Cerebral Cortex, 2024, **34**, 1–23
https://doi.org/10.1093/cercor/bhae032
Advance access publication date 3 February 2024
Original Article

Fractionating difficulty during sentence comprehension using functional neuroimaging

Malathi Thothathiri[1,*], Jana Basnakova[2], Ashley G. Lewis[2,3], Josephine M. Briand[1]

[1]Department of Speech, Language and Hearing Sciences, The George Washington University, Washington, DC, United States,
[2]Radboud University, Donders Institute for Brain, Cognition, and Behaviour, Nijmegen, The Netherlands,
[3]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
*Corresponding author: Department of Speech, Language and Hearing Sciences, The George Washington University, Washington, DC 20052.
Email: malathi@gwu.edu

Sentence comprehension is highly practiced and largely automatic, but this belies the complexity of the underlying processes. We used functional neuroimaging to investigate garden-path sentences that cause difficulty during comprehension, in order to unpack the different processes used to support sentence interpretation. By investigating garden-path and other types of sentences within the same individuals, we functionally profiled different regions within the temporal and frontal cortices in the left hemisphere. The results revealed that different aspects of comprehension difficulty are handled by left posterior temporal, left anterior temporal, ventral left frontal, and dorsal left frontal cortices. The functional profiles of these regions likely lie along a spectrum of specificity to generality, including language-specific processing of linguistic representations, more general conflict resolution processes operating over linguistic representations, and processes for handling difficulty in general. These findings suggest that difficulty is not unitary and that there is a role for a variety of linguistic and non-linguistic processes in supporting comprehension.

Key words: attention; cognitive control; left frontal cortex; syntactic ambiguity; working memory.

Introduction

The ability to construct and understand sentences (cf. individual sounds or symbols) is an essential feature that separates human language from other forms of human and non-human communication. Most sentences are not stored and reproduced in whole. Every day, we generate new sentences to communicate ideas and thoughts, and understand new sentences produced by others. Although highly practiced, sentence comprehension is a complex ability that is supported by multiple processes, such as retrieving words from the mental lexicon, building syntactic structures, and resolving conflict between competing representations. The role of these different processes can be determined by studying sentences that are more difficult to process than the simplest cases. So-called "garden-path" sentences have a long history in psycholinguistics and neurolinguistics because they allow researchers to separate different linguistic processes (syntax, semantics, pragmatics) and also to determine how broader functions like cognitive control and working memory assist comprehension (Just and Carpenter 1992; Christianson et al. 2001; Novick et al. 2005; den Ouden et al. 2016). Consider a garden-path sentence like (1):

(1) *As the men wrestled the rivals arrived at the gymnasium.*

As readers and listeners process the sentence, they initially tend to interpret "rivals" as the object of wrestling (i.e. that the men wrestled the rivals). However, upon encountering "arrived" and the subsequent words, the language comprehension system experiences conflict with this interpretation, which should trigger syntactic re-analysis and re-interpretation (i.e. that the men were wrestling someone else and not the rivals, who just arrived). This kind of tripping-up makes garden-path sentences more difficult than say, a sentence like "The cat sat on the mat." But what exactly is "difficulty"? Is it a unitary concept, or are different aspects of a difficult comprehension situation handled by different neural systems? In this paper, we address this issue by measuring neural activation for different types of sentences, which were designed to isolate the contribution of different processes, in the same set of individuals.

Prior literature has identified a key set of processes that could be recruited to different extents for different types of sentences. These include syntactic processing, semantic/pragmatic processing, cognitive control, and working memory. Historically, syntax was considered the predominant process that is used for sentence comprehension because the meaning of a sentence ultimately depends on its syntactic structure. Syntax allows us to differentiate between "The dog chased the cat" and "The cat chased the dog" despite the two sentences containing the same words. However, several studies have now shown that semantics and pragmatics can also exert a powerful influence on sentence interpretation, sometimes even overriding the meaning indicated by syntax (see e.g. Altmann and Kamide 1999; Ferreira 2003; Kim and Osterhout 2005). For example, a sentence like "The dog was bitten by the man" may be interpreted instead as the dog biting the man due to our semantic and pragmatic knowledge about the world. Cognitive control may be relevant for supporting accurate comprehension especially in demanding situations such as ambiguous sentences, conflicting interpretations, and potentially noisy input conditions (Novick et al. 2005; Thothathiri et al. 2012a; Peelle and Wingfield 2022). Working memory is relevant for integrating

Received: August 14, 2023. **Revised:** January 15, 2024. **Accepted:** January 16, 2024

incoming words into a cohesive structural interpretation (Just and Carpenter 1992; Lewis et al. 2006).

