The Neural Basis of Non-Literal Language Processing:

A Meta-Analysis of Functional Magnetic Resonance Imaging Studies

A Senior Honors Thesis for the Cognitive & Brain Sciences Major, Department of Psychology

Miriam Hauptman

Tufts University, 2019

Thesis Committee: Idan Blank, Ph.D., Evelina Fedorenko, Ph.D.,
Ariel Goldberg, Ph.D., Jayden Ziegler

Abstract

Researchers of non-literal language (e.g., humor, metaphor, indirect speech) have traditionally sought to establish whether different cognitive mechanisms support non-literal vs. literal language processing. Previous approaches treated processing time and hemispheric lateralization as indicators of the recruitment of different cognitive mechanisms in non-literal language processing. In the present meta-analysis, we attempted to overcome methodological challenges of these paradigms by exploring the extent to which non-literal language processing is associated with neural activation in three distinct functional networks. Specifically, we analyzed whether activation peaks from 71 fMRI experiments (770 total peaks) reported in contrasts of non-literal vs. literal language processing fell into the language, multiple demand (MD), and theory of mind (ToM) networks. These networks become robustly and selectively activated when individuals perform tasks that engage cognitive processes presumably involved in non-literal language processing. Our analysis indicated that 1) peaks from non-literal vs. literal language contrasts fell into the language and ToM networks at above-chance levels, 2) a high number of the peaks are located in the right hemisphere, and 3) different types of non-literal language phenomena can be separated into three groupings based on their relationships to the functional networks of interest. Together, our findings suggest that different combinations of cognitive resources may indeed support non-literal vs. literal language processing, though more work is needed to clarify the exact nature of these differences. Future investigations into the neural basis of non-literal language processing should continue to incorporate functional network-based analyses in order to further explore the relationships between linguistic processing, executive function, and mentalizing and non-literal language processing.

The Neural Basis of Non-Literal Language Processing:

A Meta-Analysis of Functional Magnetic Resonance Imaging Studies

Understanding non-literal language processing is no piece of cake

A baked confection or a lack of difficulty? The idiomatic phrase “a piece of cake” takes on different meanings in different linguistic and social contexts. It is precisely this sensitivity to information about speaker intent, social circumstances, linguistic patterns, and world knowledge that defines non-literal or “pragmatic” language (Glucksberg, 1991; Wilson & Sperber, 2003; Wilson & Sperber, 2012). Research on non-literal language has traditionally sought to identify the cognitive mechanisms that drive non-literal vs. literal language processing. Although evidence from cognitive psychology, neuropsychology, and neuroscience has contributed important insights to this debate, there is still a considerable amount of ambiguity surrounding current understandings of the cognitive mechanisms involved in non-literal language processing (Kasparian, 2013; Yang, Edens, Simpson & Krawczyk, 2009). The present neuroimaging meta-analysis will attempt to improve these understandings by employing a novel analysis approach. Specifically, we will examine the extent to which non-literal language processing is associated with neural activation in distinct functional networks as measured by functional magnetic resonance imaging (fMRI).

In the following sections, we first explain what is meant by “non-literal language” before outlining several cognitive processes whose contributions to non-literal language processing deserve further investigation. We subsequently review past approaches to studying non-literal language processing and describe how the present study offers a different perspective on this topic. Ultimately, the complexity of scientific discourse on non-literal language processing suggests that understanding the mechanisms that support it will be far from a piece of cake.

Non-literal language: a cheat sheet

The "piece of cake" idiom is one example of non-literal language among many. From idioms to irony, non-literal language is marked by its dependence on contextual information and, therefore, its importance in social communication (Sperber & Wilson, 1986). Consistent with previous meta-analyses of non-literal language (e.g., Bohrn, Altmann, & Jacobs, 2012; Reyes-Aguilar, Valles-Capetillo, & Giordano, 2018), the present study will treat multiple non-literal language phenomena as a single class. Descriptions of these phenomena are displayed in Table 1. While they can be defined separately, all non-literal language types listed in Table 1 share a dependence on contextual information, or information that cannot be derived solely from the literal meaning of their linguistic components. Note, however, that grouping these phenomena may prove inappropriate, as it is possible that they represent distinct kinds. We will return to the issue of grouping non-literal language types in the discussion.

Metaphor, a classic example of non-literal language, has been studied most heavily. The process of linking a concrete source (e.g., "money") with an abstract target (e.g., "time"), such as in the example, "You need to budget your time" (Lakoff & Johnson, 1980, p. 456), requires that speakers retrieve information from memory and relate it to the current context (Glucksberg, McGlone, Grodzinsky, & Amunts, 2001). Less prototypical types of non-literal language, such as prosody and text/discourse comprehension, bear similarly crucial relationships to contextual information. For instance, prosody, or vocal tone (Hellbernd & Sammler, 2016), serves as a cue to meaning in situations in which other cues are lacking (Snedeker & Trueswell, 2003). During text and discourse comprehension, listeners or readers maintain a "mental model" (Johnson-Laird, 1983) of the current situation by making social and linguistic inferences that allow for the construction of a text or conversation's broader thematic meaning (Graesser, Singer, & Trabasso, 1994; Mason & Just, 2009).

Which cognitive processes support non-literal language processing?

Non-literal language, like any other form of language, carries linguistic meaning that must be derived from its orthographic or phonological form. It would appear, then, that at least some of the cognitive resources that contribute to literal language processing are also recruited for non-literal language processing. Consistent with this prediction, correlations between non-literal and literal language skills in both normal and clinical populations are well-documented. Individuals with strong literal language abilities (e.g., large vocabularies) tend to excel at creating and understanding non-literal language (Beaty & Silvia, 2013; Whyte & Nelson, 2015), and patients with general language impairments perform poorly on non-literal language comprehension tasks (Papagno, Tabossi, Colombo, & Zampetti, 2004; Papagno & Genoni, 2004). In addition, past neuroimaging studies of non-literal language processing often report increased activation in a classic language region called the left inferior frontal gyrus (IFG), commonly known as Broca's area, during non-literal vs. literal language comprehension (e.g., Jin et al., 2009; Spotorno, Koun, Prado, Van Der Henst, & Noveck, 2012; Matsui et al., 2016; Feng et al., 2017). In short, non-literal language is still language.

While literal and non-literal language processing seem to rely on a similar set of cognitive machinery, researchers of non-literal language have traditionally hypothesized that non-literal language processing draws upon additional processing resources. The distinction between literal and non-literal language processing was first captured by Grice (1975) and Searle's (1979) theories on non-literal language processing, which comprise what is now called the standard pragmatic model. In accordance with the scientific trends of the time (e.g., Frazier

& Rayner, 1982; Coltheart, 1978), the standard pragmatic model describes that similar yet separate cognitive mechanisms enable non-literal vs. literal language processing (Grice, 1975; Searle, 1979). Specifically, non-literal language processing involves the accessing of literal before non-literal meaning in a serial manner (Clark & Lucy, 1975). Though many have criticized the idea of strictly serial linguistic processing (e.g., Marlsen-Wilson, 1987; MacDonald, Pearlmutter, & Seidenberg, 1994), insights from the standard pragmatic model have inspired decades of research aimed at understanding the extent to which additional cognitive resources support non-literal language processing.

Executive function and non-literal language processing

Scholars of non-literal language extending back to Grice (1975) have explained differences in literal and non-literal language processing by referencing the cognitively demanding nature of non-literal language processing. Echoing the idea that non-literal language processing relies on cognitive mechanisms beyond those recruited for literal language processing, Jung-Beeman (2005) posited that semantic integration, or the calculation of “semantic overlap among multiple semantic fields,” (p. 515) becomes especially demanding during non-literal language processing. Across different types of non-literal language, relating disparate semantic concepts and updating models of global context may prove taxing for semantic integration mechanisms (Jung-Beeman, 2005). Regarding more domain-general (i.e., not language-specific) cognitive functions, non-literal language processing ability is correlated with fluid intelligence (Silvia & Beaty, 2012; Dennis & Barnes, 1990), a measure of problem-solving ability (Cattell, 1963), and working memory capacity predicts pragmatic language skills in individuals with schizophrenia (Herold, Tenyi, Simon, & Trixler, 2004; Champagne-Lavau & Stip 2010; Varga et al., 2014) and autism (Schuh, Eigsti, & Mirman, 2016).

Many have attempted to explain the putative link between executive function, a general term used to describe inhibition, working memory, and cognitive flexibility (Diamond, 2013), and non-literal language processing by describing the cognitively taxing nature of individual non-literal language phenomena. According to these theories, idiom processing places a high inferential load on the comprehender (Colle et al., 2013, Honan, McDonald, Gowland, Fisher, & Randall, 2015), humor appreciation requires a “frame-shift” during which information from working memory and stored knowledge is compared (Coulson & Kutas, 2001), and discourse comprehension involves the maintenance and updating of common-ground information between interlocutors (Brown-Schmidt, 2009).

In their discussion on the role of inhibition in language processing, Gernsbacher and Robertson (1999) present a unified theory on the relationship between executive function and non-literal language processing. The authors believe that all possible interpretations of an utterance become activated at once, thus making the selection of the correct interpretation—and the suppression of the rest—a matter of inhibition. Inhibition becomes especially important when comprehending or producing non-literal language, whose alternate interpretations can vary widely and differ starkly from the intended meaning. Channon and Watts (2003) expand upon this idea by arguing that inhibition additionally supports the ability to consider social consequences of selecting a given interpretation. Evidence from neuropsychological investigations of executive function and non-literal language processing appears to support claims about the role of inhibition in non-literal language processing. Individuals with brain damage have demonstrated increased difficulty in understanding sarcasm (McDonald & Pearce, 1996) and inferring listener reluctance in conversation (McDonald & Pearce, 1998) as a function of their performance on tests of inhibition. Using multiple measures of executive function, Channon and Watts (2003) similarly demonstrated that patients' performance on inhibition, but not working memory or multitasking tasks predicted the extent of their inferencing and text comprehension abilities.

Mentalizing and non-literal language processing

Beyond executive function, the role of mentalizing may also help distinguish non-literal from literal language processing. The term "theory of mind," coined first by Premack and Woodruff (1978), and later referred to as "mentalizing" by Frith and Frith (1999), refers to the capacity to attribute mental states to others. A defining human quality, theory of mind (ToM) serves as a crucial tool in social communication (Saxe, 2006). Given the relationship between non-literal language and social communication (Sperber & Wilson, 1986), many theories of non-literal language processing have focused on the role of mentalizing. Wilson and Sperber (2012) specifically described that a specialized "sub-module" formed out of broader ToM abilities is employed during non-literal language processing (p. 270). The sub-module is not employed during non-linguistic tasks, such as when decoding others' visual focus (i.e., an application of ToM abilities that is required during joint attention tasks), but instead helps members of a communicative exchange compute the relative value, or relevance, of using contextual information to interpret an utterance (for the original version of Relevance Theory, see Sperber & Wilson, 1986). Understanding the environmental stimulus that will attract a conversational partner's attention, the background knowledge she is likely to use while processing this stimulus, and the resulting inferences she may make during communication allows us to engage in pragmatic exchanges while minimizing cognitive load and maximizing mutual comprehensibility (Wilson & Sperber, 2012).

As in the case of executive function, scientists have attempted to anchor more abstract theories on the relationship between mentalizing and non-literal language processing by examining individual types of non-literal language. Upon hearing an indirect request, for instance, listeners draw upon a model of the speaker's mind in order to distinguish requests from literal statements (e.g., to determine whether, "It's hot in here" means, "Please open the window" or simply, "It is warm in this room") (Van Ackeren, Casasanto, Bekkering, Hagoort, & Rueschemeyer, 2012). Similarly, Colston and Gibbs (2002) showed that individuals engage in mentalizing during irony comprehension, and Winner and Leekam (1991) reported that mentalizing allows children to distinguish between irony and deception. With respect to humor, Vrticka, Black, and Reiss (2013) defined two core processes of humor appreciation: an emotional component, or mirth, and a cognitive component, involving the ability to attribute mental states to others (i.e., ToM). Finally, metaphor processing is thought to depend partially upon the comprehender's ability to guess the speaker's communicative intentions (Eviatar & Just, 2006).

A possible connection between theory of mind capabilities and non-literal language processing is especially apparent in studies of individuals with brain damage and neurological disorders. Evidence from neuropsychology has repeatedly indicated that individuals with impaired ToM ability perform worse on tests of non-literal language processing (Martin & McDonald, 2005; Winner, Brownell, Happé, Blum, & Pincus, 1998). Similarly, ToM ability has been correlated with non-literal language processing deficits in individuals with schizophrenia (Champagne-Lavau & Stip, 2010; Gavilan & Garcia-Albea, 2011; Varga et al., 2013) and autism (Happé, 1995). Happé (1995) makes strong predictions about the role of ToM in simile, metaphor, and irony comprehension in individuals with autism, arguing that ToM impairments largely explain their deficits in using and understanding non-literal language. Similar discussions about the central role of ToM in non-literal language processing have arisen in studies of individuals with schizophrenia. A striking fMRI study on irony processing reported that normal controls showed activation in traditional ToM areas (Saxe & Powell, 2006) such as the left temporo-parietal junction (TPJ), while individuals with schizophrenia did not (Varga et al., 2013). However, when offered "linguistic help" (i.e., an adverb that described the mental state of speakers in the stimuli set, such as "Tom *angrily* remarks..."), individuals with schizophrenia improved on the irony comprehension task, and significant differences in activation between individuals with and without schizophrenia disappeared completely. The authors interpreted this result as evidence both for the role of mentalizing in irony comprehension in neurotypical people and for the effects of impaired ToM ability on non-literal language processing in individuals with schizophrenia (Varga et al., 2013). Consistent with these findings, neuroimaging studies on non-literal language processing in neurotypical individuals have repeatedly reported increased neural activation in cortical areas associated with ToM during non-literal vs. literal language comprehension (e.g., Mason & Just, 2009; Van Ackeren et al., 2012; Spotorno et al., 2012; Bašnáková et al., 2016).

Past and present methodological approaches

The relationship between executive function and theory of mind in non-literal language processing supports the notion that non-literal language, though similar to literal language in certain respects, draws upon a different combination of cognitive resources. Investigations concerned directly with identifying these differences have historically leveraged a variety of methodologies. One early intuition was that increased processing time signaled the recruitment of additional cognitive resources. However, as we will discuss in the following section, operationalizing differences in non-literal and literal language processing using behavioral reaction time faced notable challenges. Inspired by a series of intriguing neuropsychological findings, many argued that the right hemisphere's relationship to non-literal language processing could alternatively indicate the recruitment of additional processing resources. A review of this literature will reveal that evidence for differential hemispheric involvement in literal vs. non-literal language processing suffers from a lack of methodological consistency. Despite concerns with the lateralization approach, however, other methods investigating the neural underpinnings of non-literal language processing promise a more effective and sophisticated look at the cognitive mechanisms involved in non-literal language processing.

One such method involves examining the involvement of different functional networks, or collections of brain regions that show robust and selective activation when individuals perform a related set of tasks (Saxe, Schulz, & Jiang, 2006). The present study will specifically explore the role of the language, Multiple Demand (MD), and Theory of Mind (ToM) networks in non-literal language processing. These three networks support cognitive functions—linguistic processing, executive function, and mentalizing—said to contribute to non-literal language processing. In examining the role of the language, MD, and ToM networks in non-literal language processing, we aim to avoid the methodological weaknesses characteristic of past empirical approaches (Townsend, 1989; Kasparian, 2013) and thus improve current understandings of the cognitive mechanisms involved in non-literal language processing.

In this section, we introduced the concept of non-literal language and described several linguistic and non-linguistic cognitive processes associated with non-literal language processing. In the following sections, we will review the behavioral reaction time and hemispheric lateralization methodologies and defend the merit of the current approach.

A matter of time: behavioral evidence on non-literal language

In the years following Grice (1975) and Searle's (1979) proposal, scientists used behavioral reaction time to test whether individuals apprehend literal and non-literal meanings simultaneously or sequentially. If understanding the idiom "kick the bucket" necessarily involved the accessing of its literal (to physically kick a bucket) followed by its non-literal (to die) meaning, researchers predicted that reading a non-literal phrase should take longer than reading a literal one. Importantly, the existence of multiple processing steps (e.g., the accessing of literal meaning, the accessing of non-literal meaning) does not necessarily imply the need for additional processing time. It is only when these processing steps are executed serially, or when one step cannot begin until the previous step has concluded, that more processing time is traditionally expected (Marlsen-Wilson, 1987). As a serial processing model, the standard pragmatic model indeed offered the prediction that non-literal language processing is performed more slowly than literal language processing.

Early researchers of non-literal language operationalized expected differences in processing time by comparing behavioral reaction time measures. Contrary to the predominant hypotheses of the time, numerous experiments reported that participant reaction time to literal vs. non-literal language stimuli was identical (Ortony, Schallert, Reynolds & Antos, 1978; Gibbs, 1980). Gibbs (1980) even found that participants were faster at comprehending the non-literal vs. literal interpretations of conventional idioms. Additionally, Keysar (1989) observed that participants took longer to complete a sentence verification task when both literal and non-literal interpretations of a text were plausible, suggesting that both interpretations became available in parallel.

As participant reaction times to non-literal and literal stimuli were repeatedly shown to be identical across different experimental paradigms (e.g., Gibbs, 1984; Glucksberg, 2003; Ahrens et al., 2007), it appeared that the initial predictions of the standard pragmatic model could not be substantiated. That is, given putative evidence for parallel processing in non-literal language comprehension, a common mechanism was presumed to support both literal and non-literal language processing (Cacciara & Glucksberg, 1994). However, a call to reject the idea of separate mechanisms in non-literal language processing based on a single type of evidence merits a degree of skepticism. Indeed, there is some doubt about the internal validity of behavioral reaction time measures. For instance, behavioral reaction time naturally increases with age (Woods, Wyma, Yund, Herron, & Reed, 2015), which could confound experimental findings. Even in the most controlled experimental settings, there remains a strong possibility that reaction time data reflects a mere "local," or experiment-specific, phenomenon that cannot speak to broad questions about serial vs. parallel processing (Townsend, 1989, p. 52).

In the case that reaction time indeed serves as a direct measure of processing time, however, time-course represents only one dimension by which non-literal and literal language processing might differ. In the spirit of Grice (1975) and Searle's (1979) proposals, many scientists continued to argue that distinct combinations of cognitive mechanisms support literal vs. non-literal processing—this distinction may simply manifest in ways other than time-course. Evidence from the field of neuropsychology, for instance, suggested the potential relevance of hemispheric involvement in the effort to disentangle the processing mechanisms involved in non-literal vs. literal language processing. This will serve as the focus of the following section.

A tale of two hemispheres: non-literal language and lateralization

For over two decades, neuropsychological investigations shaped the prevailing theories on differences between literal and non-literal language processing. The push to study the non-literal language abilities of individuals with brain damage was initially prompted by Winner and Gardner's (1977) seminal findings. The authors observed that participants with right-hemisphere damage (RHD) performed worse on a metaphor comprehension task than did those with left-hemisphere damage (LHD), despite that the left hemisphere's central role in language processing had been well-established (Broca, 1861). Subsequent investigations documented deficits in individuals with RHD across many types of non-literal language, including idioms (Myers & Linebaugh, 1981; Van Lancker & Kempler, 1987), metaphors (Brownell, 1988), jokes (Brownell, Potter, Bihrle, & Gardner, 1986), indirect speech acts (Stemmer, Giroux, & Joanette, 1994; McDonald & van Sommers, 1993), and sarcasm (Giora, Zaidel, Soroker, Batori, & Kasher, 2000). In addition, patients with RHD demonstrated difficulties when completing tasks involving text and discourse comprehension (Gardner, Brownell, Wapner, & Michelow, 1983; Delis, Wapner, Gardner, & Moses, 1983; Ferstl, Walther, Guthke, & von Cramon, 2005).

These findings garnered attention not only due to the extent of impairments reported in people with RHD, but also because of the skills that these individuals often retained. In his study on RHD and inferencing, Beeman (1993) invoked a striking quotation from a participant with RHD who claimed that he seemed, "to be missing the subtleties, the complex mosaic of meaning that is language" (p. 113). Indeed, many individuals with RHD demonstrated normal literal language abilities; their deficits seemed to emerge only in more "cognitively sophisticated" linguistic environments (Johns, Tooley & Traxler, 2008, p. 1038). Interestingly, some people with RHD successfully identified the need to reinterpret non-literal language apart from its literal meaning but seemed unable to *carry out* the necessary frame-shift (for a discussion on frame-shifting in humorous language, see Coulson, 2001). For instance, Brownell et al. (1986) showed that patients with RHD chose non-sequiturs over both literal and comedic endings when attempting a joke completion task. That is, the patients seemed to recognize that selecting a literal ending was inappropriate but could not judge the relevance of non-sequiturs vs. non-literal endings. McDonald and van Sommers (1993) similarly reported the case of an individual with RHD who could interpret the non-literal meaning of indirect requests (e.g., “Can you pass the salt?”), but could not reject a literal interpretation when the experimenters presented it as an alternative. Brownell et al. (1986) and McDonald and van Sommers’ (1993) studies, along with several others (e.g., Bihrle, Brownell, Powelson, & Gardner, 1986), implicated the role of the right hemisphere in executing a final decision step in non-literal interpretation selection.

