1016/j.neuropsychologia.2005.03.019

Mashal, N., Faust, M., & Hendler, T. (2005b). The role of the right hemisphere in processing nonsalient metaphorical meanings: Application of Principal Components Analysis to fMRI data. *Neuropsychologia*, *43*(14), 2084–2100. https://doi.org/10.1016/j.neuropsychologia.2005.03.019

Mashal, N., Faust, M., Hendler, T., & Jung-Beeman, M. (2007). An fMRI investigation of the neural correlates underlying the processing of novel metaphoric expressions. *Brain and Language*, *100*(2), 115–126. https://doi.org/10.1016/j.bandl.2005.10.005

Mason, R. A., & Just, M. A. (2011). Differentiable cortical networks for inferences concerning people's intentions versus physical causality. *Human Brain Mapping*, *32*(2), 313–329. https://doi.org/10.1002/hbm.21021

Mather, M., Cacioppo, J. T., & Kanwisher, N. (2013). How fMRI Can Inform Cognitive Theories. *Perspectives on Psychological Science*, *8*(1), 108–113. https://doi.org/10.1177/1745691612469037

Matsui, T., Nakamura, T., Utsumi, A., Sasaki, A. T., Koike, T., Yoshida, Y., Harada, T., Tanabe, H. C., & Sadato, N. (2016). The role of prosody and context in sarcasm comprehension: Behavioral and fMRI evidence. *Neuropsychologia*, *87*, 74–84. https://doi.org/10.1016/j.neuropsychologia.2016.04.031

Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual Differences in Children's Pragmatic Ability: A Review of Associations with Formal Language, Social Cognition, and Executive Functions. *Language Learning and Development*, *14*(3), 186–223. https://doi.org/10.1080/15475441.2018.1455584

McDonald, S., & Pearce, S. (1998). Requests That Overcome Listener Reluctance: Impairment Associated with Executive Dysfunction in Brain Injury. *Brain and Language*, *61*(1), 88–104. https://doi.org/10.1006/brln.1997.1846

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

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. https://doi.org/10.1177/0956797612437427

Myers, P.S., Linebaugh, C.W., 1981. Comprehension of idiomatic expressions by right-hemisphere-damaged adults. In R. H. Brookshire (Ed.), Clinical aphasiology: Conference proceedings pp. 254–261. Minneapolis: BRK.

Myers, P. (1998). Right hemisphere damage: Disorders of communication and cognition. San Diego, CA: Singular.

Nagels, A., Kauschke, C., Schrauf, J., Whitney, C., Straube, B., & Kircher, Ti. (2013). Neural substrates of figurative language during natural speech perception: An fMRI study. *Frontiers in Behavioral Neuroscience*, *7*. https://doi.org/10.3389/fnbeh.2013.00121

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

Obert, A., Gierski, F., Calmus, A., Flucher, A., Portefaix, C., Pierot, L., Kaladjian, A., & Caillies, S. (2016). Neural Correlates of Contrast and Humor: Processing Common Features of Verbal Irony. *PLOS ONE*, *11*(11), e0166704. https://doi.org/10.1371/journal.pone.0166704

Obert, A., Gierski, F., Calmus, A., Portefaix, C., Declercq, C., Pierot, L., & Caillies, S. (2014). Differential bilateral involvement of the parietal gyrus during predicative metaphor processing: An auditory fMRI study. *Brain and Language*, *137*, 112–119. https://doi.org/10.1016/j.bandl.2014.08.002

Oliveri, M., Romero, L., & Papagno, C. (2004). Left But Not Right Temporal Involvement in Opaque Idiom Comprehension: A Repetitive Transcranial Magnetic Stimulation Study. *Journal of Cognitive Neuroscience*, *16*(5), 848–855. https://doi.org/10.1162/089892904970717

Ouden, D.-B. den, Dickey, M. W., Anderson, C., & Christianson, K. (2016). Neural correlates of early-closure garden-path processing: Effects of prosody and plausibility. *The Quarterly Journal of Experimental Psychology*, *69*(5), 926–949. https://doi.org/10.1080/17470218.2015.1028416