In functional neuroimaging (fMRI) studies, contrasts between different conditions that vary on a particular dimension are used to identify the neural substrates associated with a particular process. For example, syntactically complex sentences may be compared to syntactically simpler sentences in an attempt to locate regions that are relevant to syntactic processing. However, herein lay a challenge. Sentences that are difficult along one dimension are also usually difficult along other dimensions. For example, compare the garden-path sentence (1) above to a non-garden-path sentence like (2):

(2) *While the zookeeper fed the ponies the stallion stomped its hoof.*

For (1), syntactic re-analysis is needed in order to arrive at the correct interpretation (i.e. that the men did not necessarily wrestle the rivals). In contrast, in (2), the structure is unambiguous—it is clear that the zookeeper fed the ponies and not the stallion, and no re-analysis is needed. Thus, a contrast between (1, 2) might be expected to identify regions involved in syntactic analysis. However, this is not the only possible differential process in this comparison: (1) may also trigger additional semantic/pragmatic processing to reconcile the different conflicting interpretations, cognitive control to resolve conflict, and working memory to re-process the sentence. Separating these different dimensions of difficulty is important for understanding the various subcomponents of sentence processing and identifying the neural substrates of each component.

We used a functional profiling approach to fractionate the processing of difficult-to-comprehend sentences. Specifically, we examined how regions of interest (ROIs) implicated in processing garden-path versus non-garden-path sentences are activated for *other* sentences that vary in their syntactic re-analysis, semantic/pragmatic processing, cognitive control, and working memory demands. Below, we briefly review prior literature on the processes used for sentence comprehension in general and the literature on garden-path sentences in particular before describing the design of the present study and its contributions.

Prior literature on sentence comprehension processes

Sentences are composed of multiple words connected together by structural or syntactic rules. Accordingly, early neurolinguistic studies focused on identifying the locus of syntactic operations in the brain (see e.g. Caplan et al. 2002 and Dapretto and Bookheimer (1999) for a summary). Typical experimental designs contrasted sentences with different syntactic complexity, sentences containing syntactic versus other kinds of errors, or tasks that emphasized syntactic versus other kinds of processing (Caplan et al. 2002;Dapretto and Bookheimer 1999 ; Embick et al. 2000). Many studies pointed towards a role for the left frontal cortex, especially Brodmann areas 44 and/or 45, collectively known as Broca's area (Dapretto and Bookheimer 1999; Embick et al. 2000). However, some early as well as more recent studies have documented the involvement of left posterior temporal and inferior parietal regions (Caplan et al. 2002; Thothathiri et al. 2012b; Wartenburger et al. 2004; Yokoyama et al. 2007). Damage to the temporo-parietal and not the frontal cortex predicts syntactic comprehension deficits in aphasia (Thothathiri et al. 2012b. See also Fridriksson et al. 2018). Based on a meta-analysis of more than 35 neuroimaging (fMRI/PET) studies of sentence comprehension, Walenski et al. (2019) found evidence for an association between syntactic processing and both left frontal and left posterior temporal regions. Overall, there is growing consensus for the posterior temporal lobe's involvement in syntactic processing during comprehension. For the frontal lobe, there is debate about whether its role may be best described as being more specific to sentence production than comprehension or broader resources like cognitive control or working memory (Matchin et al. 2020; Thothathiri et al. 2012a, 2012b; Walenski et al. 2019. See also Diachek et al. 2020).

Sentence comprehension is ultimately about understanding meaning. The meanings of individual words must be combined to compute the compositional meaning of the sentence. In contrast to the syntactic processing that is tied to more posterior temporal regions, semantic composition has been linked to the left anterior temporal lobe (ATL). Left ATL has been linked to semantic memory based on word-level evidence from semantic dementia, compositional processing at the level of phrases and sentences, and to interactions between word-level and phrasal-level information (Lambon Ralph et al. 2010; Westerlund and Pylkkänen 2014). Although damage to this region is not routinely associated with sentence comprehension deficits, its involvement in semantic processing could make it relevant for language comprehension under some conditions (Walenski et al. 2019).