In light of the promulgation of strong claims regarding non-literal language and the right hemisphere, the idea that the “right hemisphere is special” (Kasparian 2013, p. 2) in non-literal language processing likely deserves a degree of scrutiny. Perhaps the most obvious concern is that empirical findings do not always substantiate these claims. Several neuropsychological studies have reported that individuals with RHD and LHD show similar deficits in metaphor comprehension (Tompkins, 1990; Giora et al., 2000; Zaidel, Kasher, Soroker, & Batori, 2002; Gagnon, Goulet, Giroux, & Joanette, 2003). A recent study by Cardillo, McQuire and Chatterjee (2018) found no difference in the ability of 20 individuals with RHD and 20 individuals with LHD to accurately match a metaphorical phrase to its meaning; both groups demonstrated significant impairment. Similarly, neuroimaging studies investigating differences in neural activation during non-literal vs. literal language comprehension do not uniformly provide evidence for right hemisphere involvement. In fact, many studies do not report any activation in the right hemisphere during non-literal vs. literal language comprehension at all (e.g., Rapp, Leube, Erb, Grodd, & Kircher, 2004; Lee & Dapretto, 2006; Citron & Goldberg, 2014; Samur, Lai, Hagoort, & Willems, 2015; Yi et al., 2017). It is not surprising, then, that meta-analyses of non-literal language have repeatedly described bilateral yet highly left-lateralized networks in non-literal vs. literal language comprehension (Bohrn et al., 2012; Reyes-Aguilar et al., 2018), with similar left vs. right hemisphere activation peak breakdowns as observed in a meta-analysis of literal language processing (Rapp, Mutschler & Erb, 2012; see Binder, Desai, Graves, & Conant, 2009 for the literal language meta-analysis).

Despite the emergence of a more complicated picture regarding the right hemisphere's role in non-literal language processing, the 1990s and early 2000s saw a continuation of support for the "right hemisphere is special" position. Hypotheses from the fields of psycholinguistics (Giora, 1997) and cognitive neuroscience (Beeman & Chiarello, 1998; Jung-Beeman 2005) continued to point to the right hemisphere as a vital component in the comprehension and production of (especially novel or unusual) non-literal language. Consistent with these frameworks, many neuroimaging studies of non-literal language reported increased activation in the right hemisphere when comparing neurotypical participants' neural responses to non-literal vs. literal linguistic stimuli (e.g., Bottini et al., 1994; Mashal, Faust, & Hendler, 2005; Mashal, Faust, Hendler, & Jung-Beeman, 2007; Schmidt, DeBuse, & Seger, 2007; Zempleni, Haverkort, Renken, & Stowe, 2007; Bašnáková, van Berkum, Weber, & Hagoort, 2016). Bottini et al.'s (1994) pioneering PET study of metaphor comprehension, for example, revealed a correlation between metaphor comprehension and increased neural activation in the right hemisphere that did not arise during literal sentence comprehension conditions. Lezak's (1995) proposal that the right hemisphere processes visual information more holistically than the left further encouraged theories of right hemisphere involvement in non-literal language—the appreciation of the relationship between language and context (Joanette, Goulet, Hannequin, & Boeglin, 1990), or the diffuse activation of semantic information (Jung-Beeman, 2005), seemed to fit nicely with the idea of the right hemisphere as a holistic processor (Martin & McDonald, 2003).

Methodological issues may help explain the apparent heterogeneity in previous investigations of non-literal language. First, many early neuropsychological studies of non-literal language did not report the extent or location of right hemisphere damage (Schmidt, Kranjec, Cardillo, & Chatterjee, 2010). Experts in neuropsychology maintain that averaging the performance of individuals grouped only because of grossly defined damage is highly problematic (McCloskey & Caramazza, 1988). Additionally, neuroimaging and neuropsychological investigations tend to make use of different types of stimuli (sentences vs. word pairs, respectively), which complicates comparisons—or claims of consistency—between the two sets of empirical work (Cardillo et al., 2018). A particularly troubling aspect of neuroimaging research concerns the potential presence of task-related artifacts (Kasparian, 2013). In a study devoted specifically to measuring the effect of task on activation during non-literal language processing, Yang et al. (2009) found that task difficulty appeared to modulate activation in the right hemisphere. This finding is consistent with the observation that neuroimaging studies that control for the length and processing difficulty of their stimuli often report weaker effects in the right hemisphere (e.g., Rapp et al., 2004; Citron & Goldberg, 2014). Finally, it should be noted that Winner and Gardner's (1977) famous study—which arguably inspired decades of research on RHD and non-literal language—described that patients with RHD performed poorly on a pictorial matching task involving metaphor, whereas their performance on a verbal metaphor task was spared. Rinaldi, Marangolo, and Baldassarri (2004) replicated these results while controlling for visuospatial deficits in a group of 50 participants with RHD, indicating that RHD may be associated with general deficits in integrating information across different representation modalities rather than with a specific deficit in processing non-literal language.

The argument for the special role of the right hemisphere in non-literal language processing initially found support among an interdisciplinary cross-section of the research community. It appears, however, that several crucial features of the lateralization debate have sometimes been overlooked. Beyond the fact that many studies do not implicate the role of the right hemisphere in non-literal language processing, a lack of standardization in neuropsychological and neuroscientific methods makes comparing work between the two fields a difficult task. Taken together, these concerns pose a considerable challenge to the once-predominant narrative regarding right hemisphere involvement in non-literal language processing. Indeed, in line with a recent shift in cognitive neuroscience toward the idea of bilateral distributed processing models (Bambini, Gentili, Ricciardi, Bertinetto, & Pietrini, 2011), many researchers of non-literal language have adopted milder views on the role of the right hemisphere in non-literal language processing (e.g., Kuperberg, McGuire, Bullmore, & Brammer, 2000; Rapp et al., 2004; Kuperberg, Lakshmanan, Caplan, & Holcomb, 2006). An increasingly common hypothesis suggests that the right hemisphere may not be tuned specifically for non-literal language but is instead recruited in especially demanding experimental settings (Johns et al., 2008; Schmidt & Seger, 2009). With decades of methodological and theoretical concerns complicating the debate on hemispheric specialization in non-literal language processing, however, discussions on the role of the right hemisphere are unlikely to lead to fruitful conclusions about the mechanisms that drive non-literal language processing. It is for this reason that the present study aims to investigate non-literal language processing from a different perspective—that of functional networks.

Birds of a feather flock together: functional networks

McDonald and Pearce (1998) offered an intriguing observation about non-literal language processing in their neuropsychological study of individuals with RHD. According to the authors, the preservation of literal language abilities in individuals with non-literal language deficits indicates that these deficits may stem from more general cognitive impairments that emerge during "the application of [non-literal] language skills" (p. 89). In other words, non-literal language processing may involve a variety of cognitive operations drawn from both linguistic and non-linguistic domains. In the second section of the introduction ("A cheat sheet"), we described the relationship between non-literal language processing and linguistic processing, executive function, and mentalizing. These cognitive processes are associated with distinct neural signatures. In the present study, we will capitalize on the existence of separable *functional networks* involved in language processing, executive function, and mentalizing to investigate how each of these three cognitive functions may be associated with non-literal language processing.

First, what is a functional network? Inspired by the scientific impulse to decompose a phenomenon into its component parts (Bechtel, 2002), cognitive neuroscientists have sought to identify functional networks, or collections of brain regions that show robust and selective activation when individuals perform a related set of tasks (Saxe et al., 2006). Crucially, the existence of a functional network does not necessarily point to the encapsulation or innateness (cf. Fodor, 1983) of the system in question, but instead may reflect the storage of "domain-specific knowledge representations that mediate" a specialized set of cognitive operations

(Fedorenko & Varley 2016, p. 145). Described in computational terms, the computations carried out in functional networks seem to take only certain type of data structure as input. In their discussion of the language network, Fedorenko and Varley (2016) acknowledge the possibility that the cognitive function that engages it (i.e., linguistic processing) could be comprised of multiple separable subprocesses, though limited evidence currently supports this notion. Still, due in part to methodological limitations, the existence of a functional network does not necessarily indicate that all components of the network operate uniformly. We will now describe each functional network of interest in turn.

Executive function: The MD network

The "multiple demand" (MD) network includes areas of cortex in the prefrontal and parietal regions (Duncan & Owen, 2000; Duncan, 2010; Fig. 1). Specifically, the MD network is marked by selective and robust activation "in and around the posterior part of the inferior frontal sulcus, in the anterior insula and adjacent frontal operculum, in the presupplementary motor area and adjacent dorsal anterior cingulate, and in and around the intraparietal sulcus" during a wide variety of tasks involving attention, working memory, perceptual categorization, and problem solving (Duncan 2010, p. 172; Federenko, Duncan, & Kanwisher 2013). Activation throughout these regions is modulated by stimulus complexity (Duncan 2010; Downar, Crawley, Mikulis, & Davis, 2001; Downar, Crawley, Mikulis, & Davis, 2000), perceptual difficulty (i.e., degree of stimulus degradation) (Duncan & Owen, 2000), and task novelty/switching (Dove, Pollmann, Schubert, Wiggins, & von Cramon, 2000; Dumontheil, Thomson, & Duncan, 2011) across verbal (Fedorenko, 2014) and nonverbal tasks, and is even observed during transitions between musical movements (Sridharan et al., 2007; Kurby & Zacks, 2008).

Work from a range of scientific disciplines provides converging evidence for the existence of a domain-general and functionally distinct MD network. Multi-voxel pattern analysis (Haxby et al., 2001) has implicated these regions in the encoding of categorical decision-making (Li et al., 2007). On a cellular level, neurons in the prefrontal cortex engage in "adaptive coding," meaning that they "adapt their properties to carry specifically information that is relevant to current concerns" (Duncan 2001, p. 820). Neuropsychological research has suggested a causal relationship between MD regions and fluid intelligence abilities. Damage to many of the regions outlined by Duncan (2010), including the intraparietal sulcus and inferior frontal sulcus, predicted fluid intelligence deficits in a study of 80 patients with brain damage (Woolgar et al., 2010). Crucially, lesion extent in language regions (discussed in the following section) did *not* predict fluid intelligence deficits (Woolgar et al., 2018), indicating that these two networks serve distinct functional purposes. Using intersubject correlation (ISC) as an index of input tracking, Blank and Fedorenko (2017) found that activations in language regions were highly synchronized across participants during a naturalistic story comprehension task, while activations in the MD regions of different participants were less consistent. They interpreted this result as evidence for the possibility that the MD network does not treat linguistic information as input, though the MD network may become engaged idiosyncratically as it "comes to the rescue" during instances of particularly demanding linguistic processing (Blank & Fedorenko, 2017, p. 10008). In addition to language regions, the default network (Grecius, Krasnow, Reiss, & Menon, 2003), which becomes activated during rest and deactivated during task completion, appears to be functionally separable from the MD network (Fox et al., 2005; Hugdahl, Raichle, Mitra & Specht, 2015; Mineroff, Blank, Mahowald, & Fedorenko, 2018).

Linguistic processing: The language network

The language network is comprised of a set of regions in the frontal and temporal lobes (Binder et al., 1997), including the inferior frontal gyrus ("Broca's area"), that responds robustly and selectively during linguistic processing (Fedorenko, Hsieh, Nieto-Castañón, Whitfield-Gabrieli, & Kanwisher, 2010; Fig. 2). Specifically, the language network is engaged during what is called "high-level linguistic processing," or the extraction of linguistic meaning (Fedorenko & Varley, 2016), and can be operationalized by comparing neural activation when participants read sentences vs. nonwords (Fedorenko et al., 2010; 2016). Recent evidence indicates that the classic idea of a syntax/semantics dichotomy does not emerge in the language network. Mollica et al. (2019) showed that syntax degradedness did not predict lowered activation in language regions when mutual information remained high. That is, sentences containing scrambled words that retained some level of composite meaning did not elicit a different neural response than unscrambled (i.e., normal) sentences. Only sentences scrambled to the point of total meaninglessness resulted in decreased activation in the language network. Consistent with this finding, Fedorenko et al. (2018) reported that certain regions within the language network appeared more involved in semantic vs. syntactic processing, though no regions were modulated more by syntactic processing. Together, these studies suggest that the language network is engaged chiefly by meaningful linguistic input.

As in the case of the MD network, numerous findings indicate that the language network engages in a selective set of cognitive processes. Fedorenko, Behr, and Kanwisher (2011) defined language regions in individual participants by subtracting activation during sentence vs. non-word comprehension and found that arithmetic, logic-puzzle, visuospatial working memory, and Stroop tasks did not elicit significant activation in these regions. In accordance with this finding, a neuropsychological study reported that severely aphasic individuals nevertheless demonstrated intact mathematical skills (Varley, Klessinger, Romanowski & Siegal, 2005). In addition, viewing videos of hands, faces, and objects did not appear to engage the language network (Pritchett et al., 2018). Mineroff et al. (2018) recently found that activation effect sizes in the language network were highly correlated within the language network, but did not correlate with activation effect sizes in the MD and default mode networks. In sum, the type of processing that engages the language network appears separable from many aspects of cognition (Varley & Fedorenko, 2016; Monti, Parsons & Osherson, 2012).

Mentalizing: The ToM network

The ToM network, including the temporo-parietal junction (TPJ) and medial prefrontal cortices, becomes robustly activated when humans reason about others' mental states, engage in shared attention tasks, and experience feelings of emotional empathy (Saxe & Kanwisher, 2003; Saxe, 2006; Saxe & Powell, 2006; Fig. 3). These regions appear to serve a set of cognitive functions apart from executive function and high-level linguistic processing. Non-overlapping brain regions are associated with executive function and ToM (Saxe et al., 2006), and several cases of aphasic individuals with intact performance on nonverbal false belief tasks have been reported (Apperly, Samson, Carroll, Hussain, & Humphreys, 2006; Varley & Siegal, 2001; Varley & Siegal, 2000). Yet, performance on ToM tasks is correlated with knowledge of metacognitive vocabulary (believe, feel, etc.) in children (Lockl & Schneider, 2006; Antonietti, Sempio, & Marchetti, 2006), and ToM-related processing appears to modulate a small region in the language network called the superior temporal sulcus (Deen, Koldewyn, Kanwisher, & Saxe, 2015). Ultimately, however, the existence of functionally selective networks does not imply a lack of interfacing between these networks. It is precisely this possibility of interaction that motivates the present investigation of the relationship between the language, MD, and ToM networks during non-literal language processing.

Cutting to the chase: the present study

The present study adopts a novel approach to neuroimaging meta-analysis by comparing the locations of activation peaks reported in previous studies with the locations of the language, MD, and ToM networks. In this section, we attempt to justify the efficacy of this methodology before outlining hypotheses about potential findings.

Neuroimaging meta-analysis

Meta-analysis is an important methodological tool in increasing the statistical power of neuroimaging experiments. In response to an explosion of publications on fMRI in the past few decades, many scientists feel a “growing need to integrate research findings and establishing consistency across labs and...scanning procedures” (Wager, Lindquist & Kaplan, 2007, p. 150). An improvement over comparing broad anatomical labels across studies, voxel-based meta-analysis increases statistical power and decreases the chance of observing false positives by aggregating activation peak data from a larger participant pool (Wager et al., 2007; Costafreda, 2009). Two common voxel-based neuroimaging meta-analysis techniques are Effect Size Signed Differential Mapping (ES-SDM) (Radua et al., 2012) and Activation Likelihood Estimation (ALE) (Turkeltaub et al., 2002; Laird et al., 2005). The former incorporates activation peak effect sizes while estimating the distribution of activation across experiments. ALE, as its name suggests, provides an estimate of “the likelihood that at least one of the foci in the literature should have been reported at this voxel” (Turkeltaub & Coslett, 2010, p. 2).

There are a few important disadvantages associated with conducting these types of neuroimaging meta-analysis. Typically, neuroimaging studies report significant results only, which leaves meta-analysis algorithms with no choice but to treat all unknown activations as possessing an effect size of zero, thereby biasing findings (Costafreda, 2009). Moreover, traditional neuroimaging meta-analysis and individual group analysis techniques suffer from a similar set of problems. In a discussion on the potential limitations of their meta-analysis, Rapp et al. (2012) acknowledge that ALE may wash out clusters containing more spatially distributed voxels. In other words, this approach appears to suffer from reduced sensitivity, or the ability to detect activation when it is truly present (see the following section for a discussion on challenges in traditional group analysis).

In the past decade, multiple neuroimaging meta-analyses of non-literal language have been conducted using ES-SDM and ALE. These include meta-analyses on text comprehension (Ferstl, Neumann, Bogler & von Cramon, 2007), metaphor (Vartanian, 2011; Yang, 2014), humor (Vrticka et al., 2013) and lying (Lisofsky, Kazzer, Heekeren & Prehn, 2014), as well as several that aggregate results from experiments involving multiple types of non-literal language (Bohrn et al., 2012; Rapp et al., 2012; Reyes-Aguilar et al., 2018). The latter group of studies report a predominantly left-lateralized non-literal language processing network with consistent activation in traditional language, MD, and ToM areas across included studies (Bohrn et al., 2012; Rapp et al., 2012; Reyes-Aguilar et al., 2018). Previous meta-analyses of non-literal language have focused at maximum on only 4 types of non-literal language across 50 different experiments (Reyes-Aguilar et al., 2018), and their inclusion criteria are sometimes not sufficiently stringent (i.e., experiments using highly variable paradigms and contrasts with low-level baselines are included) (e.g., Bohrn et al., 2012; Rapp et al., 2012; Yang, 2014). In the present study, we seek to expand upon previous meta-analyses of non-literal language by including a larger number of studies across a wider range of non-literal language phenomena while adhering to selective exclusion criteria and incorporating information about functional networks in our analysis.

Why functional networks?

A central practice in neuroimaging research involves comparing functional profiles across individual brains in order to increase the replicability and generalizability of findings. Typically, this comparison is carried out by fitting, or normalizing, each participant's brain to a three-dimensional coordinate (i.e., stereotaxic) space (e.g., the Montreal Neurological Institute [MNI] template, Evans et al., 1993; Collins, Neelin, Peters, & Evans, 1994) before conducting a random effects analysis (Holmes & Friston, 1998) to identify group-based functional Regions of Interest (fROIs). Statistically significant activation is determined based on the proportion of participants who show activation at a given coordinate in common space.

While normalization represents an important attempt to generalize functional data, concerns about the use of this technique have increased in recent years. Just as our faces look different, so too do our brains (Saxe, Brett & Kanwisher, 2006): each individual brain differs in size, shape, cytoarchitectonic boundaries, and folding patterns (Ono, Kubik, Abernathey, & Yasargil, 1990; Paus et al., 1996; Tomaiuolo et al., 1999; Juch et al., 2005)—especially in high-level cortical areas in the temporal (Jones & Powell, 1970; Gloor, 1997; Wise et al., 2001; Tomaiuolo et al 1999), parietal (Caspers et al., 2006; Caspers et al., 2008; Scheperjans et al., 2008) and frontal (Amunts et al., 1999; Miller et al., 2002) lobes. Anatomical variability poses a considerable challenge to normalization techniques, as even the most advanced methods (e.g., Fischl, Sereno, Tootell, & Dale, 1999) cannot overcome all dimensions of variability (Frost &

Goebel, 2012). As a result, sensitivity, or the extent to which significant activation is detected when truly present, and functional resolution, or the ability to keep adjacent but functionally distinct neural responses separate, both suffer (Swallow, Braver, Snyder, Speer, & Zacks, 2003; Nieto-Castañón & Fedorenko, 2012). A lack of sensitivity and functional resolution may leave traditional group analysis techniques underpowered (Frost & Goebel, 2012), especially in studies of high-level cognitive functions such as language (Nieto-Castañón & Fedorenko, 2012).