Papagno, C., & Genoni, A. (2004). The role of syntactic competence in idiom comprehension: A study on aphasic patients. *Journal of Neurolinguistics*, *17*(5), 371–382. https://doi.org/10.1016/j.jneuroling.2003.11.002

Paunov, A. M. (2018). *FMRI studies of the relationship between language and theory of mind in adult cognition* [Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/121828

Paunov, A. M., Blank, I. A., & Fedorenko, E. (2019). Functionally distinct language and Theory of Mind networks are synchronized at rest and during language comprehension. *Journal of Neurophysiology*, *121*(4), 1244–1265. https://doi.org/10.1152/jn.00619.2018

Paunov, A. M., Blank, I. A., Jouravlev, O., Mineroff, Z., Gallée, J., & Fedorenko, E. (2021). *Differential tracking of linguistic vs. Mental state content in naturalistic stimuli by language and Theory of Mind (ToM) brain networks* (p. 2021.04.28.441724). https://doi.org/10.1101/2021.04.28.441724

Perrone□Bertolotti, M., Dohen, M., Lœvenbruck, H., Sato, M., Pichat, C., & Baciu, M. (2013). Neural correlates of the perception of contrastive prosodic focus in French: A functional magnetic resonance imaging study. *Human Brain Mapping*, *34*(10), 2574–2591. https://doi.org/10.1002/hbm.22090

Piñango, M. M., Zhang, M., Foster□Hanson, E., Negishi, M., Lacadie, C., & Constable, R. T. (2017). Metonymy as Referential Dependency: Psycholinguistic and Neurolinguistic Arguments for a Unified Linguistic Treatment. *Cognitive Science*, *41*(S2), 351–378. https://doi.org/10.1111/cogs.12341

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

Poldrack, R. A. (2011). Inferring Mental States from Neuroimaging Data: From Reverse Inference to Large-Scale Decoding. *Neuron*, *72*(5), 692–697. https://doi.org/10.1016/j.neuron.2011.11.001

Pomp, J., Bestgen, A.-K., Schulze, P., Müller, C. J., Citron, F. M. M., Suchan, B., & Kuchinke, L. (2018). Lexical olfaction recruits olfactory orbitofrontal cortex in metaphorical and literal contexts. *Brain and Language*, *179*, 11–21. https://doi.org/10.1016/j.bandl.2018.02.001

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Rapp, A. M., Erb, M., Grodd, W., Bartels, M., & Markert, K. (2011). Neural correlates of metonymy resolution. *Brain and Language*, *119*(3), 196–205. https://doi.org/10.1016/j.bandl.2011.07.004

Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., & Kircher, T. T. J. (2004). Neural correlates of metaphor processing. *Cognitive Brain Research*, *20*(3), 395–402. https://doi.org/10.1016/j.cogbrainres.2004.03.017

Rapp, A. M., Mutschler, D. E., & Erb, M. (2012). Where in the brain is nonliteral language? A coordinate-based meta-analysis of functional magnetic resonance imaging studies. *NeuroImage*, *63*(1), 600–610. https://doi.org/10.1016/j.neuroimage.2012.06.022

Rapp, A. M., Mutschler, D. E., Wild, B., Erb, M., Lengsfeld, I., Saur, R., & Grodd, W. (2010). Neural correlates of irony comprehension: The role of schizotypal personality traits. *Brain and Language*, *113*(1), 1–12. https://doi.org/10.1016/j.bandl.2009.11.007

Regev, T. I., Affourtit, J., Chen, X., Schipper, A. E., Bergen, L., Mahowald, K., & Fedorenko, E. (2021). *High-level language brain regions are sensitive to sub-lexical regularities* (p. 2021.06.11.447786). https://doi.org/10.1101/2021.06.11.447786

Reyes-Aguilar, A., Valles-Capetillo, E., & Giordano, M. (2018). A Quantitative Meta-analysis of Neuroimaging Studies of Pragmatic Language Comprehension: In Search of a Universal Neural Substrate. *Neuroscience*, *395*, 60–88. https://doi.org/10.1016/j.neuroscience.2018.10.043

Rints, A., McAuley, T., & Nilsen, E. S. (2015). Social Communication Is Predicted by Inhibitory Ability and ADHD Traits in Preschool-Aged Children: A Mediation Model. *Journal of Attention Disorders*, *19*(10), 901–911. https://doi.org/10.1177/1087054714558873

Samur, D., Lai, V. T., Hagoort, P., & Willems, R. M. (2015a). Emotional context modulates embodied metaphor comprehension. *Neuropsychologia*, *78*, 108–114. https://doi.org/10.1016/j.neuropsychologia.2015.10.003

Samur, D., Lai, V. T., Hagoort, P., & Willems, R. M. (2015b). Emotional context modulates embodied metaphor comprehension. *Neuropsychologia*, *78*, 108–114. https://doi.org/10.1016/j.neuropsychologia.2015.10.003

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." *NeuroImage*, *19*(4), 1835–1842. https://doi.org/10.1016/S1053-8119(03)00230-1

Saxe, R., Moran, J. M., Scholz, J., & Gabrieli, J. (2006). Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Social Cognitive and Affective Neuroscience*, *1*(3), 229–234. https://doi.org/10.1093/scan/nsl034

Saxe, R., & Powell, L. J. (2006). It's the Thought That Counts: Specific Brain Regions for One Component of Theory of Mind. *Psychological Science*, *17*(8), 692–699. https://doi.org/10.1111/j.1467-9280.2006.01768.x

Schmidt, G. L., & Seger, C. A. (2009). Neural correlates of metaphor processing: The roles of figurativeness, familiarity and difficulty. *Brain and Cognition*, *71*(3), 375–386. https://doi.org/10.1016/j.bandc.2009.06.001

Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E. N., & Saxe, R. (2009). Distinct Regions of Right Temporo-Parietal Junction Are Selective for Theory of Mind and Exogenous Attention. *PLOS ONE*, *4*(3), e4869. https://doi.org/10.1371/journal.pone.0004869

Schuil, K. D. I., Smits, M., & Zwaan, R. A. (2013). Sentential Context Modulates the Involvement of the Motor Cortex in Action Language Processing: An fMRI Study. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00100

Scott, T. L., Gallée, J., & Fedorenko, E. (2017). A new fun and robust version of an fMRI localizer for the frontotemporal language system. *Cognitive Neuroscience*, *8*(3), 167–176. https://doi.org/10.1080/17588928.2016.1201466

Searle, J. (1979). Metaphor. In A. Ortony (Ed.), *Metaphor and thought*. New York: Cambridge University Press.

Shain, C., Blank, I. A., van Schijndel, M., Schuler, W., & Fedorenko, E. (2020). FMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia*, *138*, 107307. https://doi.org/10.1016/j.neuropsychologia.2019.107307

Shain, C., Paunov, A., Chen, X., Lipkin, B., & Fedorenko, E. (2022). No evidence of theory of mind reasoning in the human language network. *bioRxiv*. https://doi.org/10.1101/2022.07.18.500516

Shashidhara, S., Mitchell, D. J., Erez, Y., & Duncan, J. (2019). Progressive Recruitment of the Frontoparietal Multiple-demand System with Increased Task Complexity, Time Pressure, and Reward. *Journal of Cognitive Neuroscience*, *31*(11), 1617–1630. https://doi.org/10.1162/jocn_a_01440

Shibata, M., Abe, J., Itoh, H., Shimada, K., & Umeda, S. (2011). Neural processing associated with comprehension of an indirect reply during a scenario reading task. *Neuropsychologia*, *49*(13), 3542–3550. https://doi.org/10.1016/j.neuropsychologia.2011.09.006

Shibata, M., Abe, J., Terao, A., & Miyamoto, T. (2007). Neural mechanisms involved in the comprehension of metaphoric and literal sentences: An fMRI study. *Brain Research*, *1166*, 92–102. https://doi.org/10.1016/j.brainres.2007.06.040