Fedorenko et al. (2010) present a promising alternative to such techniques called group-constrained, subject-specific (GcSS) analysis. In GcSS analysis, probabilistic overlap maps constructed by comparing the relative number of participants who show significant activation at a given location are used to derive group-level partitions via a segmentation algorithm (Meyer, 1991). The group-level partitions can then be used to carry out a variety of further analyses. GcSS analysis improves upon traditional group analysis approaches in several respects (see Nieto-Castañón & Fedorenko, 2012 for a discussion). Most important for the purposes of the present study, GcSS benefits from increased sensitivity by relying on probabilistic overlap maps instead of normalization techniques to identify group-level partitions. This means that activation that does not overlap in large numbers of participants still contributes to the definition of the group-level partitions. Increased sensitivity increases the statistical power of neuroimaging experiments, which has become a major goal for neuroimaging research in light of the replication crisis (Button et al., 2013; Poldrack et al., 2017).

The current approach

In the present analysis, we will test activation peaks derived via traditional group analysis in existing studies against the group-level partitions and probabilistic overlap maps associated with the partitions that comprise the language, MD, and ToM networks. To address the scientific debate on the cognitive mechanisms involved in non-literal vs. literal language processing, we will only include activation peak data from non-literal vs. literal language contrasts. This contrast targets the neural resources recruited above those required during literal language processing and might therefore help inform theories on the role of additional cognitive mechanisms in non-literal language processing. As mentioned in the previous section, the language, MD, and ToM networks become selectively activated during a related set of tasks, meaning that the voxels contained within them do not show robust response profiles under many other conditions (e.g., Fedorenko et al., 2011; Pritchett et al., 2018; Blank & Fedorenko, 2017; Saxe et al., 2006). For instance, robust activation is observed in the language network when individuals read sentences vs. nonwords, but not when they complete arithmetic, logic-puzzle, visuospatial working memory, and Stroop tasks (Fedorenko et al., 2011). Importantly, the functional selectivity of functional networks helps to circumvent the “reverse inference” problem (Poldrack, 2006), allowing scientists to make more compelling observations about the relationship between the localization of activation and the engagement of certain cognitive functions (Blank, Kanwisher & Fedorenko, 2014).

Following previous cross-methodological meta-analyses (e.g., Woolgar et al., 2018), the methodological approach employed in the present meta-analysis should benefit from general advantages associated with meta-analysis (e.g., increased statistical power due to larger participant size), while exploring potential alternatives to traditional group-based neuroimaging meta-analysis and generating unique interpretations about the nature of the cognitive mechanisms that support non-literal language processing. The use of functional networks in neuroimaging studies of language has become especially critical as the field of language research follows the example of other fields (e.g., visual processing, Kanwisher et al., 1997) in distancing itself from traditional group analysis methods. Ultimately, sophisticated understandings of scientific phenomena emerge when evidence from different methodological camps converge. In this experiment, we aim to contribute to methodological triangulation by presenting evidence generated from a cross-methodological meta-analysis technique.

Predictions

The purpose of the present study is to investigate the cognitive mechanisms involved in non-literal language processing by exploring three specific research questions. First, which functional networks, if any, are dominant (i.e., engaged more heavily) during non-literal language processing? Second, is the right hemisphere recruited more heavily during non-literal language processing? Finally, which non-literal language phenomena are most similar to each other based on their relationships to the functional networks of interest?

Previous theoretical and behavioral investigations have implicated the role of linguistic processing (e.g., Jung-Beeman, 2005), executive function (e.g., Gernsbacher & Robertson, 1999), and mentalizing (e.g., Wilson & Sperber, 2012) in non-literal language processing. In accordance with this work, many neuroimaging investigations report increased activation in traditional MD (e.g., Kristensen, Wang, Petersson, & Hagoort, 2013), language (e.g., Jin et al., 2009; Spotono, Koun, Prado, Van Der Henst, & Noveck, 2012; Matsui et al., 2016; Feng et al., 2017), and ToM (e.g., Mason & Just, 2009; Van Ackeren et al., 2012; Spotorno et al., 2012; Bašnáková et al., 2016) regions during non-literal vs. literal language processing. We therefore predict that a substantial number of activation peaks from non-literal vs. literal language contrasts will fall in each network of interest.

Though the role of the right hemisphere in non-literal language processing has often been invoked to explain differences in non-literal vs. literal language processing mechanisms (e.g., Giora, 1997; Beeman, 1993), serious methodological concerns undermine the strength of this argument (Cardillo et al., 2018; Kasparian, 2013). Additionally, previous neuroimaging experiments do not consistently report significant activation in the right hemisphere (e.g., Rapp et al., 2004; Citron & Goldberg, 2014). Thus, contrary to theories on the specialization of the right hemisphere in non-literal language processing, and consistent with previous neuroimaging meta-analyses of non-literal language (Bohrn et al., 2012; Rapp et al., 2012; Reyes-Aguilar et al., 2018), we expect to find that the majority of activation peaks are left-lateralized.

The present analysis is exploratory in nature. Its primary goal is to summarize existing neuroimaging studies on non-literal language processing by using a novel meta-analysis technique and, in doing so, to contribute to discussions on the cognitive mechanisms involved in non-literal language processing. Findings will be used to inform more detailed hypotheses in future experiments.

Table 1

Types of non-literal language investigated in the present analysis

Name	Example	Relevant Literature
Humor	"She went on a fourteen-day diet but she only lost two weeks"	Coulson, Urbach, & Kutas (2006), p. 242; Ruch (2008); Wyer & Collins (1992)
Idioms	"Kick the bucket"	Jackendoff (2002), p. 154
Indirect Speech	A: "What did you think of my. presentation? B: "It's hard to give a good presentation"	Holtgraves (1999), p. 521; Lee & Pinker (2010); Sperber & Wilson (1986)
Irony	Child repeatedly begs A for ice cream. A: "Are you sure you want ice cream?"	Attardo (2000), p. 818; Grice (1975); Gibbs (2000)
Metaphor	"You need to budget your time"	Lakoff & Johnson (1980), p. 456; Glucksberg, McGlone, Grodzinsky, & Amunts (2001)
Metonymy	"Room 23 called for more ice"	Piñango et al. (2017), p. 352
Sarcasm	"Do you have to make that noise while you are eating?"	McDonald (1999), p. 488
Prosody	"Tap the frog with the flower" vs. "Tap the frog (PAUSE) with the flower"	Snedeker & Trueswell (2003), p. 107; Hellbernd & Sammler (2016); Kjelgaard & Speer (1999)
Proverbs	"A rolling stone gathers no moss"	Gibbs, Johnsson, & Colston (2009), p. 234
Text/Discourse Comprehension	"Brad had no money but he just had to have the beautiful ruby ring for his wife. Seeing no salespeople around, he quietly made his way closer to the ring on the counter. He was seen running out the door"	Mason & Just (2009), p. 159; Johnson-Laird (1983); Graesser, Singer, & Trabasso (1994)

Method

Selection criteria

The literature search was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009). Studies were identified through Boolean searches in NeuroSynth, Google Scholar, PsycInfo, and PubMed databases using combinations of the keywords, “humor,” “jokes,” “comedy,” “idioms,” “indirect speech,” “indirect request,” “speech act,” “irony,” “ironic,” “lying,” “metaphor,” “metonymy,” “sarcasm,” “sarcastic,” “prosody,” “proverbs,” “text coherence,” “text comprehension,” “discourse comprehension,” “pun,” “paradox,” “hyperbole,” “oxymoron,” “understatement,” “figure of speech,” “figurative language,” “non-literal language,” “platitude,” “saying,” “personification,” “synecdoche,” “anaphora,” “anthropomorphism,” “pragmatics,” “narrative,” “fMRI,” “brain,” and “neuroimaging.” Only papers published after the year 2000 were considered. Additionally, we checked the reference lists of ten neuroimaging meta-analyses on non-literal language (Bohrn, Altmann, & Jacobs, 2012; Ferstl, Neumann, Bogler, & von Cramon, 2010; Lisofsky, Kazzer, Heekeren, & Prehn, 2014; Mason & Just, 2009; Rapp, Mutschler, & Erb, 2012; Reyes-Aguilar, Valles-Capetillo, & Giordano, 2018; Vartanian 2012; Vrticka, Blank, & Reiss, 2013; Yang 2014; Yang & Shu, 2016) for relevant articles.

200 articles were initially selected based on the content of their abstracts for full-text screening, after which 129 were excluded. The remaining 71 studies, conducted between the years 2001 and 2018, met the following criteria: (1) fMRI, not EEG or PET, was used, (2) participants were neurotypical adults, (3) participants were native speakers of the language in which the experiment was conducted, (4) peak activations were reported using Talairach

(Talairach & Tournoux, 1988) or Montreal Neurological Institute (MNI) coordinates, (5) data analysis was a standard random effects analysis (Holmes & Friston, 1998) performed at a whole brain level, and (6) contrasts of non-literal language processing against a literal language baseline were reported. One activation peak was excluded due to a perceived misprint in the published dataset (Perrone-Bertolotti et al., 2013). In total, 770 activation peaks (94 contrasts) from 1343 total participants aged 18 to 55 (average age = 24.55, range = 18-55; average number of participants per study = 19, range = 8-39, SD = 6.1) were included in the analysis. An overview of the literature is provided in Table 2 and Table 3.

Table 2

Included articles by non-literal language type

Type	Number of Studies	Number of Peaks
Humor	8	107
Idiom	4	33
Indirect Speech	9	152
Irony	10	68
Metaphor	20	209
Metonymy	2	10
Prosody	4	62
Proverb	1	9
Sarcasm	1	1
Text Coherence	12	119
TOTAL	71	770

Functional localization of the language, MD, and ToM networks

The topography of the language, MD, and ToM networks was characterized through whole-brain probabilistic maps and binary masks. Both representations were functionally defined using group-level representations of individual-level data from large samples of participants who performed localizer tasks (e.g., Saxe et al., 2006; Fedorenko et al., 2010; Nieto-Castañón & Fedorenko, 2012) designed to identify language, MD, and ToM regions, as described below. For each network, the whole-brain probabilistic map represented, for each voxel, the percentage of participants that showed significant activation for the relevant localizer contrast in that voxel at $p < .001$, uncorrected. The binary masks were derived by parcellating the probabilistic maps into large areas of cortex within each of which most participants (≥10%) showed significant activations using a segmentation algorithm (Meyer, 1991) as described in Fedorenko et al. (2010).

The language localizer task is available at http://web.mit.edu/evelina9/www/funcloc/funcloc_localizers.html and is discussed in detail in Fedorenko et al. (2010). Briefly, participants read sentences and sequences of pronounceable nonwords one word/nonword at a time in a standard blocked design with a counterbalanced order across runs. At the end of each trial, a button-pressing task was included to help participants remain alert. The sentences > non-words contrast identifies regions comprising a fronto-temporal network sensitive to high-level linguistic processes including, but not limited to, lexical and combinatorial processing of both form (syntax) and content (semantics) (e.g., Fedorenko et al., 2010, 2012, 2018; Blank et al., 2016). This localizer has been shown to generate similar patterns of neural activity regardless of materials, task, and modality of presentation (Fedorenko et al., 2010; Fedorenko, 2014). A whole-brain probabilistic activation map, as well as binary masks delineating the gross locations of activations in the sentences > non-words contrast, were derived based on individual activation maps from 220 participants. As in Blank et al. (2014), left hemisphere masks were mirror-projected onto the right hemisphere, creating right hemisphere masks. This step was performed because the language network is left-lateralized, such that 1) masks created based on weaker activations in the right hemisphere might be under-inclusive; and 2) averaging activations from one strongly and one weakly activated hemisphere might similarly distort the underlying topography of activation effects (note that only masks were mirror-projected; the whole-brain probabilistic map retained the hemispheric differences in activation patterns). Overall, twelve binary masks were defined, with six located in each hemisphere. These included six masks on the lateral surface of the frontal cortex in the inferior frontal gyrus (L/R IFG) and its orbital part (L/R IFGorb) as well as in the middle frontal gyrus (L/R MFG); and six fROIs on the lateral surface of the temporal and parietal cortex, in the anterior temporal cortex (L/R AntTemp), posterior temporal cortex (L/R PostTemp), and angular gyrus (L/R AngG). These masks are depicted in Figure 1 (Appendix A).

The MD network was defined based on a spatial working memory task in which neural activation from a hard version and an easy version was contrasted (Fedorenko et al., 2011). In the hard version of the task, participants kept track of eight randomly-generated locations that were sequentially flashed two at a time onto a 3 x 4 grid. In the easy version, participants kept track of only four locations (flashed one at a time). At the end of each trial, participants indicated their memory for the correct pattern in a two-alternative, forced-choice paradigm. They immediately received feedback about the accuracy of their answers. The hard and easy conditions were presented in a standard blocked design with a counterbalanced order across runs. The hard > easy contrast identifies a bilateral network of frontal, parietal, cingular and opercular regions that are causally related to fluid intelligence (Woolgar et al., 2018) and are recruited by processes related to working memory, cognitive control, and response inhibition (e.g., Duncan & Owen, 2000; Miller & Cohen, 2001; Braver et al., 2003; Cole & Schneider, 2007; Hugdahl et al., 2015). Indeed, Fedorenko, Duncan, and Kanwisher (2013) showed that the regions activated by this particular hard > easy contrast are also activated by contrasts between hard and easy conditions across a wide range of other tasks and stimuli. Data from 197 participants was used to generate the whole-brain probabilistic map and the binary masks comprising the MD network. To ensure that these group-level representations were symmetric across the two hemispheres (reflecting the bilateral nature of this network), individual activation maps were first averaged across hemispheres. Overall, eighteen masks were defined, nine in each hemisphere. These included the opercular part of the inferior frontal gyrus (L/R IFGop), the middle frontal gyrus (L/R MFG) and its orbital part (L/R MFGorb), the precentral gyrus (L/R PrecG), the insular cortex (L/R Insula), the supplementary motor area (L/R SMA), the inferior parietal cortex (L/R InfPar), the superior parietal cortex (L/R SupPar), and the anterior cingulate cortex (L/R AntCing). The MD masks are depicted in Figure 2 (Appendix A).

The ToM network was defined based on a task consisting of reading short paragraphs followed by a comprehension question, in which situations that required mental-state inference were contrasted with situations that did not. Specifically, activation associated with a “false belief” condition, in which a character held an outdated belief, was contrasted with activation during a “false photo” condition, in which an outdated representation was contained in an inanimate object (condition order was counter-balanced across runs). This contrast targets the brain regions involved in ToM processing (Dodell-Feder, Koster-Hale, Bedny, & Saxe, 2011; Jacoby, Bruneau, Koster-Hale, & Saxe, 2016; Saxe & Kanwisher, 2003). The whole-brain probabilistic map and binary masks were generated based on data from 119 participants. Given that the resulting activations were right-lateralized, masks were generated based on right-hemisphere data and were then mirror-projected onto the left hemisphere to create left hemisphere homologues. Overall, eighteen masks were defined for the ToM network, nine in each hemisphere. These included the temporo-parietal junction (L/R TPJ), the middle temporal gyrus (L/R MTG), the anterior middle temporal gyrus (L/R AntMTG), the posterior lateral prefrontal cortex (L/R PostLPFC), the triangular part of the inferior frontal gyrus (L/R IFG), the anterior lateral prefrontal cortex (L/R AntLPFC), the precuneus, the superior medial prefrontal cortex (L/R SupMPFC), and the inferior middle prefrontal cortex (L/R InfMPFC). These masks are shown in Figure 3 (Appendix A).

Meta-analysis

Before analysis, included activation peaks reported in Talairach space were converted to MNI space using a converter from the Effect Size Signed Differential Mapping (ES-SDM) (Radua et al., 2012) website (https://www.sdmproject.com/utilities/?show=Coordinates). The MNI coordinates of the 770 activation peaks selected for analysis were tested against both 1) the whole-brain probabilistic maps for each of the three functional networks of interest and 2) the corresponding binary masks broadly delineating the likely locations of functional activations in each network across the population. In particular, two values were generated for each coordinate: (1) a probability value showing the percentage of the 220 (language), 197 (MD), or 119 (ToM) participants from the functional localization sample who showed activation at that location; and (2) an integer which was either the individual label for the mask that the peak fell into, or zero if the peak did not fall into any of the relevant masks.

To test whether activations were more likely to fall in some network(s) or hemispheres than in others, and to compare these results to previous meta-analyses, we performed a series of two-tailed pairwise proportion tests. First, for each pair of networks (language vs. MD, language vs. ToM, and MD vs. ToM), we tested whether the two networks differed compared to the proportion of peaks falling within their masks (based on measure (2) above). Second, we tested whether the overall proportion of peaks falling within the right hemisphere differed from the proportions reported in previous meta-analyses. In focusing on the right hemisphere, we aimed to address past work suggesting a special role for the right hemisphere in non-literal language processing. Third, for each pair of networks, we tested whether they differed in terms of the distribution of peaks across the two hemispheres (quantified as the proportion of peaks within a network that fell in the right hemisphere). Significance was assessed based on the resulting z-scores. We conducted the proportion tests using both raw proportions (i.e., proportions derived by counting) and proportions adjusted according to the size of each networks' masks (number of voxels).

To test for pairwise associations in activation across networks, we used logistic, mixed-effects regression implemented via the lme4 package (Bates, Mächler, Bolker, & Walker, 2014) in R. For a given pair of networks A and B, we predicted whether peaks fell within vs. outside of network A based on them falling within vs. outside of network B (based on measure (2) above), with a random intercept for the grouping of peaks by study (additional grouping of peaks by topics such as "humor", "irony", etc., did not improve the model). Significance was assessed using ablative testing with ANOVA, i.e., testing whether the predictor of network B significantly improved prediction for network A. p-values were FDR-corrected for multiple comparisons.

Whereas the above test evaluates the association in activation between two networks against a theoretical null hypothesis of zero association, the group-level representations of the networks themselves show considerable spatial overlap and are thus not independent in practice (even though these networks exhibit robust functional and spatial dissociation at the individual participant level). For example, if language masks substantially overlapped with ToM masks, then even a collection of peaks sampled randomly across the cortex would have a subset of peaks falling within both language and ToM masks and might therefore appear to support a significant association between the two networks. To test whether the associations in activation across a pair of networks A and B were different than what would be expected by chance given the topography of network masks, we first computed a Chi-square statistic quantifying the extent to which the distribution of peaks within vs. outside the networks (based on measure (2) above) deviated from the expected distribution if the two networks were completely independent. Then, instead of comparing this statistic to the theoretical null distribution representing zero-association, we constructed an empirical null distribution via a permutation test: for each of 10,000 iterations, we randomly sampled 770 voxels (with replacement) from across the gray matter (as defined by the AAL atlas in SPM), counted how many voxels fell within vs. outside the masks of each of the two networks, and computed the resulting Chi-square value; the resulting sampling distribution represented the Chi-square values that would be expected based on the topography of the networks' masks alone, and the real Chi-square value was compared against it to obtain a *p*-value.

To evaluate the similarity across different non-literal language phenomena in terms of the functional networks they are associated with, we used hierarchical clustering. First, we represented each phenomenon as a vector with six entries, specifying the phenomenon's average probability (across peaks) of falling within each network (language, MD, ToM) in each hemisphere (left, right) (based on measure (1) above). Only probability values from peaks that fell in at least one network contributed to the averages, as the low probability values from non-network peaks would wash out the averages and thus minimize differences between non-literal language phenomena. Next, we computed the correlations across these vectors to capture similarities across the corresponding topics. Finally, these correlations were converted to distances (by subtracting them from 1) and submitted to Matlab's linkage function to produce a hierarchical tree (we used "average" linkage, but results were qualitatively similar with other methods). Data on sarcasm were excluded due to its limited contribution to the clustering analysis (one activation peak).

Results

Of the 770 total activation peaks included in the present analysis, 368 peaks (48%) fell in at least one network. Sub-categorizations of the peaks are depicted in Figures 4-7 (Appendix B). Below, we refer to these as the peaks "relevant" to our analyses.

Which networks are activated during non-literal language processing, and in which hemisphere?

First, we tested for each network separately whether it was associated with non-literal language processing more than what would be expected by chance. The proportion of relevant peaks falling specifically (but not necessarily exclusively) in the language network (182/368) was greater than the "baseline" proportion of "relevant voxels" (voxels falling in a mask of at least one network) (13,588/66,280) that fell specifically in the language network ($p < .00001$). Similarly, the proportion of relevant peaks falling in the ToM network (235/368) was greater than the corresponding "baseline" proportion of relevant voxels that fell in the ToM network (21,868/66,280) ($p < .00001$). However, the proportion of relevant peaks falling in the MD network (121/368) was smaller than the "baseline" proportion of relevant voxels that fell in the MD network (30,824/66,280) ($p < .00001$).