Shibata, M., Terasawa, Y., & Umeda, S. (2014). Integration of cognitive and affective networks in humor comprehension. *Neuropsychologia*, *65*, 137–145. https://doi.org/10.1016/j.neuropsychologia.2014.10.025

Shibata, M., Toyomura, A., Itoh, H., & Abe, J. (2010). Neural substrates of irony comprehension: A functional MRI study. *Brain Research*, *1308*, 114–123. https://doi.org/10.1016/j.brainres.2009.10.030

Shibata, M., Toyomura, A., Motoyama, H., Itoh, H., Kawabata, Y., & Abe, J. (2012). Does simile comprehension differ from metaphor comprehension? A functional MRI study. *Brain and Language*, *121*(3), 254–260. https://doi.org/10.1016/j.bandl.2012.03.006

Siegal, M., Carrington, J., & Radel, M. (1996). Theory of Mind and Pragmatic Understanding Following Right Hemisphere Damage. *Brain and Language*, *53*(1), 40–50. https://doi.org/10.1006/brln.1996.0035

Smith, D. M., Perez, D. C., Porter, A., Dworetsky, A., & Gratton, C. (2021). Light through the fog: Using precision fMRI data to disentangle the neural substrates of cognitive control. *Current Opinion in Behavioral Sciences*, *40*, 19–26. https://doi.org/10.1016/j.cobeha.2020.12.004

Smith, N., & Levy, R. (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, No. 33).

Smith, V., Duncan, J., & Mitchell, D. J. (2021). Roles of the Default Mode and Multiple-Demand Networks in Naturalistic versus Symbolic Decisions. *The Journal of Neuroscience*, *41*(10), 2214–2228. https://doi.org/10.1523/JNEUROSCI.1888-20.2020

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142). Cambridge, MA: Harvard University Press.

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J.-B., & Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*, *63*(1), 25–39. https://doi.org/10.1016/j.neuroimage.2012.06.046

Stemmer, B., Giroux, F., & Joanette, Y. (1994). Production and Evaluation of Requests by Right Hemisphere Brain-Damaged Individuals. *Brain and Language*, *47*(1), 1–31. https://doi.org/10.1006/brln.1994.1040

Stringaris, A. K., Medford, N. C., Giampietro, V., Brammer, M. J., & David, A. S. (2007). Deriving meaning: Distinct neural mechanisms for metaphoric, literal, and non-meaningful sentences. *Brain and Language*, *100*(2), 150–162. https://doi.org/10.1016/j.bandl.2005.08.001

Tahmasebi, A. M., Davis, M. H., Wild, C. J., Rodd, J. M., Hakyemez, H., Abolmaesumi, P., & Johnsrude, I. S. (2012). Is the Link between Anatomical Structure and Function Equally Strong at All Cognitive Levels of Processing? *Cerebral Cortex*, *22*(7), 1593–1603. https://doi.org/10.1093/cercor/bhr205

Thirion, B., Thual, A., & Pinho, A. L. (2021). From deep brain phenotyping to functional atlasing. *Current Opinion in Behavioral Sciences*, *40*, 201–212. https://doi.org/10.1016/j.cobeha.2021.05.004

Tian, F., Hou, Y., Zhu, W., Dietrich, A., Zhang, Q., Yang, W., Chen, Q., Sun, J., Jiang, Q., & Cao, G. (2017). Getting the Joke: Insight during Humor Comprehension – Evidence from an fMRI Study. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01835

Tompkins, C. A. (1990). Knowledge and Strategies for Processing Lexical Metaphor after Right or Left Hemisphere Brain Damage. *Journal of Speech, Language, and Hearing Research*, *33*(2), 307–316. https://doi.org/10.1044/jshr.3302.307

Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-Analysis of the Functional Neuroanatomy of Single-Word Reading: Method and Validation. *NeuroImage*, *16*(3, Part A), 765–780. https://doi.org/10.1006/nimg.2002.1131

Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., & Fox, P. (2012). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Human Brain Mapping*, *33*(1), 1–13. https://doi.org/10.1002/hbm.21186

van Ackeren, M. J., Casasanto, D., Bekkering, H., Hagoort, P., & Rueschemeyer, S.-A. (2012). Pragmatics in Action: Indirect Requests Engage Theory of Mind Areas and the Cortical Motor Network. *Journal of Cognitive Neuroscience*, *24*(11), 2237–2247. https://doi.org/10.1162/jocn_a_00274

van Ackeren, M. J., Smaragdi, A., & Rueschemeyer, S.-A. (2016). Neuronal interactions between mentalising and action systems during indirect request processing. *Social Cognitive and Affective Neuroscience*, *11*(9), 1402–1410. https://doi.org/10.1093/scan/nsw062

Van Lancker, D. R., & Kempler, D. (1987). Comprehension of familiar phrases by left- but not by right-hemisphere damaged patients. *Brain and Language*, *32*(2), 265–277. https://doi.org/10.1016/0093-934X(87)90128-3

Varga, E., Schnell, Zs., Tényi, T., Németh, N., Simon, M., Hajnal, A., Horváth, R. A., Hamvas, E., Járai, R., Fekete, S., & Herold, R. (2014). Compensatory effect of general cognitive skills on non-literal language processing in schizophrenia: A preliminary study. *Journal of Neurolinguistics*, *29*, 1–16. https://doi.org/10.1016/j.jneuroling.2014.01.001

Varga, E., Simon, M., Tényi, T., Schnell, Zs., Hajnal, A., Orsi, G., Dóczi, T., Komoly, S., Janszky, J., Füredi, R., Hamvas, E., Fekete, S., & Herold, R. (2013). Irony comprehension and context processing in schizophrenia during remission – A functional MRI study. *Brain and Language*, *126*(3), 231–242. https://doi.org/10.1016/j.bandl.2013.05.017

Vartanian, O. (2012). Dissociable neural systems for analogy and metaphor: Implications for the neuroscience of creativity. *British Journal of Psychology*, *103*(3), 302–316. https://doi.org/10.1111/j.2044-8295.2011.02073.x

Virtue, S., Haberman, J., Clancy, Z., Parrish, T., & Jung Beeman, M. (2006). Neural activity of inferences during story comprehension. *Brain Research*, *1084*(1), 104–114. https://doi.org/10.1016/j.brainres.2006.02.053

Virtue, S., Parrish, T., & Jung-Beeman, M. (2008). Inferences during Story Comprehension: Cortical Recruitment Affected by Predictability of Events and Working Memory Capacity. *Journal of Cognitive Neuroscience*, *20*(12), 2274–2284. https://doi.org/10.1162/jocn.2008.20160

Vrticka, P., Black, J. M., & Reiss, A. L. (2013). The neural basis of humour processing. *Nature Reviews Neuroscience*, *14*(12), 860–868. https://doi.org/10.1038/nrn3566

Wakusawa, K., Sugiura, M., Sassa, Y., Jeong, H., Horie, K., Sato, S., Yokoyama, H., Tsuchiya, S., Inuma, K., & Kawashima, R. (2007). Comprehension of implicit meanings in social situations involving irony: A functional MRI study. *NeuroImage*, *37*(4), 1417–1426. https://doi.org/10.1016/j.neuroimage.2007.06.013

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Developmental changes in the neural basis of interpreting communicative intent. *Social Cognitive and Affective Neuroscience*, *1*(2), 107–121. https://doi.org/10.1093/scan/nsl018

Wehbe, L., Blank, I. A., Shain, C., Futrell, R., Levy, R., von der Malsburg, T., Smith, N., Gibson, E., & Fedorenko, E. (2021). Incremental language comprehension difficulty predicts activity in the language network but not the multiple demand network. *Cerebral Cortex*, *31*(9), 4006–4023. https://doi.org/10.1093/cercor/bhab065

Weylman, S. T., Brownell, H. H., Roman, M., & Gardner, H. (1989). Appreciation of indirect requests by left- and right-brain-damaged patients: The effects of verbal context and conventionality of wording. *Brain and Language*, *36*(4), 580–591. https://doi.org/10.1016/0093-934X(89)90087-4