When we repeated the tests above only for the right-hemisphere peaks (139 relevant peaks), the same pattern was observed: the proportion of peaks falling specifically in the language network (64/139) was greater than the "baseline" proportion of right-hemisphere "relevant voxels" that fell specifically in the language network (6794/33,212) ($p < .00001$). A similar result obtained for the ToM peaks (91/139 vs. 11,006/33,212) ($p < .00001$). In contrast, the proportion of peaks falling specifically in the MD network (49/139) was smaller than the "baseline" proportion of right-hemisphere "relevant voxels" that fell specifically in the MD network (18,786/33,212) (p = .008513). The distribution of relevant peaks in the right hemisphere did not differ significantly across networks when the overall size of each network was *not* accounted for (ToM vs. language: p = 0.455788, ToM vs. MD: p = 0.745938, language vs. MD: p = 0.455788).

Next, we compared the networks to one another to test for differences in their level of association with non-literal language processing. Of the 368 relevant peaks, significantly more peaks fell in the ToM network than in either the MD (p < .00001) or the language (p < .000468) networks, followed by the peaks falling in the language network which in turn outnumbered those in the MD network (p < .00001). However, when the number of the 368 peaks in each network was divided by the overall size of each network (total number of voxels across all masks), significantly more peaks fell in the language network compared to either the MD (p < .00001) or ToM (p = 0.024576) networks, followed by the peaks falling in the ToM network which in turn outnumbered those in the MD network (p < .00001). When repeating the latter analysis for the right hemisphere only, the language vs. ToM networks did not differ significantly in their respective proportions of peaks (p = 0.421686), and both contained significantly more peaks than the MD network (both ps < .00001).

Across all 770 peaks, 39% (297) are located in the right hemisphere (see Figures 8-11 [Appendix B] for graphical depictions of results regarding the networks in which the right hemisphere peaks fell). This proportion is significantly larger than both the proportion of right hemisphere peaks reported in Binder et al. (2009)'s meta-analysis of literal language processing, in which 32% (362/1133) of peaks fell in the right hemisphere (p = .002887), and the proportion of right hemisphere peaks reported in Rapp et al. (2012)'s meta-analysis of non-literal language, in which 32% (129/409) of peaks fell in the right hemisphere (p = .016752).

Can recruitment of some networks for non-literal language processing predict the recruitment of other networks?

Our next analysis addressed whether falling within one network changed the likelihood of a peak to fall in another network. Of all relevant peaks (both right and left hemisphere), those that fell in the language network had an increased chance of falling in the ToM network (estimate based on mixed-effect logistic regression: OR = 18.2, $p < .00001$), and a decreased chance of falling in the MD network (OR = 0.47, $p < .05$). Peaks that fell in the MD network had a decreased chance of falling in the ToM network (OR = 0.5, $p < .05$). In the left hemisphere, peaks that fell in the language network had an increased chance of falling in the ToM network (OR = 18.55, $p < .00001$), and a decreased chance of falling in the MD network (OR = 0.41, $p < .05$). Peaks that fell in the MD network had a decreased chance of falling in the ToM network (OR = 0.37, $p < .05$). In the right hemisphere, peaks that fell in the language network had an increased chance of falling in the ToM network (OR = 23.32, $p < .00001$), but had neither a decreased nor an increased chance of falling in the MD network (OR = 0.53, $p = .51$). Peaks that fell in the MD network had neither a decreased nor an increased chance of falling in the ToM network (OR = 0.58, $p = .51$).

An additional analysis compared these dependencies across networks in terms of peak distribution to the overall across-network dependencies as captured by the spatial overlap between their respective masks. We found that the dependency between the language and ToM networks in the present dataset of peaks was significantly stronger than what would be expected based on spatial overlap alone; whereas this effect was not significant when both hemispheres were considered jointly ($\chi^2_{(1)} = 225.13$, $p = 0.07$), it was significant for each hemisphere separately (left hemisphere peaks only: $\chi^2_{(1)} = 315.70$, $p = 0$; right hemisphere peaks only: $\chi^2_{(1)} = 276.53$, $p = 0.0013$). In contrast, the dependency between the language and MD networks in the present dataset, combined across hemispheres, was descriptively weaker from the overall dependency as determined by spatial overlap, but this difference was not significant ($\chi^2_{(1)} = 7.31$, $p = 0.078$). No significant differences were found when each hemisphere was considered separately (left hemisphere only: $\chi^2_{(1)} = 0.002$, $p = 1$; right hemisphere only: $\chi^2_{(1)} = 1.67$, $p = 1$). Similarly, the dependency between the ToM and MD networks in the present dataset, combined across hemispheres, was descriptively weaker from, but not significantly different than, what would be expected based on spatial overlap alone ($\chi^2_{(1)} = 8.97$, $p = 0.057$). Here, too, no significant differences were found when each hemisphere was considered separately (left hemisphere-only: $\chi^2_{(1)} = 0.002$, $p = 1$; right hemisphere only: $\chi^2_{(1)} = 4.14$, $p = 0.31$).

Which subsets of non-literal language phenomena show similar patterns of network recruitment?

We assessed which “kinds” of non-literal language types, at least as classified by the topics of different studies, exhibited similar distributions of peaks across the three networks. Hierarchical clustering revealed similarities in peak distributions from studies involving proverbs, metaphors, and metonymy. This group differed greatly from two additional groupings: idioms, text coherence, and humor; and indirect speech, irony, and prosody. Differences between the latter two groupings did not appear as stark. The results from the hierarchical clustering analysis are shown in Figure 12 (Appendix B).

Discussion

The primary purpose of the present work was to examine the cognitive mechanisms involved in non-literal language processing. We approached this aim by performing a meta-analysis of existing fMRI studies on non-literal language. Meta-analysis serves an important tool in neuroimaging research in that it summarizes findings across research groups and thus benefits from increased statistical power (Costafreda, 2009). The analysis technique employed in the present study involved comparing the locations of activation peaks from past studies on non-literal language processing with the locations of three functional networks. The language, MD, and ToM networks become activated when individuals perform cognitive operations that appear to contribute to non-literal language processing. Adhering to stringent inclusion criteria, we ultimately included activation peak data from 71 studies (94 contrasts) conducted on 10 different types of non-literal language in our analysis (Table 2). Table 3 contains a complete listing of included studies.

In our investigation of the cognitive mechanisms involved in non-literal language processing, we addressed three main research questions. First, we asked whether the included peaks fell more heavily in any one network (language, MD, or ToM). Next, we examined the extent to which the included peaks fell more heavily in the left or right hemisphere. Finally, we explored how individual non-literal language types could be grouped together based on their functional profiles. Following previous work on the importance of linguistic processing, executive function, and mentalizing in non-literal language processing (e.g., Jung-Beeman, 2005; Gernsbacher & Robertson, 1999; Wilson & Sperber, 2012), we predicted that a substantial number of activation peaks from existing fMRI studies on non-literal language would fall in each of the language, MD, and ToM networks. In response to methodological concerns regarding past research on non-literal language processing and the right hemisphere (Cardillo et al., 2018; Kasparian, 2013), we expected to observe a similar number of right vs. left hemisphere peaks as reported in previous meta-analyses of literal language. The final analysis was exploratory.

Findings related to the question of network dominance indicate that included activation peaks fell in the language and ToM networks at above-chance levels. Although the most activation peaks fell in the ToM network by numbers, the most peaks fell in the language network when controlling for network size. In addition, peaks that fell in the language network had an increased chance of also falling in the ToM network, and this dependency between the language and ToM networks was greater than what was expected based on the topography of the two networks. Contrary to expectation, the MD network appeared to play a diminished role during non-literal vs. literal language processing. Regarding the hemispheric lateralization question, we observed more right hemisphere peaks in the present analysis than in past meta-analyses on both literal and non-literal language. This result appears consistent with past theories that assert that the right hemisphere plays a special role during non-literal vs. literal language processing. Finally, the hierarchical clustering analysis revealed three main groups of non-literal language phenomena, suggesting that the types of non-literal language included in the present analysis represent separable kinds.

Limitations

Several limitations of the present experiment should be acknowledged. First, it is important to consider problems related to individual fMRI experiments that could become magnified when data is pooled from multiple studies. One such problem is bias. Carp (2012) described that both experimenters and publishers contribute to bias in fMRI research. Publication bias, or the tendency to only publish positive results, and method selection bias, or the tendency to exclusively include analyses that produce positive results, warp the empirical landscape. Though bias is a concern across different types of scientific work, fMRI research suffers particularly from method selection bias, which results in large numbers of false positives. For example, Carp (2012) found that data collection and analysis methods varied widely across 241 fMRI experiments conducted between 2007 and 2012. In fact, nearly every selected study used different combinations of analysis pipelines and significance thresholds (Carp, 2012). A lack of methodological standardization may help explain why an estimated 10-20% of significant activations reported in published fMRI studies are false positives (Wager et al., 2007). The open science movement (Simmons, Nelson, & Simonsohn, 2011; Ioannidis, Munafo, Fusar-Poli, Nosek, & David, 2014; Poldrack et al., 2017), which recently emerged in response to the replication crisis (Ioannidis, 2005), aims to reduce the number of false positives reported in scientific work by encouraging methodological transparency. Unfortunately, the problem of bias in fMRI research is not addressed directly in the present analysis. It would be interesting to conduct further analyses to explore possible differences in results when grouping studies by significance threshold and/or analysis pipeline type.

There are several other ways in which the present experiment could be modified to respect categorical differences among the included studies. In the current study, we did not separate studies that use different presentation modalities (auditory vs. visual), different tasks, conventional vs. novel stimuli, or stimuli in different languages (Table 3). The distinction between conventional and novel stimuli is especially pertinent, as past theories make strong predictions about the recruitment of the right hemisphere during novel or unusual non-literal stimuli comprehension (Giora, 1997). In addition, task type (valence judgment vs. semantic judgment) was shown to modulate activation in the right hemisphere in a non-literal language comprehension study (Yang et al., 2009). Previous meta-analyses of non-literal language have similarly found slight differences in activation when comparing activation peak data from studies that use different presentation modalities, conventional vs. novel stimuli, and stimuli in different languages (Bohrn et al., 2012; Rapp et al., 2012; Reyes-Aguilar et al., 2018). Importantly, it is unclear whether the observed differences truly reflect the difference of interest or whether they instead point to the presence of other variables between the groups. In particular, modality-specific (i.e., auditory vs. visual) differences seem unlikely, as activations related to basic reading or listening mechanisms should cancel out in the non-literal vs. literal language contrasts. Regardless, since we did not repeat the present analyses while accounting for differences in stimuli and task, our results essentially treat the 71 included studies as a homogenous group. Indeed, it is possible that finer-grained differences in stimuli presentation modality, language, salience (i.e., conventionality/frequency), and in overall task partially confound the results observed in the present analysis. I will return to the issue of differences between non-literal language phenomena later in the discussion.

Another potential weakness of the present experiment concerns the use of activation peak data in neuroimaging meta-analysis. As discussed in the last section of the introduction ("Cutting to the chase"), researchers typically publish activation peak data only, which can be problematic for voxel-based meta-analysis algorithms that must consequently treat all non-peak voxels as possessing an effect size of zero (Costafreda, 2009). Though we did not use traditional meta-analysis techniques such ALE and ES-SDM in the present study, other concerns related to the use of activation peak data persist. First, activation peaks represent the most active voxels above a given significance threshold, and therefore provide a less detailed picture of the location and intensity of activation in all voxels than that provided by group activation maps. Further, the use of different significance thresholds may affect the number of voxels considered “active,” which could explain why some studies included in the present analysis report 42 peaks for a non-literal vs. literal language processing contrast (Bašnáková et al., 2015) while others report just 1 (Ahrens et al., 2007). Given many sources of variation across fMRI experiments, however, activation peak data, which can be viewed as a more conservative depiction of findings as compared to group activation maps (i.e., activation peaks reflect only the most active voxels), seems highly appropriate for the purposes of the present analysis. In any case, most researchers do not publish group activation maps, which would seriously limit the number of studies included in the present meta-analysis if group activation maps were preferred.

Finally, general criticisms about the use of functional networks in fMRI research deserve consideration. Friston, Rotshtein, Geng, Sterzer, & Henson (2006) argued that focusing on a limited number of functional networks in fMRI analysis could cause scientists to lose sight of broader relationships between regions in the brain. In other words, analysis involving functional networks may “miss the forest for a tree” (Saxe et al., 2006, p. 1092). In fact, advocates for the use of functional networks in fMRI research have repeatedly emphasized that functional network-based analysis should not replace traditional group analysis (Saxe et al., 2006; Nieto-Castañón & Fedorenko, 2012). Instead, analysis that incorporates information about functional networks serves as an additional methodological tool in understanding the neural underpinnings of cognition. Following the publication of multiple traditional group meta-analyses on non-literal language (e.g., Bohrn et al., 2012; Rapp et al., 2012; Reyes-Aguilar et al., 2018), the present meta-analysis thus offers a different perspective to the discussion on non-literal language processing and the brain.

Question 1: Network dominance

The language network

In examining the networks into which included activation peaks fell, we aimed to identify neural differences between non-literal and literal language processing that might point to the recruitment of additional or different cognitive resources during non-literal language processing. Based on our findings, the language network appears to serve the most dominant role in non-literal language processing. Not only did a statistically significant number of activation peaks fall in the language network, but more activation peaks fell in the language network than in any other network when controlling for network size.

The engagement of the language network during non-literal vs. literal language processing may initially seem somewhat counterintuitive. One would expect to see activation in language regions during non-literal language processing alone, as non-literal language is a form of language. Yet, when elements of shared linguistic processing are subtracted away in a non-literal vs. literal language contrast, the presence of remaining activation in language regions raises an interesting question. What aspects of high-level linguistic processing are involved in non-literal language processing beyond what is required for literal language processing?

Previous theories have emphasized the role of executive function (e.g., Gernsbacher & Robertson, 1999) and mentalizing (e.g., Wilson & Sperber, 2012) to argue for differences in the cognitive mechanisms responsible for non-literal vs. literal language processing. The cognitively demanding nature of non-literal language processing, or the increased need to model others' thoughts during non-literal language exchanges, lend themselves to fairly intuitive interpretations of how non-literal differs from literal language. According to these interpretations, non-literal and literal language processing draw upon similar linguistic resources, but additional contributions from other cognitive domains such as mentalizing and executive function distinguish non-literal from literal language processing. This line of reasoning is in fact consistent with serial processing theories, such as the standard pragmatic model (Grice, 1975; Searle, 1979), in which literal linguistic processing serves merely as the basis of non-literal language processing.

Perhaps the highly intuitive nature of serial processing theories—despite their potential flaws—explains why hypotheses about additional linguistic machinery required during non-literal language processing are not as prevalent in the literature. Even Jung-Beeman (2005), who focuses on linguistic processes (as opposed to mentalizing or executive function-related processes) when discussing differences in non-literal and literal language processing, talks about these differences in terms of additional "demands" on the semantic integration system (p. 515). A burdened semantic integration system would seem to engage executive functions such as attention and working memory. Many of the experiments included in the present meta-analysis describe observed activation in the left inferior frontal gyrus (Broca's area) in non-literal vs. literal language contrasts in a similar manner. For instance, Stingaris et al. (2007) posited that increased activation in the left IFG during metaphor vs. literal sentence comprehension reflects more controlled, or conscious, retrieval from semantic stores during ambiguity resolution. In their study on discourse coherence, Shibata et al. (2011) describe that the detection of true contextual violations (which ceases once participants realize that they are reading an incoherent conversation) could explain activation in language areas during coherent vs. incoherent sentences conditions. Other interpretations refer to the increased need to integrate inferences into an overarching mental model of a text or conversation (Jin et al., 2009), or to activate varying non-literal interpretations of a particular sentence or utterance (Jang et al., 2013).

Like Jung-Beeman's (2005) argument about semantic integration in non-literal language processing, these explanations can all be understood in terms of increased demands on domain-general executive function mechanisms. Specifically, increased difficulty in the retrieval, integration, and selection of semantic interpretations would seem to place increased demands on cognitive mechanisms involved in inhibition, attention, and working memory. Yet, contrary to expectation, our data show that fewer activation peaks fell into the MD network than what was expected by chance. This is especially intriguing considering the dominance of the language network as compared to the MD and ToM networks. Taken together, these findings may reflect the existence of language-specific executive processing mechanisms.

Previous empirical work has indeed demonstrated that language regions appear sensitive to cognitively demanding linguistic tasks. For instance, Thompson-Schill, D'Esposito, and Kan (1999) found that activity in the left IFG was modulated by the availability of competing semantic knowledge during a word generation task. Specifically, more activation was observed in the left IFG when competing semantic knowledge was high. More recent research has directly compared activation in MD and language regions (as defined using Fedorenko et al.'s (2010) GcSS fROI technique) during language tasks. Blank and Fedorenko (2017), for instance, found that the MD network did not track linguistic input when participants listened to naturalistic stories. The authors hypothesize that activation in MD regions during linguistic processing may arise only in typical fMRI experiments in which participants develop basic attentional strategies to complete non-naturalistic tasks. In other words, the MD network does not become engaged when domain-specific executive function mechanisms are available (i.e., during naturalistic language comprehension).

The notion of language-specific executive function mechanisms remains compatible with past theories about the role of executive function in non-literal language processing. Gernsbacher and Robertson (1999), for instance, emphasized the role of suppression in non-literal language interpretation selection. If language-specific executive function mechanisms exist, then the type of suppression needed to inhibit inappropriate interpretations of a sarcastic remark may draw upon different neural resources than the type of suppression needed to inhibit memory of previously viewed shapes during a memory task. Further, documented correlations between domain-general executive function and non-literal language processing abilities (e.g., McDonald & Pearce, 1996; McDonald & Pearce, 1998; Channon & Watts, 2003) could in fact point to a more specific correlation between domain-general and language-specific executive function mechanisms.

Additionally, the language network may become engaged during non-literal language processing due to factors beyond executive function. One alternative explanation for the present study's findings involves the notion of diffuse activation. Beeman et al. (1994) argued that the right hemisphere activates semantic fields spread across the cortex when handling distantly-related interpretations of a non-literal utterance. When two conditions are contrasted in traditional fMRI group analysis, the condition that generated more diffuse activation may result in the emergence of more significantly active peaks. Seeing as possible interpretations of non-literal language often vary widely, non-literal language conditions might generate especially high amounts of diffuse activation in language regions, which could explain the prevalence of language network peaks observed in the present study. An effective way to test this possibility would be to first compare activation when participants interpret distantly vs. closely related pairs of literal concepts before examining differences during the interpretation of non-literal vs. literal language.

The ToM network

As in the case of the language network, activation peaks fell in the ToM network at an above-chance level when the results were adjusted for network size. These results are consistent both with previous arguments concerning the need to model others' mental states during non-literal language comprehension and production (e.g., Sperber & Wilson, 1986; Wilson & Sperber, 2012), and with evidence for associations between ToM deficits and impaired non-literal language processing ability (e.g., Martin & McDonald, 2005; Winner et al., 1998; Gavilan & Garcia-Albea, 2011). An unpublished study from our lab (Kline, Gallee, Balewski, & Fedorenko) found that reading jokes over non-jokes elicited significant responses in the ToM network, and these responses were modulated by the degree of stimuli "funniness" as rated by participants. Importantly, all three functional networks explored in the present analysis (language, MD, and ToM) were defined individually in the same group of Kline et al. (submitted manuscript) participants who subsequently completed the jokes task, and only the ToM network was activated in the jokes vs. non-jokes contrast. Kline et al.'s use of the GcSS technique (Fedorenko et al., 2010) strengthens their claims about the involvement of the ToM network in joke comprehension. Future work should continue to employ this type of analysis to further explore the role of the ToM network during non-literal language processing.

An additional finding regarding the ToM network was that peaks in the language network had an increased likelihood of also falling in the ToM network, and the dependency between the language and ToM networks exceeded what was predicted by chance. This finding may indicate that a specialized subset of shared voxels between the language and ToM networks is recruited during non-literal language processing. In accordance with the finding that the dependency between ToM and language networks exceeded chance-level rates, an experiment included in the present meta-analysis reported results from a psycho-physiological interactions (PPI) analysis that showed that functional connectivity between areas in the ToM and language network increased during irony vs. literal sentence comprehension (Spotorno et al., 2012). Whether or not the particular regions in which increased functional connectivity was observed in Spotorno et al. (2012) indeed correspond with the location of peaks from the present analysis that fell in both the ToM and language networks remains an open question. In any case, consistent correlations between ToM and language development in children (e.g., Lockl & Schneider, 2006; Antonietti, Sempio, & Marchetti, 2006), and recent evidence that the ToM and language network are synchronized during both language comprehension and rest (Paunov, Blank, & Fedorenko, 2019) suggest that the connection between ToM and language regions merits further investigation.

Limitations

First, it is potentially concerning that fewer than half (47%) of the included peaks fell into at least one network. fMRI experiments reporting activation peak data often include peaks located in the midbrain, which may explain why such a large number of peaks did not fall in any of the functional networks of interest. In future iterations of the present experiment, it may be advisable to exclude activations from subcortical regions.

In addition, the analysis used in the present study focused only on whether a given activation peak fell in the functional networks of interest. We did not consider information about the exact locations of that peak within the networks. When comparing the relative number of peaks that fell in each network, we therefore cannot speak to questions about the proximity of these peaks to each other or about their specific location in the binary masks that comprise the networks (12 individual masks in language, 18 in MD, 18 in ToM). Despite that more peaks fell in the language network when controlling for network size, it remains possible that these peaks are dispersed throughout the language network, and that ToM peaks cluster more in a particular region. The latter result would perhaps constitute a more meaningful finding, as concentrated clustering within a small subset of a particular network—regardless of the size of the entire network—could point to the existence of a specialized set of neural resources within that network. In order to investigate this possibility, future research should focus on comparing individual binary mask sizes to the number of peaks that fell in the masks.

Despite that specific information about peak location was not incorporated in the present analysis, it is clear that the ToM and language networks emerged more strongly than the MD network during non-literal vs. literal language processing. Mentalizing and linguistic processing, then, seem to serve important and perhaps specialized roles in non-literal language processing.

Question 2: Lateralization

Given the historical significance of the lateralization debate and its potential to reveal differences in the mechanisms underlying non-literal vs. literal language processing, we reported the hemispheres into included peaks fell. Close to half (42%) of all included peaks were located in the right hemisphere. This proportion is significantly higher than that of right hemisphere peaks analyzed in previous meta-analyses of literal (Binder et al., 2009) and non-literal (Rapp et al., 2012) language. To claim that these results provide direct support for theories that emphasize the role of the right hemisphere in non-literal language processing (e.g., Winner & Gardner, 1977; Joanette et al., 1990; Jung-Beeman, 2005), however, is potentially problematic. Most obviously, the right hemisphere does not seem to play a *dominant* role in non-literal language processing, as over half of the activation peaks included in the present analysis were located in the left hemisphere. Instead, since the proportion of right hemisphere peaks in the present analysis exceeds the proportion of right hemisphere peaks in a previous meta-analysis of literal language, the right hemisphere appears to *contribute to* non-literal language processing.

Regardless of the exact nature of the relationship between non-literal language processing and the right hemisphere, referring to the contributions of the right hemisphere as a whole seems misguided. The origins of this practice can be traced back to methodological inconsistencies in the neuropsychological work that inspired the focus on the role of the right hemisphere in non-literal language processing. These studies rarely indicated the exact location of right hemisphere damage in patients and often aggregated data across individuals with highly variant right hemisphere lesions (Cardillo et al., 2018; Kasparian, 2013). Such practices limited the specificity with which researchers could describe contributions of the right hemisphere to non-literal language processing, which is unfortunate considering the anatomically and functionally diverse nature of the right hemisphere (Galaburda, Rosen, & Sherman, 1990; Rypma & D'Esposito, 1999; Lindell, 2006). Recent evidence has indeed suggested that the right hemisphere contains selective components of the language, MD, and ToM networks (Fedorenko et al., 2010; Duncan, 2010; Saxe, 2006).

With regard to our data, less than half of the total right hemisphere peaks (47%) fell into at least one functional network. Unfortunately, the analysis used in the present experiment did reveal the locations of non-network right hemisphere peaks. Are these peaks spread across the right hemisphere, or are they concentrated in specific regions? Do they border the functional networks of interest, or are they located very distantly from the networks? The inability to answer such questions complicates interpretations of the relatively high number of total right hemisphere peaks reported in the present analysis. Further, among the peaks that fell into at least one network, a higher proportion of peaks fell in the right hemisphere ToM and language regions than what was expected based on the size of each networks' right hemisphere regions, while a lower proportion of peaks fell in right hemisphere MD regions. The relative strength of the ToM and language networks and weakness of the MD network in the right hemisphere seriously challenge broad portrayals of the right hemisphere's contribution to non-literal language processing. As such, investigations aimed at uncovering the cognitive mechanisms that support non-literal language processing are unlikely to yield fruitful conclusions if they focus only on the role of the right hemisphere as a whole. Future research should therefore continue to explore the degree to which subcomponents of functional networks located in the right hemisphere become engaged during non-literal language processing.

If regions within the right hemisphere indeed play a special role in non-literal vs. literal language processing, then it is important to consider the ways in which the right hemisphere's contributions might differ for different types of non-literal language. The fact that the proportion of right hemisphere peaks in the present study was higher than the proportion of right hemisphere peaks in another meta-analysis of non-literal language processing (Rapp et al., 2012) introduces an intriguing possibility. Perhaps the inclusion of data pertaining to different types of non-literal language skewed the observed number of right hemisphere peaks. Rapp et al. (2012) specifically analyzed data from metaphor, irony, and idiom studies, while the present experiment included data from studies involving seven additional types of non-literal language (Table 1). It would be interesting to examine whether certain types of non-literal language modulate activation in right hemisphere regions to a greater extent. Other potential differences among non-literal language phenomena are discussed in the following section.

Question 3: Groupings of non-literal language phenomena

Our final research question concerned how different non-language phenomena could be grouped based on their network profiles. The existence of separable groupings may reflect the involvement of different combinations of cognitive mechanisms among non-literal language phenomena. The clustering analysis compared information about mean probability values associated with peaks corresponding to each non-literal language phenomenon (e.g., the mean probabilities of peaks in right hemisphere ToM regions for metaphor vs. irony vs. text coherence studies). Probability values can be understood as indicators of generalizability. Observing a high probability value of a peak in the language network, for example, allows us to be more confident that that peak would fall into the language network of any member of the population (i.e., neurotypical adults). Notably, previous meta-analyses of non-literal language processing have reported very few reliable differences in activation between different non-literal language types (Reyes-Aguilar et al., 2018). Seeing as several clear groupings emerged in the present analysis, the hierarchical clustering technique seems to benefit from increased sensitivity to differences in activation across non-literal language types and thus serves as a more effective method of identifying differences in neural signatures between non-literal language types.

The grouping analysis revealed three main groups of non-literal language phenomena: proverbs/metaphors/metonymy, idioms/text coherence/humor, and indirect speech/irony/prosody. The largest differences between the groups were observed when comparing the first group to the second two. The second two groups (idioms/text coherence/humor, indirect speech/irony/prosody) did not differ as starkly. As a result, arguments about the distinctness of proverbs/metaphors/metonymy vs. the rest of the groups likely carry more weight than arguments about the distinctness between idioms/text coherence/humor and indirect speech/irony/prosody.

Before each of the groups is discussed individually, it should be noted that the data used in the present experiment contained high numbers of peaks from humor, indirect speech, metaphor, and text coherence studies. The number of peaks included in the clustering analysis for each non-literal language type is depicted in parentheses in Figure 10. Due to inequalities in the number of studies and peaks that represent each non-literal language phenomenon, network information on phenomena associated with higher numbers of studies and peaks may be more informative. That is, the presence of more "participants" (i.e., peaks/studies) for humor, indirect speech, metaphor, and text coherence likely allows for a more accurate representation of these non-literal language types' functional profiles. Conversely, clustering results for non-literal language phenomena such as proverbs, for which only five peaks from a single experiment were included in the clustering analysis, may not be as meaningful.

Proverbs/metaphors/metonymy

Proverbs, metaphors, and metonymy share several important qualities. With regard to similarities between proverbs and metaphors, Gibbs and Beitel (1995) described proverbs as "higher-order metaphorical conceptualizations of everyday experience" (p. 133). The proverb, "don't put all your eggs in one basket," for instance, unifies the metaphors, "life is a container" and "beliefs are physical possessions." In her discussion on the distinction between irony and metaphor, Winner (1988) argues that relating concepts such as "life" and "container" serves as a distinguishing feature of metaphor comprehension. Metonymy comprehension similarly relies on a person's ability to link some attribute of a concept (the name of the hotel room in which people are staying) to that concept (the people staying in the room), such as in the example, "Room 23 called for more ice" (Piñango et al., 2017, p. 352). In contrast, other types of non-literal language, including irony and indirect speech, are more sensitive to the social context in which they are expressed than to conceptual information they encode (Winner, 1988). Seeing as all three types of non-literal language included in this grouping—proverbs, metaphors, and metonymy—constitute *conceptual linking* forms of non-literal language, the observed similarity in their network probabilities is not surprising.

The grouping of proverbs, metaphors, and metonymy alone, however, is not entirely intuitive. In particular, researchers of non-literal language have typically referred to proverbs and idioms as members of the same class, as both contain idiosyncratic form-to-meaning mappings. As a result of this idiosyncrasy, proverbs and idioms must remain somewhat fossilized in long-term memory (Jackendoff, 2002). That is, in contrast with typical combinations of verb and noun phrases, idioms such as, "cross your fingers" and proverbs such as, "a rolling stone gathers no moss" cannot be freely assembled by one individual and immediately understood by all. Given the inflexibility of both idiom and proverb development and their resulting need for long-term memory storage, it is interesting that idioms and proverbs emerged in separate groupings in the present experiment. One explanation may relate to the fact that proverbs usually comprise complete sentences, while some idioms can be modified lexically and syntactically depending on context (e.g., "She crossed her fingers" vs. "They cross their fingers"). Perhaps activation peaks from studies that use less alterable idioms as stimuli (e.g., "by and large") would more closely resemble those from studies of proverbs. This is a topic for future investigation.

Idioms/text coherence/humor and indirect speech/irony/prosody

In accordance with Winner's (1988) taxonomy of non-literal language types, text/discourse coherence, humor, indirect speech, and irony can all be considered more context-than concept-dependent types of non-literal language. As opposed to metaphor processing, successful interpretation of these types of non-literal language relies heavily on the comprehender's ability to assess social circumstances and model others' mental states (Lee & Pinker, 2010; Sperber & Wilson, 2002; Wyer & Collins, 1992). This notion is supported by evidence from fMRI research. Previous investigations on the involvement of the ToM network (defined individually in the same participants who subsequently completed a target task) during text (Jacoby & Fedorenko, 2018) and humor (Kline et al., submitted manuscript) comprehension have reported increased activation in the ToM network in non-literal vs. literal language conditions. Jacoby & Fedorenko (2018) specifically found that reading texts devoid of references to characters' mental states (e.g., a factual description of a type of fruit) engaged the ToM network. This finding is especially striking considering Jacoby & Fedorenko's (2018) use of stimuli that did not directly target mentalizing, and therefore suggests that the ToM network may become engaged as individuals attempt to build coherence when encountering all kinds of texts and discourse.

The importance of ToM in humor and text comprehension is mirrored by the mean probability values in the right hemisphere ToM network associated with these non-literal language phenomena. Humor was associated with the highest average right hemisphere ToM network value, 0.35, and text coherence had the second highest value, 0.27. Crucially, the right hemisphere contains a classic ToM region called the right temporo-parietal junction (TPJ) (Saxe, 2006). Though we did not run tests to evaluate statistical differences between these two mean probabilities and the rest of the mean probability values, it seems likely that the involvement of right hemisphere ToM regions in text and humor comprehension may have contributed to the linking of the two phenomena in the hierarchical clustering analysis. Further research is needed to establish the degree to which text coherence and humor processing bear a special relationship to the ToM network.

Factors outside of ToM may also explain the emergence of the indirect speech/irony/prosody and idioms/text coherence/humor clusters. In fact, the grouping of indirect speech/irony/prosody cluster, though still fairly similar to idioms/text coherence/humor, is probably the most intuitive of all three groupings. Changes in prosody greatly influence our interpretation of indirect and ironic statements (Wichmann, 2004; Bryant, 2005; Bryant & Fox Tree, 2009; Culpeper, Boulsfield, & Wichmann, 2003; Pell, 2006). Bryant (2005), for instance, found that ironic utterances were spoken more slowly in naturalistic contexts. Pell (2006) observed that individuals with brain damage struggled to gauge the confidence and politeness of speakers as a function the prosody the speakers used while expressing indirect statements. More generally, we have all probably experienced confusion regarding the ironic quality of an utterance when it is spoken without the "right amount" of ironic intonation. Given the apparent connection between prosody and irony/indirect speech, it would be interesting to explore whether irony and indirect speech still cluster with prosody when data from visual-modality studies of irony and indirect speech (i.e., studies in which participants read rather than hear stimuli) are exclusively incorporated in the clustering analysis.

Conclusion

The purpose of the present study was to investigate the cognitive mechanisms involved in non-literal vs. literal language processing. Previous work treated differences in time-course and hemispheric involvement as indicators of the recruitment of different cognitive resources in non-literal language processing. Due to low temporal resolution in fMRI, results from the present meta-analysis cannot speak to questions about variation in processing time between non-literal and literal language processing. Regarding hemispheric differences in non-literal vs. literal language processing, our dataset contained more right hemisphere peaks than those used in previous studies of literal and non-literal language. While this finding is consistent with theories that emphasize the special role of the right hemisphere in non-literal language processing, broad depictions of the right hemisphere's contributions to non-literal language processing do not respect its functional and anatomical diversity.

In order to disentangle the specific cognitive mechanisms recruited during non-literal language processing, we compared the locations of included activation peaks from non-literal vs. literal contrasts with the locations of selective and robustly active language, MD, and ToM networks. We selected these networks due to prior evidence indicating the involvement of linguistic processing, executive function, and mentalizing in non-literal language processing. Our most important findings ultimately emerged from the functional network-based analyses. First, more peaks fell in the language and ToM networks than the number predicted by chance, suggesting that aspects of linguistic processing and mentalizing may be recruited in non-literal over literal language processing. In particular, the dependency between the ToM and language networks raises the possibility that a specialized subset of voxels shared between these networks—perhaps representing an integrated social-linguistic processing mechanism (Wilson & Sperber, 2012)—becomes engaged during non-literal language processing. In addition, the observation that a low number of peaks fell in the MD network and a high number of peaks fell in the language network may point to the involvement of language-specific executive function mechanisms in non-literal language processing. Finally, when comparing the network profiles of different non-literal language types, we identified three groupings, suggesting that different cognitive mechanisms may be recruited to varying degrees among different non-literal language phenomena. For instance, mentalizing may be engaged more during indirect speech than metaphor comprehension. In sum, our findings are consistent with the claim that multiple cognitive mechanisms are differentially engaged in non-literal vs. literal language processing.

Future work should continue to include functional network-based analyses in order to explore neural evidence for the recruitment of specialized cognitive mechanisms in non-literal language processing. One proposal would be to repeat our analyses using activation peaks from literal vs. non-literal language processing contrasts and compare the results to our findings. Though only approximately 20 such contrasts were reported in included experiments (cf. the 94 included non-literal vs. literal contrasts), analyzing the literal vs. non-literal activation peaks may help reveal whether literal language processing also draws upon a specialized set of cognitive resources. Evidence for the existence of cognitive mechanisms used exclusively in literal language processing would challenge the traditional idea that literal language processing serves merely as the basis of non-literal language processing (Grice, 1975; Searle, 1979). Beyond meta-analysis, future fMRI experiments on non-literal language should incorporate information about individually-defined functional networks in analysis as in Jacoby & Fedorenko (2018). Comparing activation from a target task and from functional network localizer tasks in the same set of participants will strengthen claims regarding the involvement of the language, MD, and ToM networks in non-literal language processing.

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., … 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

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

Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H. B. M., & Zilles, K. (1999). Broca's region revisited: Cytoarchitecture and intersubject variability. *Journal of Comparative Neurology*, *412*(2), 319–341. https://doi.org/10.1002/(SICI)1096-9861(19990920)412:2<319::AID-CNE10>3.0.CO;2-7

Antonietti, A., Sempio, O. L., & Marchetti, A. (2006a). *Theory of Mind and Language in Developmental Contexts*. Springer Science & Business Media.

Apperly, I. A., Samson, D., Carroll, N., Hussain, S., & Humphreys, G. (2006). Intact first- and second-order false belief reasoning in a patient with severely impaired grammar. *Social Neuroscience*, *1*(3–4), 334–348. https://doi.org/10.1080/17470910601038693

Attardo, S. (2000). Irony as relevant inappropriateness. *Journal of Pragmatics*, *32*(6), 793–826. https://doi.org/10.1016/S0378-2166(99)00070-3

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *ArXiv E-Prints*, *1406*, arXiv:1406.5823.

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

Bechtel, W. (2002). Decomposing the Mind-Brain: A Long-Term Pursuit. *Brain and Mind*, *3*(2), 229–242. https://doi.org/10.1023/A:1019980423053

Beeman, M. (1993). Semantic Processing in the Right Hemisphere May Contribute to Drawing Inferences from Discourse. *Brain and Language*, *44*(1), 80–120. https://doi.org/10.1006/brln.1993.1006

Beeman, M. J., & Chiarello, C. (1998). Complementary Right- and Left-Hemisphere Language Comprehension. *Current Directions in Psychological Science*, *7*(1), 2–8. Retrieved from JSTOR.

Beeman, Mark, Friedman, R. B., Grafman, J., Perez, E., Diamond, S., & Lindsay, M. B. (1994). Summation Priming and Coarse Semantic Coding in the Right Hemisphere. *Journal of Cognitive Neuroscience*, *6*(1), 26–45. https://doi.org/10.1162/jocn.1994.6.1.26

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

Benedek, M., Beaty, R., Jauk, E., Koschutnig, K., Fink, A., Silvia, P. J., … Neubauer, A. C. (2014). Creating metaphors: The neural basis of figurative language production. *NeuroImage*, *90*, 99–106. https://doi.org/10.1016/j.neuroimage.2013.12.046

Bihrle, A. M., Brownell, H. H., Powelson, J. A., & Gardner, H. (1986). Comprehension of humorous and nonhumorous materials by left and right brain-damaged patients. *Brain and Cognition*, *5*(4), 399–411. https://doi.org/10.1016/0278-2626(86)90042-4

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cerebral Cortex (New York, NY)*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human Brain Language Areas Identified by Functional Magnetic Resonance Imaging. *Journal of Neuroscience*, *17*(1), 353–362. https://doi.org/10.1523/JNEUROSCI.17-01-00353.1997

Blank, I. A., & Fedorenko, E. (2017). Domain-General Brain Regions Do Not Track Linguistic Input as Closely as Language-Selective Regions. *Journal of Neuroscience*, *37*(41), 9999–10011. https://doi.org/10.1523/JNEUROSCI.3642-16.2017

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

Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., … Frith, D. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language A positron emission tomography activation study. *Brain*, *117*(6), 1241–1253. https://doi.org/10.1093/brain/117.6.1241

Brownell, H. H. (1988). Appreciation of Metaphoric and Connotative Word Meaning by Brain-Damaged Patients. In C. Chiarello (Ed.), *Right Hemisphere Contributions to Lexical Semantics* (pp. 19–31). https://doi.org/10.1007/978-3-642-73674-2_2

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

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, *16*(5), 893–900. https://doi.org/10.3758/PBR.16.5.893

Bryant, G. A. (2010). Prosodic Contrasts in Ironic Speech. *Discourse Processes*, *47*(7), 545–566. https://doi.org/10.1080/01638530903531972

Bryant, G. A., & Tree, J. E. F. (2002). Recognizing Verbal Irony in Spontaneous Speech. *Metaphor and Symbol*, *17*(2), 99–119. https://doi.org/10.1207/S15327868MS1702_2

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Cacciari, C., & Glucksberg, S. (1994). Understanding figurative language. In *Handbook of psycholinguistics* (pp. 447–477). San Diego, CA, US: Academic Press.

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

Caramazza, A., & McCloskey, M. (1988). The case for single-patient studies. *Cognitive Neuropsychology*, *5*(5), 517–527. https://doi.org/10.1080/02643298808253271

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

Carp, J. (2012). The secret lives of experiments: Methods reporting in the fMRI literature. *NeuroImage*, *63*(1), 289–300. https://doi.org/10.1016/j.neuroimage.2012.07.004

Caspers, S., Eickhoff, S. B., Geyer, S., Scheperjans, F., Mohlberg, H., Zilles, K., & Amunts, K. (2008). The human inferior parietal lobule in stereotaxic space. *Brain Structure and Function*, *212*(6), 481–495. https://doi.org/10.1007/s00429-008-0195-z

Caspers, S., Geyer, S., Schleicher, A., Mohlberg, H., Amunts, K., & Zilles, K. (2006). The human inferior parietal cortex: Cytoarchitectonic parcellation and interindividual variability. *NeuroImage*, *33*(2), 430–448. https://doi.org/10.1016/j.neuroimage.2006.06.054

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. https://doi.org/10.1037/h0046743

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. (2015). 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., Liao, Y.-J., Tu, C.-H., & Chen, H.-C. (2016). Neural Correlates of Hostile Jokes: Cognitive and Motivational Processes in Humor Appreciation. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00527

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

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

Clark, H. H., & Lucy, P. (1975). Understanding what is meant from what is said: A study in conversationally conveyed requests. *Journal of Verbal Learning and Verbal Behavior*, *14*(1), 56–72. https://doi.org/10.1016/S0022-5371(75)80006-5

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

Colle, L., Angeleri, R., Vallana, M., Sacco, K., Bara, B. G., & Bosco, F. M. (2013). Understanding the communicative impairments in schizophrenia: A preliminary study. *Journal of Communication Disorders*, *46*(3), 294–308. https://doi.org/10.1016/j.jcomdis.2013.01.003

Collins, D. L., Neelin, P., Peters, T. M., & Evans, A. C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography*, *18*(2), 192–205.