Whyte, E. M., & Nelson, K. E. (2015). Trajectories of pragmatic and nonliteral language development in children with autism spectrum disorders. *Journal of Communication Disorders*, *54*, 2–14. https://doi.org/10.1016/j.jcomdis.2015.01.001

Willems, R. M., de Boer, M., de Ruiter, J. P., Noordzij, M. L., Hagoort, P., & Toni, I. (2010). A Dissociation Between Linguistic and Communicative Abilities in the Human Brain. *Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0956797609355563

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Winner, E., Brownell, H., Happé, F., Blum, A., & Pincus, D. (1998). Distinguishing Lies from Jokes: Theory of Mind Deficits and Discourse Interpretation in Right Hemisphere Brain-Damaged Patients. *Brain and Language*, *62*(1), 89–106. https://doi.org/10.1006/brln.1997.1889

Winner, E., & Gardner, H. (1977). The comprehension of metaphor in brain-damaged patients. *Brain*, *100*(4), 717–729. https://doi.org/10.1093/brain/100.4.717

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Basil Blackwell.

Woolgar, A., Parr, A., Cusack, R., Thompson, R., Nimmo-Smith, I., Torralva, T., Roca, M., Antoun, N., Manes, F., & Duncan, J. (2010). Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, *107*(33), 14899–14902. https://doi.org/10.1073/pnas.1007928107

Wu, X., Guo, T., Zhang, C., Hong, T.-Y., Cheng, C.-M., Wei, P., Hsieh, J.-C., & Luo, J. (2021). From "Aha!" to "Haha!" Using Humor to Cope with Negative Stimuli. *Cerebral Cortex*, *31*(4), 2238–2250. https://doi.org/10.1093/cercor/bhaa357

Yang, F. G., Edens, J., Simpson, C., & Krawczyk, D. C. (2009). Differences in task demands influence the hemispheric lateralization and neural correlates of metaphor. *Brain and Language*, *111*(2), 114–124. https://doi.org/10.1016/j.bandl.2009.08.006

Yang, J. (2014). The role of the right hemisphere in metaphor comprehension: A meta-analysis of functional magnetic resonance imaging studies. *Human Brain Mapping*, *35*(1), 107–122. https://doi.org/10.1002/hbm.22160

Yang, J., Li, P., Fang, X., Shu, H., Liu, Y., & Chen, L. (2016). Hemispheric involvement in the processing of Chinese idioms: An fMRI study. *Neuropsychologia*, *87*, 12–24. https://doi.org/10.1016/j.neuropsychologia.2016.04.029

Yang, J., & Shu, H. (2016). Involvement of the Motor System in Comprehension of Non-Literal Action Language: A Meta-Analysis Study. *Brain Topography*, *29*(1), 94–107. https://doi.org/10.1007/s10548-015-0427-5

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

Yarkoni, T., Speer, N. K., & Zacks, J. M. (2008). Neural substrates of narrative comprehension and memory. *NeuroImage*, *41*(4), 1408–1425. https://doi.org/10.1016/j.neuroimage.2008.03.062

Yi, Y. G., Kim, D. Y., Shim, W. H., Oh, J. Y., Kim, S. H., & Kim, H. S. (2017). Neural correlates of Korean proverb processing: A functional magnetic resonance imaging study. *Brain and Behavior*, *7*(10), e00829. https://doi.org/10.1002/brb3.829

Zaidel, E., Kasher, A., Soroker, N., & Batori, G. (2002). Effects of Right and Left Hemisphere Damage on Performance of the "Right Hemisphere Communication Battery." *Brain and Language*, *80*(3), 510–535. https://doi.org/10.1006/brln.2001.2612

Zempleni, M.-Z., Haverkort, M., Renken, R., & A. Stowe, L. (2007). Evidence for bilateral involvement in idiom comprehension: An fMRI study. *NeuroImage*, *34*(3), 1280–1291. https://doi.org/10.1016/j.neuroimage.2006.09.049