Colston, H. L., & Gibbs, R. W. G. (2002). Are Irony and Metaphor Understood Differently? *Metaphor and Symbol*, *17*(1), 57–80. https://doi.org/10.1207/S15327868MS1701_5

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151–216). London: Academic Press.

Costafreda, S. G. (2009). Pooling fMRI data: meta-analysis, mega-analysis and multi-center studies. *Frontiers in Neuroinformatics*, *3*. https://doi.org/10.3389/neuro.11.033.2009

Coulson, S. (2001). *Semantic Leaps: Frame-Shifting and Conceptual Blending in Meaning Construction*. Cambridge University Press.

Coulson, S., & Kutas, M. (2001). Getting it: human event-related brain response to jokes in good and poor comprehenders. *Neuroscience Letters*, *316*(2), 71–74. https://doi.org/10.1016/S0304-3940(01)02387-4

Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition*, *30*(6), 958–968. https://doi.org/10.3758/BF03195780

Culpeper, J., Bousfield, D., & Wichmann, A. (2003). Impoliteness revisited: with special reference to dynamic and prosodic aspects. *Journal of Pragmatics*, *35*(10), 1545–1579. https://doi.org/10.1016/S0378-2166(02)00118-2

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., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional Organization of Social Perception and Cognition in the Superior Temporal Sulcus. *Cerebral Cortex (New York, N.Y.: 1991)*, *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

Dennis, M., & Barnes, M. A. (1990). Knowing the meaning, getting the point, bridging the gap, and carrying the message: Aspects of discourse following closed head injury in childhood and adolescence. *Brain and Language*, *39*(3), 428–446. https://doi.org/10.1016/0093-934X(90)90149-B

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diaz, M. T., & Hogstrom, L. J. (2011). 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

Dodell-Feder, D., Koster-Hale, J., Bedny, M., & Saxe, R. (2011). fMRI item analysis in a theory of mind task. *NeuroImage*, *55*(2), 705–712. https://doi.org/10.1016/j.neuroimage.2010.12.040

Dove, A., Pollmann, S., Schubert, T., Wiggins, C. J., & Yves von Cramon, D. (2000). Prefrontal cortex activation in task switching: an event-related fMRI study. *Cognitive Brain Research*, *9*(1), 103–109. https://doi.org/10.1016/S0926-6410(99)00029-4

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nature Neuroscience*, *3*(3), 277–283. https://doi.org/10.1038/72991

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2001). The Effect of Task Relevance on the Cortical Response to Changes in Visual and Auditory Stimuli: An Event-Related fMRI Study. *NeuroImage*, *14*(6), 1256–1267. https://doi.org/10.1006/nimg.2001.0946

Dumontheil, I., Thompson, R., & Duncan, J. (2010). Assembly and Use of New Task Rules in Fronto-parietal Cortex. *Journal of Cognitive Neuroscience*, *23*(1), 168–182. https://doi.org/10.1162/jocn.2010.21439

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience*, *2*(11), 820–829. https://doi.org/10.1038/35097575

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., & 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

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

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

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

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

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., Mineroff, Z., Siegelman, M., & Blank, I. (2018). Word meanings and sentence structure recruit the same set of fronto-temporal regions during comprehension. *BioRxiv*, 477851. https://doi.org/10.1101/477851

Fedorenko, E., Nieto-Castañon, A., & Kanwisher, N. (2012). Lexical and syntactic representations in the brain: An fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia*, *50*(4), 499–513. https://doi.org/10.1016/j.neuropsychologia.2011.09.014

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

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., Rinck, M., & von Cramon, D. Y. (2005). Emotional and temporal aspects of situation model processing during text comprehension: An event-related fMRI study. *Journal of Cognitive Neuroscience*, *17*(5), 724–739. https://doi.org/10.1162/0898929053747658

Ferstl, E. C., & von Cramon, D. Y. (2002). What Does the Frontomedian Cortex Contribute to Language Processing: Coherence or Theory of Mind? *NeuroImage*, *17*(3), 1599–1612. https://doi.org/10.1006/nimg.2002.1247

Ferstl, E. C., Walther, K., Guthke, T., & Cramon, D. Y. von. (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

Fischl, B., Sereno, M. I., Tootell, R. B. H., & Dale, A. M. (1999). High-resolution inter-subject averaging and a surface-based coordinate system. *Hum Brain Mapp*, *8*, 272-284.

Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Essen, D. C. V., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, *102*(27), 9673–9678. https://doi.org/10.1073/pnas.0504136102

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

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

Friston, K. J., Rotshtein, P., Geng, J. J., Sterzer, P., & Henson, R. N. (2006). A critique of functional localisers. *NeuroImage*, *30*(4), 1077–1087. https://doi.org/10.1016/j.neuroimage.2005.08.012

Frith, C. D., & Frith, U. (1999). Interacting Minds—A Biological Basis. *Science*, *286*(5445), 1692–1695. https://doi.org/10.1126/science.286.5445.1692

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

Galaburda, A. M., Rosen, G. D., & Sherman, G. F. (1990). Individual variability in cortical organization: Its relationship to brain laterality and implications to function. *Neuropsychologia*, *28*(6), 529–546. https://doi.org/10.1016/0028-3932(90)90032-J

Gardner H., Brownell H., Wapner W., & Michelow D. (1983). Missing the point: The role of the right hemisphere in the processing of complex linguistic materials. In: E. Perecman (Ed.), *Cognitive processing in the right hemisphere*. New York: Academic Press, pp. 169-191.

Gavilán, J. M., & García-Albea, J. E. (2011). Theory of mind and language comprehension in schizophrenia: Poor mindreading affects figurative language comprehension beyond intelligence deficits. *Journal of Neurolinguistics*, *24*(1), 54–69. https://doi.org/10.1016/j.jneuroling.2010.07.006

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

Gibbs, R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition*, *8*(2), 149–156. https://doi.org/10.3758/BF03213418

Gibbs, R. W. (1984). Literal meaning and psychological theory. *Cognitive Science*, *8*(3), 275–304. https://doi.org/10.1016/S0364-0213(84)80004-X

Gibbs, R. W. (2000). Irony in Talk Among Friends. *Metaphor and Symbol*, *15*(1–2), 5–27. https://doi.org/10.1080/10926488.2000.9678862

Gibbs, R. W., & Beitel, D. (1995). What proverb understanding reveals about how people think. *Psychological Bulletin*, *118*(1), 133–154. https://doi.org/10.1037/0033-2909.118.1.133

Gibbs, R. W., Colston, H. L., & Johnson, M. D. (1996). Proverbs and the metaphorical mind. *Metaphor and Symbol*, *11*(3), 207-216.

Giora, R. (1997). Discourse coherence and theory of relevance: Stumbling blocks in search of a unified theory. *Journal of Pragmatics*, *27*(1), 17–34. https://doi.org/10.1016/0378-2166(95)00065-8

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

Gloor, P. (1997). *The Temporal Lobe and Limbic System*. Oxford University Press.

Glucksberg, S. (1991). Beyond literal meanings: The psychology of allusion. *Psychological Science*, *2*(3), 146–152. https://doi.org/10.1111/j.1467-9280.1991.tb00122.x

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences*, *7*(2), 92–96. https://doi.org/10.1016/S1364-6613(02)00040-2

Glucksberg, S., McGlone, M. S., Grodzinsky, Y., & Amunts, K. (2001). *Understanding Figurative Language: From Metaphor to Idioms*. Oxford University Press, USA.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, *100*(1), 253–258. https://doi.org/10.1073/pnas.0135058100

Happé, F. G. E. (1995). The Role of Age and Verbal Ability in the Theory of Mind Task Performance of Subjects with Autism. *Child Development*, *66*(3), 843–855. https://doi.org/10.1111/j.1467-8624.1995.tb00909.x

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Hellbernd, N., & Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language*, *88*, 70–86. https://doi.org/10.1016/j.jml.2016.01.001

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., Tényi, T., Simon, M., & Trixler, M. (2004). [The connection between mentalization deficit and pragmatic language skills and neurocognition in schizophrenia]. *Neuropsychopharmacologia Hungarica : a Magyar Pszichofarmakologiai Egyesulet lapja = official journal of the Hungarian Association of Psychopharmacology*, *6*(2), 72–78.

Herold, Róbert, 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

Holmes, A. P., & Friston, K. J. (1998). Generalisability, Random Effects & Population Inference. *NeuroImage*, *7*. https://doi.org/10.1016/S1053-8119(18)31587-8

Holtgraves, T. (1999). Comprehending indirect replies: When and how are their conveyed meanings activated? *Journal of Memory and Language*, *41*(4), 519–540. https://doi.org/10.1006/jmla.1999.2657

Honan, C. A., McDonald, S., Gowland, A., Fisher, A., & Randall, R. K. (2015). Deficits in comprehension of speech acts after TBI: The role of theory of mind and executive function. *Brain and Language*, *150*, 69–79. https://doi.org/10.1016/j.bandl.2015.08.007

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

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Ioannidis, J. P. A., Munafò, M. R., Fusar-Poli, P., Nosek, B. A., & David, S. P. (2014). Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. *Trends in Cognitive Sciences*, *18*(5), 235–241. https://doi.org/10.1016/j.tics.2014.02.010

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. 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

Jacoby, N., & Fedorenko, E. (2018). Discourse-level comprehension engages medial frontal Theory of Mind brain regions even for expository texts. *Language, Cognition and Neuroscience*, *0*(0), 1–17. https://doi.org/10.1080/23273798.2018.1525494

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

Jin, H., Liu, H.-L., Mo, L., Fang, S.-Y., Zhang, J. X., & Lin, C.-D. (2009). Involvement of the left inferior frontal gyrus in predictive inference making. *International Journal of Psychophysiology*, *71*(2), 142–148. https://doi.org/10.1016/j.ijpsycho.2008.08.009

Joanette, Y., Goulet, P., Hannequin, D., & Boeglin, J. (1990). *Right hemisphere and verbal communication*. In *Right Hemisphere and Verbal Communication*. https://doi.org/10.1007/978-1-4612-4460-8

Johns, C. L., Tooley, K. M., & Traxler, M. J. (2008). Discourse Impairments Following Right Hemisphere Brain Damage: A Critical Review. *Language and Linguistics Compass*, *2*(6), 1038–1062. https://doi.org/10.1111/j.1749-818X.2008.00094.x

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Harvard University Press.

Jones, E. G., & Powell, T. P. S. (1970). AN ANATOMICAL STUDY OF CONVERGING SENSORY PATHWAYS WITHIN THE CEREBRAL CORTEX OF THE MONKEY. *Brain*, *93*(4), 793–820. https://doi.org/10.1093/brain/93.4.793

Juch, H., Zimine, I., Seghier, M. L., Lazeyras, F., & Fasel, J. H. D. (2005). Anatomical variability of the lateral frontal lobe surface: implication for intersubject variability in language neuroimaging. *NeuroImage*, *24*(2), 504–514. https://doi.org/10.1016/j.neuroimage.2004.08.037

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences*, *9*(11), 512–518. https://doi.org/10.1016/j.tics.2005.09.009

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Kasparian, K. (2013). Hemispheric differences in figurative language processing: Contributions of neuroimaging methods and challenges in reconciling current empirical findings. *Journal of Neurolinguistics*, *26*(1), 1–21. https://doi.org/10.1016/j.jneuroling.2012.07.001

Keysar, B. (1989). On the functional equivalence of literal and metaphorical interpretations in discourse. *Journal of Memory and Language*, *28*(4), 375–385. https://doi.org/10.1016/0749-596X(89)90017-X

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic Facilitation and Interference in the Resolution of Temporary Syntactic Closure Ambiguity. *Journal of Memory and Language*, *40*(2), 153–194. https://doi.org/10.1006/jmla.1998.2620

Kline M., Gallée J., Balewski Z., & Fedorenko E. Understanding jokes relies on the theory of mind system. Submitted. (Website for the pre-registration: https://osf.io/mzyjd/register/564d31db8c5e4a7c9694b2be)

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., McGuire, P. K., Bullmore, E. T., Brammer, M. J., Rabe-Hesketh, S., Wright, I. C., … 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

Kuperberg, Gina 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

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72–79. https://doi.org/10.1016/j.tics.2007.11.004

Laird, A. R., Fox, P. M., Price, C. J., Glahn, D. C., Uecker, A. M., Lancaster, J. L., … Fox, P. T. (2005). ALE meta-analysis: Controlling the false discovery rate and performing statistical contrasts. *Human Brain Mapping*, *25*(1), 155–164. https://doi.org/10.1002/hbm.20136

Lakoff, G., & Johnson, M. (1980). Conceptual Metaphor in Everyday Language. *The Journal of Philosophy*, *77*(8), 453–486. https://doi.org/10.2307/2025464

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, J. J., & Pinker, S. (2010). Rationales for indirect speech: the theory of the strategic speaker. *Psychological Review*, *117*(3), 785–807. https://doi.org/10.1037/a0019688

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

Lezak, M. (1995). Neuropsychological assessment (3rd ed.). New York: Oxford University Press.

Li, S., Ostwald, D., Giese, M., & Kourtzi, Z. (2007). Flexible Coding for Categorical Decisions in the Human Brain. *Journal of Neuroscience*, *27*(45), 12321–12330. https://doi.org/10.1523/JNEUROSCI.3795-07.2007

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

Lindell, A. K. (2006). In Your Right Mind: Right Hemisphere Contributions to Language Processing and Production. *Neuropsychology Review*, *16*(3), 131–148. https://doi.org/10.1007/s11065-006-9011-9

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

Lockl, K., & Schneider, W. (2006). Precursors of metamemory in young children: the role of theory of mind and metacognitive vocabulary. *Metacognition and Learning*, *1*(1), 15–31. https://doi.org/10.1007/s11409-006-6585-9

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution [corrected]. *Psychological Review*, *101*(4), 676–703.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1), 71–102. https://doi.org/10.1016/0010-0277(87)90005-9

Martin, I., & McDonald, S. (2003). Weak coherence, no theory of mind, or executive dysfunction? Solving the puzzle of pragmatic language disorders. *Brain and Language*, *85*(3), 451–466. https://doi.org/10.1016/S0093-934X(03)00070-1

Martin, I., & McDonald, S. (2005). Evaluating the causes of impaired irony comprehension following traumatic brain injury. *Aphasiology*, *19*(8), 712–730. https://doi.org/10.1080/02687030500172203

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. (2005). 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. (2009). The Role of the Theory-of-Mind Cortical Network in the Comprehension of Narratives. *Language and Linguistics Compass*, *3*(1), 157–174. https://doi.org/10.1111/j.1749-818X.2008.00122.x

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

Matsui, T., Nakamura, T., Utsumi, A., Sasaki, A. T., Koike, T., Yoshida, Y., … 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

McDonald, S. (1999). Exploring the Process of Inference Generation in Sarcasm: A Review of Normal and Clinical Studies. *Brain and Language*, *68*(3), 486–506. https://doi.org/10.1006/brln.1999.2124

McDonald, S., & Pearce, S. (1996). Clinical Insights into Pragmatic Theory: Frontal Lobe Deficits and Sarcasm. *Brain and Language*, *53*(1), 81–104. https://doi.org/10.1006/brln.1996.0038

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

McDonald, S., & Sommers, P. van. (1993). Pragmatic language skills alter closed head injury: Ability to negotiate requests. *Cognitive Neuropsychology*, *10*(4), 297–315. https://doi.org/10.1080/02643299308253466

Meyer. *Un Algorithme Optimal Pour La Ligne De Partage Des Eaux. 8me Congrès de Reconnaissance des Formes et Intelligence Artificielle, Lyon, France,* 1991.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Miller, M. B., Van Horn, J. D., Wolford, G. L., Handy, T. C., Valsangkar-Smyth, M., Inati, S., … Gazzaniga, M. S. (2002). Extensive Individual Differences in Brain Activations Associated with Episodic Retrieval are Reliable Over Time. *Journal of Cognitive Neuroscience*, *14*(8), 1200–1214. https://doi.org/10.1162/089892902760807203

Mineroff, Z., Blank, I. A., Mahowald, K., & Fedorenko, E. (2018). A robust dissociation among the language, multiple demand, and default mode networks: Evidence from inter-region correlations in effect size. *Neuropsychologia*, *119*, 501–511. https://doi.org/10.1016/j.neuropsychologia.2018.09.011

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

Mollica, F., Siegelman, M., Diachek, E., Piantadosi, S. T., Mineroff, Z., Futrell, R., & Fedorenko, E. (2018). High local mutual information drives the response in the human language network. *BioRxiv*, 436204. https://doi.org/10.1101/436204

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 *Clinical Aphasiology: Proceedings of the Conference 1981* (pp. 254-261). BRK Publishers.

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., … 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

Ono, M., Kubik, S., & Abernathey, C. D. (1990). *Atlas of the Cerebral Sulci*. Thieme.

Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. J. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior*, *17*(4), 465–477. https://doi.org/10.1016/S0022-5371(78)90283-9

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

Papagno, C., Tabossi, P., Colombo, M. R., & Zampetti, P. (2004). Idiom comprehension in aphasic patients. *Brain and Language*, *89*(1), 226–234. https://doi.org/10.1016/S0093-934X(03)00398-5

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

Paus, T., Tomaiuolo, F., Otaky, N., MacDonald, D., Petrides, M., Atlas, J., … Evans, A. C. (1996). Human Cingulate and Paracingulate Sulci: Pattern, Variability, Asymmetry, and Probabilistic Map. *Cerebral Cortex*, *6*(2), 207–214. https://doi.org/10.1093/cercor/6.2.207

Pell, M. D. (2006). Judging emotion and attitudes from prosody following brain damage. In S. Anders, G. Ende, M. Junghofer, J. Kissler, & D. Wildgruber (Eds.), *Progress in Brain Research* (pp. 303–317). https://doi.org/10.1016/S0079-6123(06)56017-0

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., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., … Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*(2), 115–126. https://doi.org/10.1038/nrn.2016.167

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

Prado, J., Spotorno, N., Koun, E., Hewitt, E., Van der Henst, J.-B., Sperber, D., & Noveck, I. A. (2014). Neural Interaction between Logical Reasoning and Pragmatic Processing in Narrative Discourse. *Journal of Cognitive Neuroscience*, *27*(4), 692–704. https://doi.org/10.1162/jocn_a_00744

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

Pritchett, B. L., Hoeflin, C., Koldewyn, K., Dechter, E., & Fedorenko, E. (2018). High-level language processing regions are not engaged in action observation or imitation. *Journal of Neurophysiology*, *120*(5), 2555–2570. https://doi.org/10.1152/jn.00222.2018

Radua, J., Mataix-Cols, D., Phillips, M. L., El-Hage, W., Kronhaus, D. M., Cardoner, N., & Surguladze, S. (2012). A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. *European Psychiatry*, *27*(8), 605–611. https://doi.org/10.1016/j.eurpsy.2011.04.001

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

Rapp, Alexander 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, Alexander 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, Alexander 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

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

Rypma, B., & D'Esposito, M. (1999). The roles of prefrontal brain regions in components of working memory: Effects of memory load and individual differences. *Proceedings of the National Academy of Sciences*, *96*(11), 6558–6563. https://doi.org/10.1073/pnas.96.11.6558

Samur, D., Lai, V. T., Hagoort, P., & Willems, R. M. (2015). 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, Rebecca. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*(2), 235–239. https://doi.org/10.1016/j.conb.2006.03.001

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

Saxe, Rebecca, & 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

Saxe, Rebecca, Schulz, L. E., & Jiang, Y. V. (2006). Reading minds versus following rules: Dissociating theory of mind and executive control in the brain. *Social Neuroscience*, *1*(3–4), 284–298. https://doi.org/10.1080/17470910601000446

Scheperjans, F., Eickhoff, S. B., Hömke, L., Mohlberg, H., Hermann, K., Amunts, K., & Zilles, K. (2008). Probabilistic Maps, Morphometry, and Variability of Cytoarchitectonic Areas in the Human Superior Parietal Cortex. *Cerebral Cortex*, *18*(9), 2141–2157. https://doi.org/10.1093/cercor/bhm241

Schmidt, Gwen L., DeBuse, C. J., & Seger, C. A. (2007). Right hemisphere metaphor processing? Characterizing the lateralization of semantic processes. *Brain and Language*, *100*(2), 127–141. https://doi.org/10.1016/j.bandl.2005.03.002

Schmidt, Gwenda L., Kranjec, A., Cardillo, E. R., & Chatterjee, A. (2010). Beyond laterality: A critical assessment of research on the neural basis of metaphor. *Journal of the International Neuropsychological Society*, *16*(1), 1–5. https://doi.org/10.1017/S1355617709990543

Schmidt, Gwenda 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

Schuh, J. M., Eigsti, I.-M., & Mirman, D. (2016). Discourse comprehension in autism spectrum disorder: Effects of working memory load and common ground. *Autism Research*, *9*(12), 1340–1352. https://doi.org/10.1002/aur.1632

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

Searle, J. R. (1979). What Is an Intentional State? *Mind*, *88*(349), 74–92. Retrieved from JSTOR.

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

Silvia, P. J., & Beaty, R. E. (2012). Making creative metaphors: The importance of fluid intelligence for creative thought. *Intelligence*, *40*(4), 343–351. https://doi.org/10.1016/j.intell.2012.02.005

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*(1), 103–130. https://doi.org/10.1016/S0749-596X(02)00519-3

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

Sperber, D., & Wilson, D. (2002). Pragmatics, Modularity and Mind-reading. *Mind & Language*, *17*(1–2), 3–23. https://doi.org/10.1111/1468-0017.00186

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

Sridharan, D., Levitin, D. J., Chafe, C. H., Berger, J., & Menon, V. (2007). Neural Dynamics of Event Segmentation in Music: Converging Evidence for Dissociable Ventral and Dorsal Networks. *Neuron*, *55*(3), 521–532. https://doi.org/10.1016/j.neuron.2007.07.003

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

Swallow, K. M., Braver, T. S., Snyder, A. Z., Speer, N. K., & Zacks, J. M. (2003). Reliability of functional localization using fMRI. *NeuroImage*, *20*(3), 1561–1577. https://doi.org/10.1016/S1053-8119(03)00436-1

Talairach, J., & Tournoux, P. (1988). Co-planar stereotaxic atlas of the human brain (p. 122). New York, NY: Thieme Medical Publishers.

Thompson-Schill, S. L., D'Esposito, M., & Kan, I. P. (1999). Effects of Repetition and Competition on Activity in Left Prefrontal Cortex during Word Generation. *Neuron*, *23*(3), 513–522. https://doi.org/10.1016/S0896-6273(00)80804-1

Tian, F., Hou, Y., Zhu, W., Dietrich, A., Zhang, Q., Yang, W., … 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

Tomaiuolo, F., MacDonald, J. D., Caramanos, Z., Posner, G., Chiavaras, M., Evans, A. C., & Petrides, M. (1999). Morphology, morphometry and probability mapping of the pars opercularis of the inferior frontal gyrus: an in vivo MRI analysis. *European Journal of Neuroscience*, *11*(9), 3033–3046. https://doi.org/10.1046/j.1460-9568.1999.00718.x

Tompkins, 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

Tompkins, A., Lehman-Blake, T., Baumgaertner, A., & Fassbinder, W. (2001). Mechanisms of discourse comprehension impairment after right hemisphere brain damage. *Journal of Speech, Language, and Hearing Research*, *44*(2), 400–415. https://doi.org/10.1044/1092-4388(2001/033)

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like tweedledum and tweedledee but they can (and should) be distinguished. *Psychological Science*, *1*(1), 46–54. https://doi.org/10.1111/j.1467-9280.1990.tb00067.x

Turkeltaub, P. E., & Branch Coslett, H. (2010). Localization of sublexical speech perception components. *Brain and Language*, *114*(1), 1–15. https://doi.org/10.1016/j.bandl.2010.03.008

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

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, Z., Tényi, T., Németh, N., Simon, M., Hajnal, A., … 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, Z., Hajnal, A., Orsi, G., … 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

Varley, R. A., Klessinger, N. J. C., Romanowski, C. A. J., & Siegal, M. (2005). Agrammatic but numerate. *Proceedings of the National Academy of Sciences*, *102*(9), 3519–3524. https://doi.org/10.1073/pnas.0407470102

Varley, R., & Siegal, M. (2000). Evidence for cognition without grammar from causal reasoning and 'theory of mind' in an agrammatic aphasic patient. *Current Biology*, *10*(12), 723–726. https://doi.org/10.1016/S0960-9822(00)00538-8

Varley, R., Siegal, M., & Want, S. C. (2001). Severe Impairment in Grammar Does Not Preclude Theory of Mind. *Neurocase*, *7*(6), 489–493. https://doi.org/10.1093/neucas/7.6.489

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

Volz, K. G., Vogeley, K., Tittgemeyer, M., von Cramon, D. Y., & Sutter, M. (2015). The neural basis of deception in strategic interactions. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00027

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

Wager, T. D., Lindquist, M., & Kaplan, L. (2007). Meta-analysis of functional neuroimaging data: current and future directions. *Social Cognitive and Affective Neuroscience*, *2*(2), 150–158. https://doi.org/10.1093/scan/nsm015

Wakusawa, K., Sugiura, M., Sassa, Y., Jeong, H., Horie, K., Sato, S., … 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

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

Wichmann, A. (2004). The intonation of Please-requests: a corpus-based study. *Journal of Pragmatics*, *36*(9), 1521–1549. https://doi.org/10.1016/j.pragma.2004.03.003

Wilson D, & Sperber D. (2003). Relevance theory. In *Handbook of Pragmatics*, ed. G Ward, L Horn, pp. 607–32. Oxford, UK: Blackwell Sci.

Wilson, D., & Sperber, D. (2012). *Meaning and Relevance*. Cambridge University Press.

Winner, E. (1997). *The Point of Words: Children's Understanding of Metaphor and Irony*. Harvard University Press.

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

Winner, E., & Leekam, S. (1991). Distinguishing irony from deception: Understanding the speaker's second-order intention. *British Journal of Developmental Psychology*, *9*(2), 257–270. https://doi.org/10.1111/j.2044-835X.1991.tb00875.x

Wise, R. J. S., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural subsystems within `Wernicke's area'. *Brain*, *124*(1), 83–95. https://doi.org/10.1093/brain/124.1.83

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J., & Reed, B. (2015). Factors influencing the latency of simple reaction time. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00131

Woolgar, A., Duncan, J., Manes, F., & Fedorenko, E. (2018). Fluid intelligence is supported by the multiple-demand system not the language system. *Nature Human Behaviour*, *2*(3), 200. https://doi.org/10.1038/s41562-017-0282-3

Woolgar, A., Parr, A., Cusack, R., Thompson, R., Nimmo-Smith, I., Torralva, T., … 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

Wyer, R. S., & Collins, J. E. (1992). A theory of humor elicitation. *Psychological Review*, *99*(4), 663–688. https://doi.org/10.1037/0033-295X.99.4.663

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., 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

Table 3

Included articles

Authors	Year	Type	Language	Ppt	P/C	Modality	Peaks	Task
AbdulSabur et al.	2014	Text Coherence	English	18	C	Auditory	12	Production or no task
Adamczyk et al.	2017	Humor	Polish	20	C	Visual	5	Valence judgment
Ahrens et al.	2007	Metaphor	Mandarin	8	C	Visual	1	No task
Akimoto et al.	2014	Irony	Japanese	25	C	Visual	2	No task
Bambini et al.	2011	Metaphor	Italian	9	C	Visual	10	Semantic judgment
Bašnáková et al.	2015	Indirect Speech	Dutch	20	C	Visual	42	Social judgment
Bašnáková et al.	2014	Indirect Speech	Dutch	28	C	Auditory	29	Semantic judgment
Bekinschtein et al.	2011	Humor	English	12	C	Auditory	34	No task
Benedek et al.	2014	Metaphor	German	28	P	N/A	8	Production task
Bosco et al.	2017	Irony	Italian	12	C	Visual	6	Semantic judgment
Campbell et al.	2015	Humor	English	28	C	Visual	20	No task
Chan & Lavallee	2015	Humor	Mandarin	18	C	Visual	6	Valence judgment
Chan et al.	2016	Humor	Mandarin	24	C	Visual	5	Valence judgment
Chen et al.	2008	Metaphor	English	27	C	Visual	7	Semantic judgment

Chow et al.	2008	Text Coherence	German	25	C	Visual	8	Semantic judgment
Citron & Goldberg	2014	Metaphor	German	14	C	Visual	19	Semantic judgment
Citron et al.	2016	Metaphor	German	15	C	Visual	69	Semantic judgment
Dai et al.	2017	Humor	Mandarin	26	C	Visual	8	Valence judgment
den Ouden et al.	2016	Non-affective Prosody	English	24	C	Auditory	8	Semantic judgment
Diaz & Hogstrom	2011	Metaphor	English	27	C	Visual	6	Semantic judgment
Egorova et al.	2016	Indirect Speech	English	16	C	Visual	11	No task
Feng et al.	2017	Indirect Speech	Mandarin	16	C	Visual	7	Semantic judgment
Ferstl & von Cramon	2001	Text Coherence	German	18	C	Auditory	2	Semantic judgment
Ferstl & von Cramon	2002	Text Coherence	German	19	C	Visual	12	Semantic judgment
Friese et al.	2008	Text Coherence	German	12	C	Auditory	1	Semantic judgment
Hellbernd & Sammler	2018	Non-affective Prosody	German	9	C	Auditory	24	Semantic judgment
Herold et al.	2018	Irony	Hungarian	13	C	Visual	7	Semantic judgment
Hillert & Buracas	2009	Idiom	English	22	C	Auditory	11	Semantic judgment
Jang et al.	2013	Indirect Speech	Korean	12	C	Auditory	16	Semantic judgment

Kristensen et al.	2013	Non-affective Prosody	Dutch	20	C	Visual	22	Semantic and perceptual task
Kuperberg et al.	2006	Text Coherence	English	23	C	Visual	8	Semantic judgment
Lauro et al.	2008	Idiom	Italian	24	C	Auditory	10	Semantic judgment
Lee & Dapretto	2006	Metaphor	English	15	C	Visual	3	Semantic judgment
Lin et al.	2018	Text Coherence	Mandarin	22	C	Visual	13	Semantic judgment
Martín-Loeches et al.	2008	Text Coherence	Spanish	12	C	Visual	13	No task
Mashal et al.	2007	Metaphor	Hebrew	39	C	Visual	8	Semantic judgment
Mashal et al.	2009	Metaphor	Hebrew	23	C	Visual	1	Semantic judgment
Mason & Just	2011	Text Coherence	English	15	C	Auditory	30	Semantic judgment
Matsui et al.	2016	Sarcasm	Japanese	15	C	Auditory	1	Semantic judgment
Nagels et al.	2013	Metaphor	German	14	C	Auditory	2	Semantic judgment
Obert et al.	2016	Irony	French	21	C	Auditory	8	Categorical judgment
Obert et al.	2014	Metaphor	French	16	C	Auditory	3	Categorical judgment
Perrone-Bertolotti et al.	2013	Non-affective Prosody	French	21	C	Visual	8	Semantic judgment
Piñango et al.	2016	Metonymy	English	19	C	Visual	2	Semantic judgment

Pomp et al.	2018	Metaphor	German	24	C	Visual	2	Semantic judgment
Prado et al.	2015	Indirect Speech	French	16	C	Visual	6	Semantic judgment
Rapp et al.	2010	Irony	German	18	C	Visual	6	Perceptual judgment
Rapp et al.	2011	Metonymy	German	20	C	Visual	8	Semantic judgment
Samur et al.	2015	Metaphor	Dutch	15	C	Visual	3	Semantic judgment
Schmidt & Seger	2009	Metaphor	English	14	C	Visual	17	No task
Schuil et al.	2013	Metaphor	Dutch	20	C	Visual	1	Semantic judgment
Shibata et al.	2014	Humor	Japanese	11	C	Visual	17	Valence judgment
Shibata et al.	2011	Indirect Speech	Japanese	20	C	Visual	10	Valence judgment
Shibata et al.	2010	Irony	Japanese	20	C	Visual	4	Categorical judgment
Shibata et al.	2012	Metaphor	Japanese	15	C	Visual	20	Check understanding Y/N
Shibata et al.	2007	Metaphor	Japanese	13	C	Visual	5	Check understanding Y/N
Spotorno et al.	2012	Irony	French	24	C	Visual	11	Semantic judgment
Stingaris et al.	2007	Metaphor	English	13	C	Visual	9	Check understanding Y/N
Tian et al.	2017	Humor	Mandarin	20	C	Visual	12	Valence judgment

van Ackeren et al.	2012	Indirect Speech	Dutch	11	C	Auditory	26	Semantic judgment
van Ackeren et al.	2016	Indirect Speech	English	24	C	Auditory	5	Semantic judgment
Varga et al.	2013	Irony	Hungarian	13	C	Auditory	18	Semantic judgment
Virtue et al.	2006	Text Coherence	English	22	C	Auditory	3	No task
Virtue et al.	2008	Text Coherence	English	24	C	Auditory	9	No task
Wakusawa et al.	2007	Irony	Japanese	38	C	Visual	2	Semantic judgment
Wang et al.	2006	Irony	English	12	C	Auditory	4	Attentional cue
Yang et al.	2016	Idiom	Mandarin	19	C	Visual	4	Perceptual judgment
Yang et al.	2009	Metaphor	English	18	C	Visual	15	Valence judgment
Yarkoni et al.	2008	Text Coherence	English	29	C	Visual	8	Memory task
Yi et al.	2017	Proverb	Korean	15	C	Visual	9	No task
Zempleni et al.	2007	Idiom	Dutch	15	C	Visual	8	Semantic judgment

Note. Ppt = participants, P = production, C = comprehension.

Appendix A

The Language, MD, and ToM Networks

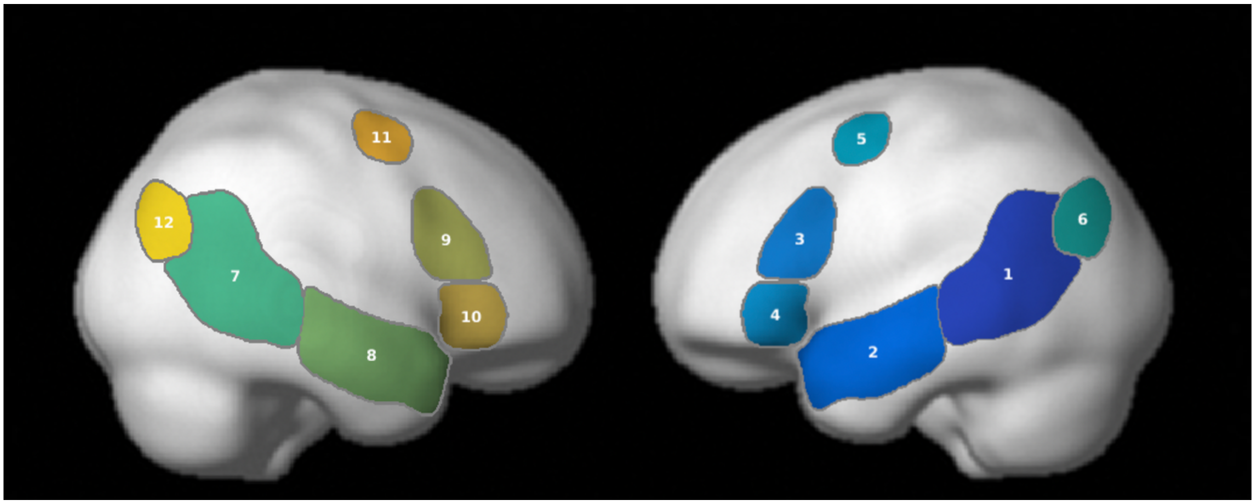

Figure 1. The language network. This figure illustrates the regions comprising the language network.

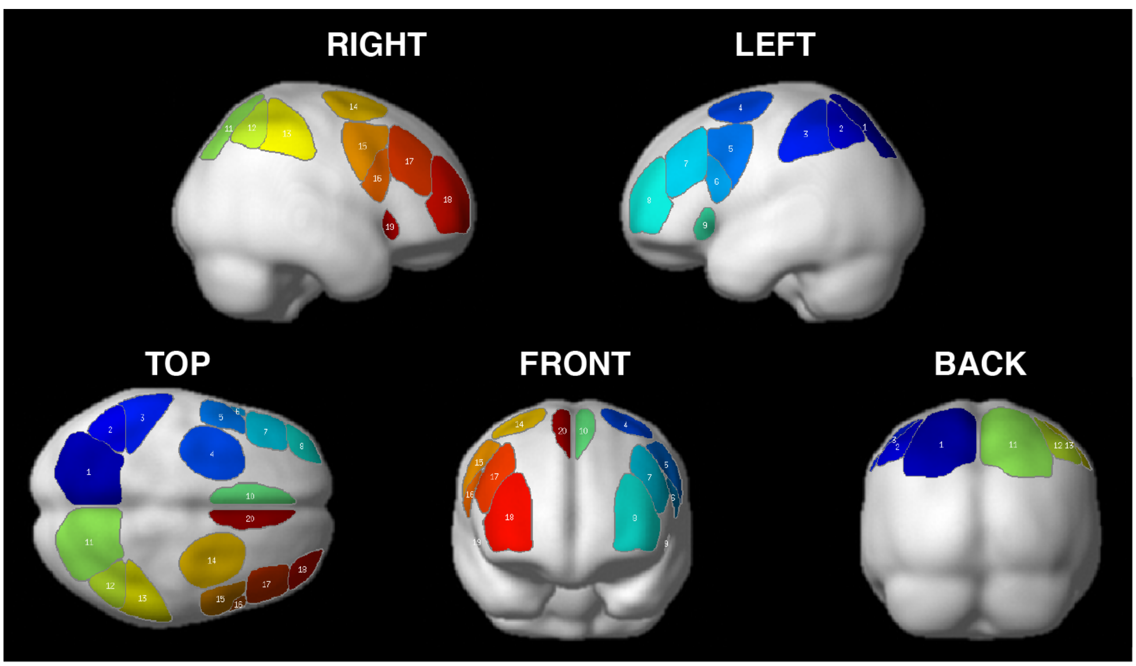

Figure 2. The MD network. This figure illustrates the regions comprising the Multiple Demand network.

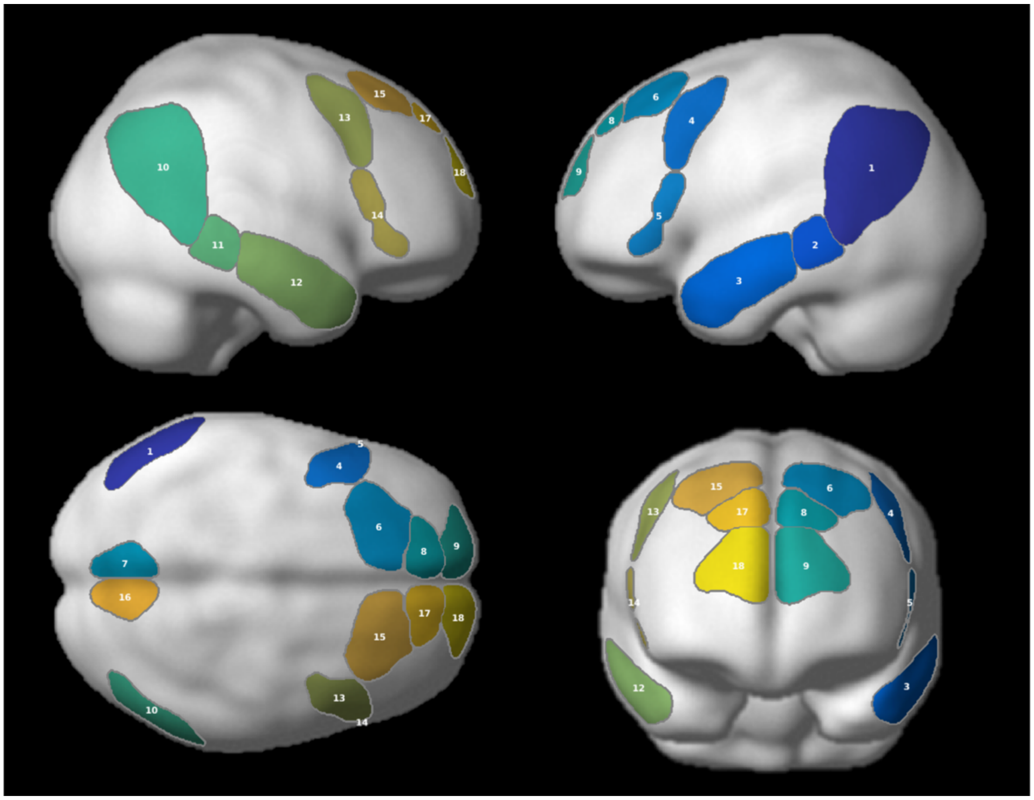

Figure 3. The ToM network. This figure illustrates the regions comprising the Theory of Mind network.

Appendix B

Graphical Depictions of Results

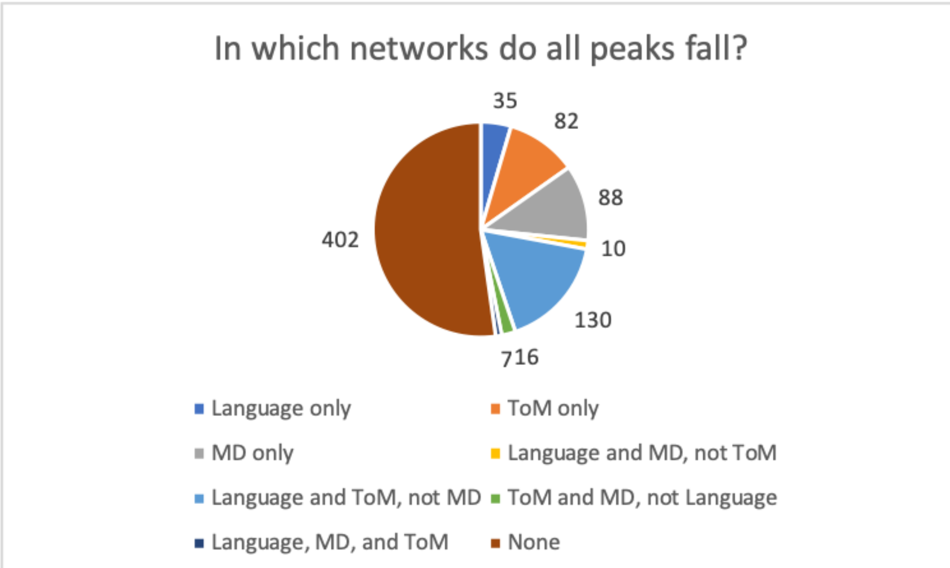

Figure 4. Breakdown of all activation peaks. This figure illustrates the networks into which all included activation peaks fall.

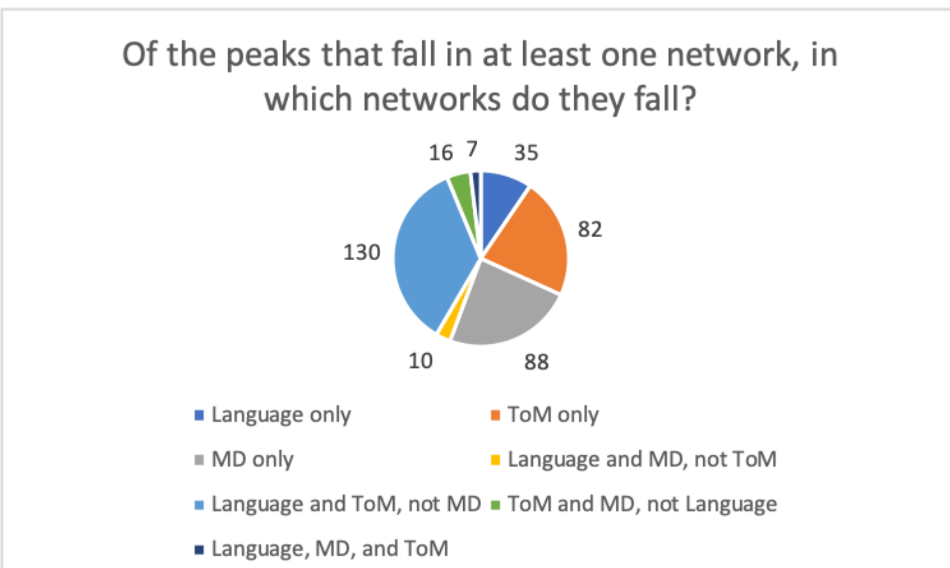

Figure 5. Breakdown of all relevant activation peaks. This figure illustrates the locations of the activation peaks that fell into at least one functional network.

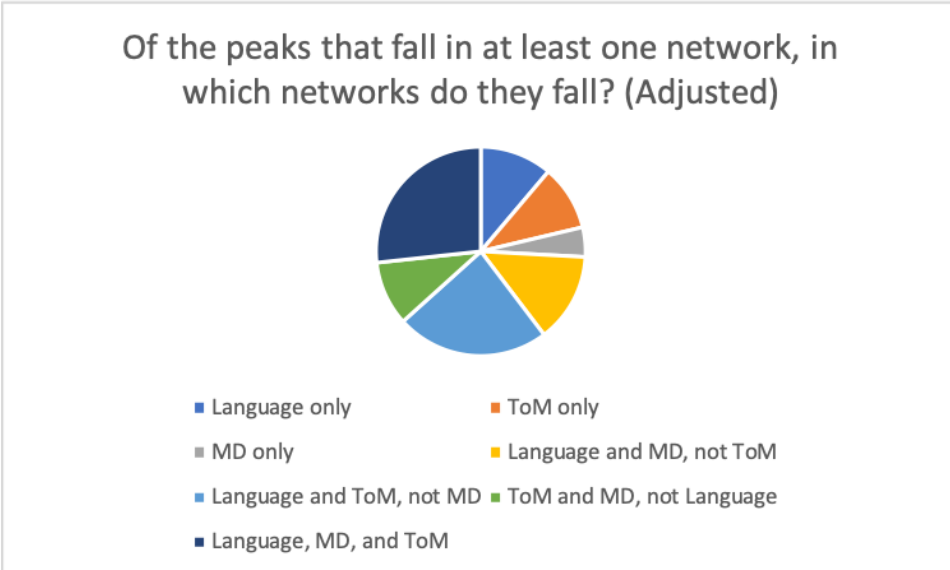

Figure 6. Breakdown of all relevant activation peaks, adjusted. This figure illustrates the locations of the activation peaks that fell into at least one network, adjusted for network size.

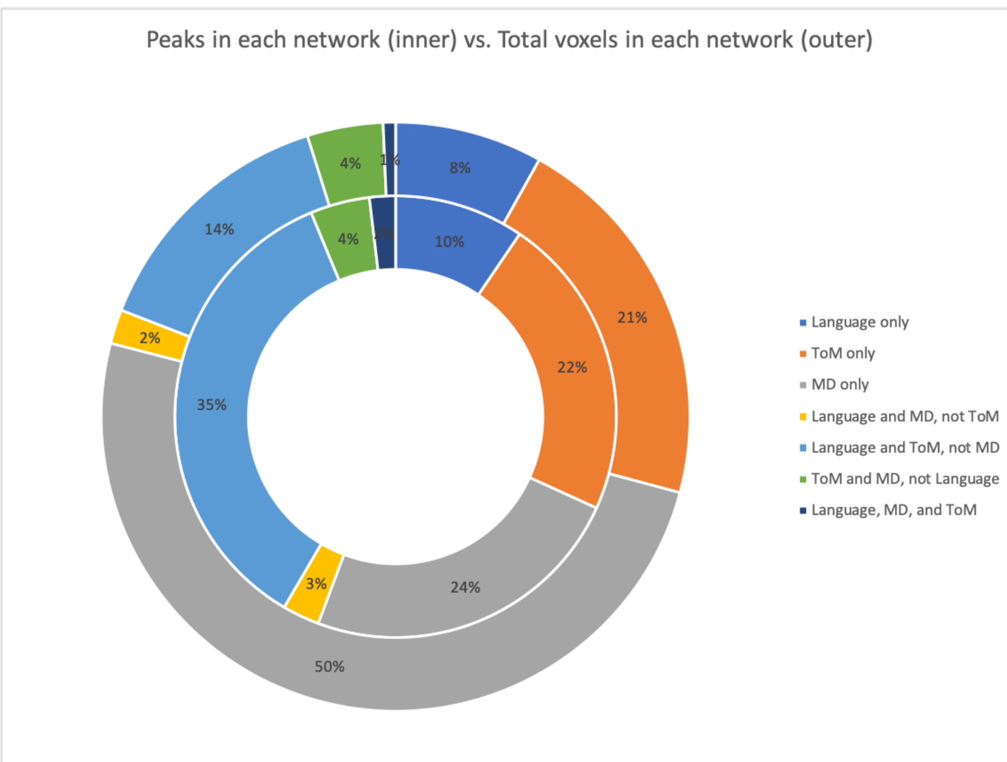

Figure 7. Breakdown of all relevant peaks vs. total voxels in each network. This figure compares the proportions of activation peaks and total voxels in each network.

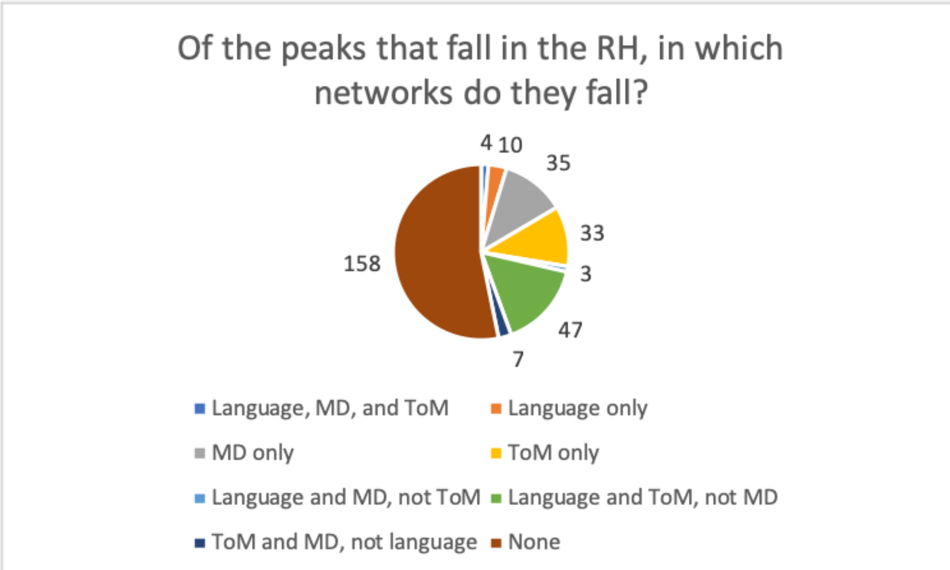

Figure 8. Breakdown of all RH activation peaks. This figure illustrates the networks into which all included RH activation peaks fall.

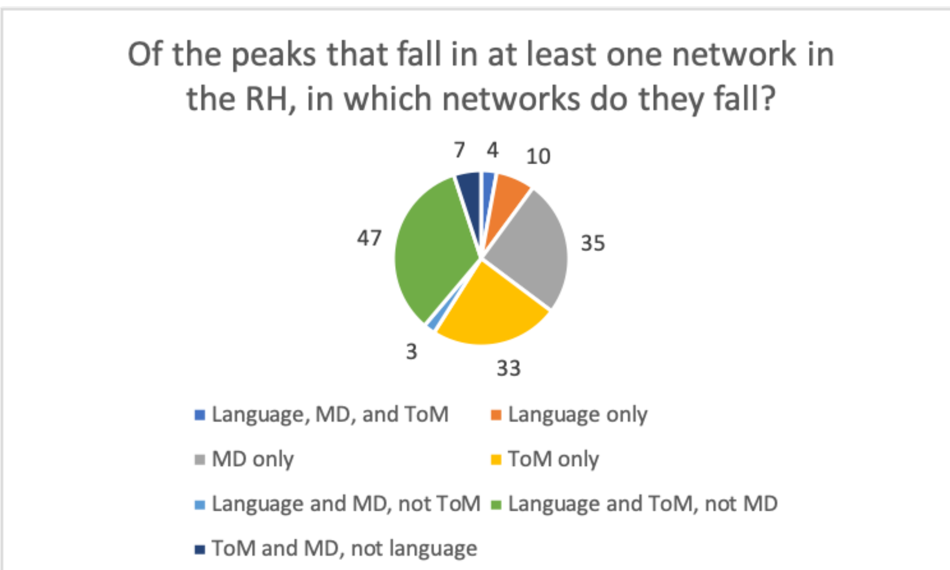

Figure 9. Breakdown of all relevant RH activation peaks. This figure illustrates the locations of the RH activation peaks that fell into at least one functional network.

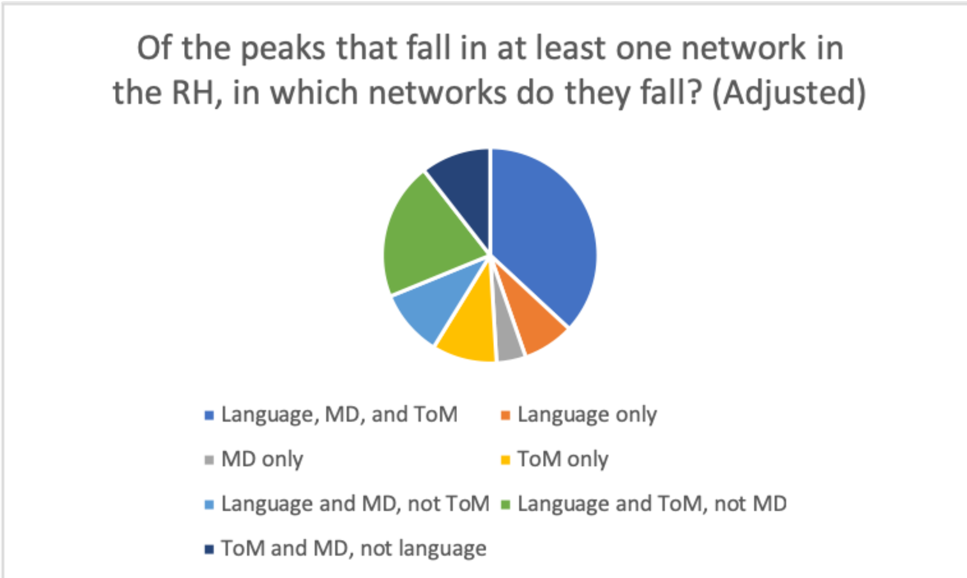

Figure 10. Breakdown of all relevant RH activation peaks, adjusted. This figure illustrates the locations of the RH activation peaks that fell into at least one network, adjusted for network size.

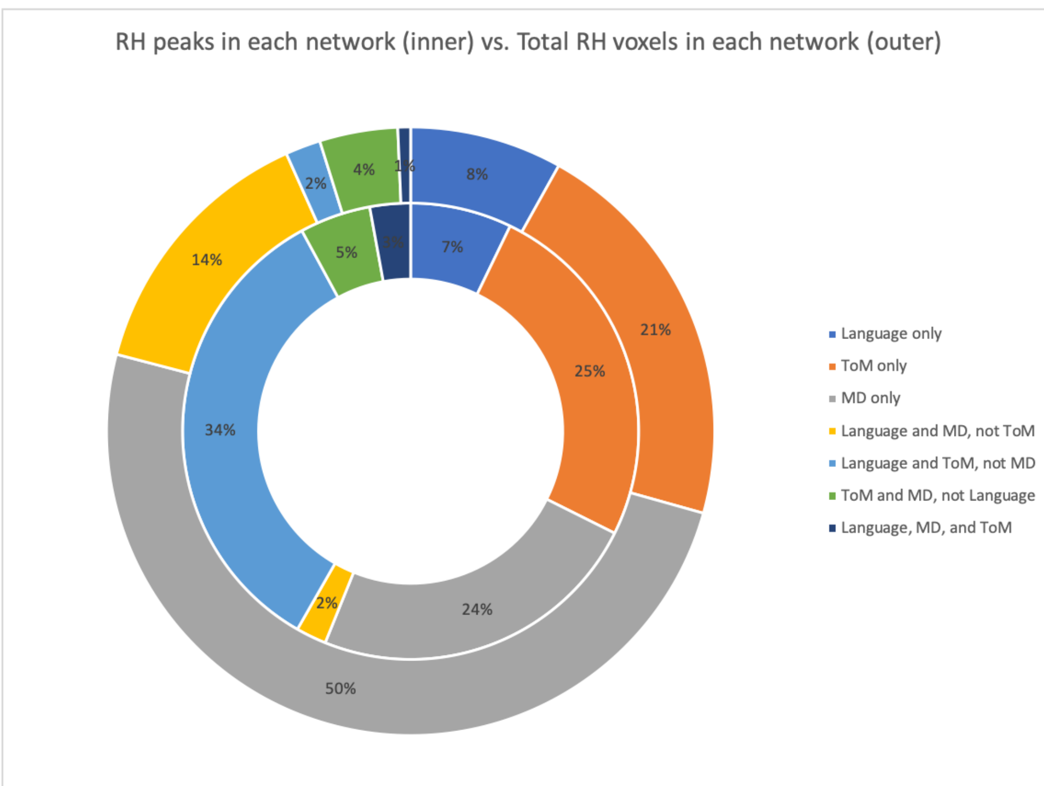

Figure 11. Breakdown of all relevant RH peaks vs. total RH voxels in each network. This figure compares the proportions of RH activation peaks and total RH voxels in each network.

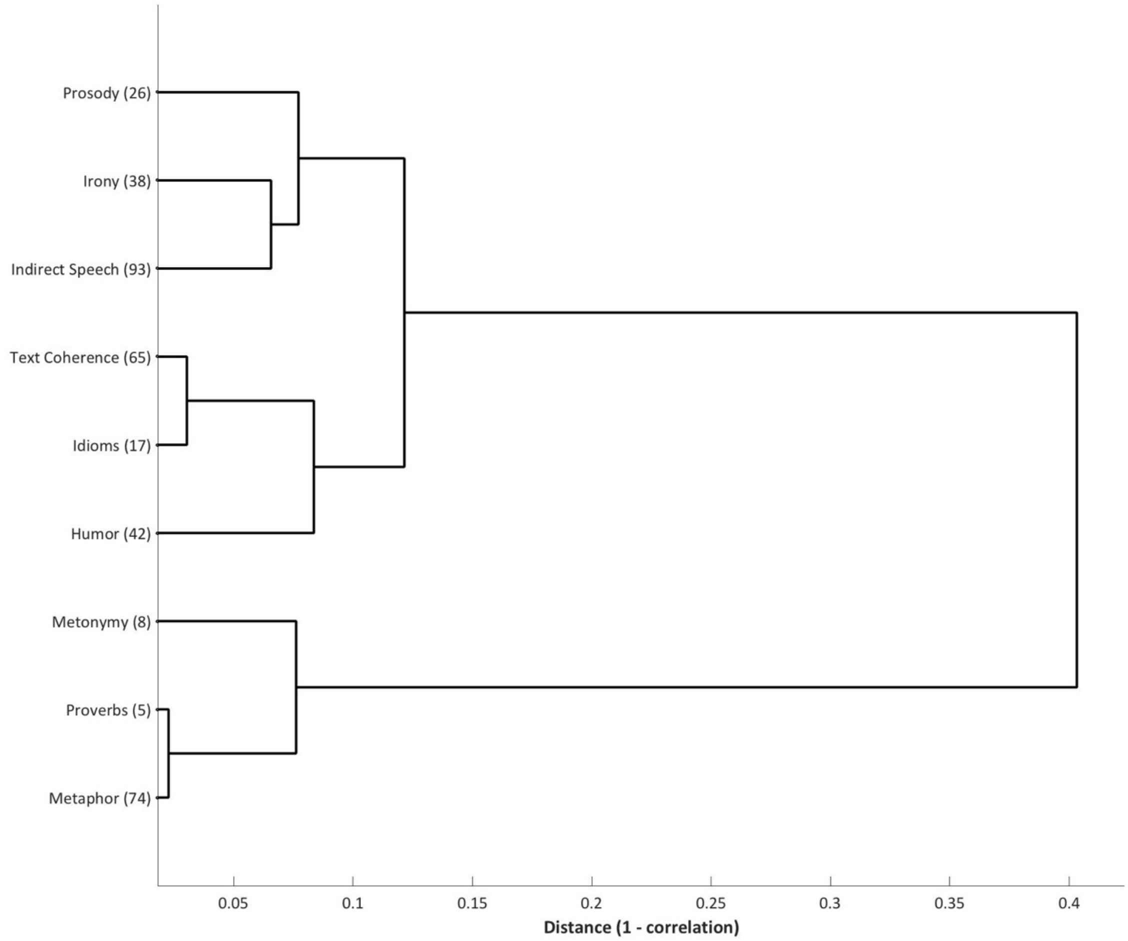

Figure 12. Hierarchical clustering of non-literal language types. This figure illustrates the results of the hierarchical clustering analysis.