Sentence comprehension difficulty is not only about challenges in syntactic or semantic processing but could also be about broader cognitive resources like cognitive control and working memory. Cognitive control, or the ability to resolve conflict between competing representations, has been argued to be relevant especially for more difficult-to-comprehend sentences. Garden-path sentences like (1) above create conflict between the original interpretation (e.g. that the men wrestled the rivals) and other information from the sentence (e.g. that the rivals just arrived) and cognitive control can be useful for resolving such conflict. Therefore, at least some difficult sentences could be difficult due to their cognitive control demands. Sentence processing also requires storing and building structured representations from words as the sentence unfolds incrementally i.e. working memory (Lewis et al. 2006; Shain et al. 2022). For example, understanding a sentence like "We sang a song that our daughter really likes" requires holding on to the word "song" until it can be linked to the verb "likes" at the end (cf. "Our daughter really likes a song that we sang"). Thus, syntactically difficult sentences may require more working memory resources than their simpler counterparts. Prior studies have provided broad support for the role of both cognitive control and working memory in sentence comprehension (Cognitive control as defined here is distinct from working memory (Friedman and Miyake 2017). The garden-path stimuli in the present study involved ambiguity and conflict and required conflict resolution but the working memory stimuli did not. Instead, the latter sentence types involved routine sentence processing operations that required holding in memory and binding together words in a syntactic dependency. Some complex working memory tasks used in the literature (e.g. N-back with interference) involve cognitive control in addition to working memory because they create conflict but this does not apply to the working memory stimuli used here.), based on neuroanatomical co-localization of those functions with sentence processing and causal links between damage to or upregulation of those abilities and better sentence comprehension (e.g. Ye and Zhou 2009; Vuong and Martin 2015; Hsu et al. 2017; Thothathiri et al. 2018; Horne et al. 2022). The left frontal cortex has been linked to both cognitive control and working memory, as part of networks working in tandem with temporal, parietal, and medial frontal regions (e.g. see Botvinick et al. 2001; Hsu et al. 2017; Shain et al. 2022).

Different neurolinguistic frameworks of language processing all accord a role for the processes discussed above in sentence comprehension. But they differ in the weighting allocated to different components. Hagoort (2005)'s Memory, Unification and Control (MUC) model proposes that language processing utilizes representations stored in memory in the left temporal cortex. Unification of different linguistic representations is hypothesized to be coordinated by the left inferior frontal cortex. Last but not least, control operations in dorsolateral and medial frontal cortices support goal-directed use of language in different contexts. The MUC model explicitly allows for interactive, concurrent processing of different sources of information (e.g. syntactic, semantic and pragmatic). It does not prioritize syntax. It also allows for a role for domain-general control operations in language use. By comparison, Friederici (2002)'s neuroanatomical model of sentence processing proposes a syntax-first view wherein syntactic processes precede semantic processes initially and the two interact only during later phases. Semantic representations are hypothesized to be stored in the temporal cortex with the frontal cortex supporting controlled or strategic use of those representations. For syntax, the temporal cortex and the most inferior parts of the frontal cortex are thought to be relevant for syntactic operations with other less inferior parts of the frontal cortex engaged for working memory.

More recently, Fedorenko and colleagues have used a single-subject localization approach to argue that most of language processing occurs within a bilateral temporo-frontal language-selective network. They distinguish some lateral frontal areas within this network from other nearby lateral frontal regions that they argue are part of a more domain-general Multiple Demand (MD) network, which is not specific to language processing (Fedorenko et al. 2012; Diachek et al. 2020; Shain et al. 2022). Although many language tasks activate both the language and the MD networks, these authors have suggested that the latter is primarily engaged when there is an explicit secondary task going beyond passive sentence comprehension (Diachek et al. 2020). Thus, left frontal regions in this framework are split between those engaged in language-specific operations within the language network and those involved in more domain-general processes within the MD network.

To summarize, converging evidence in the field suggests that both temporal and frontal regions within the left hemisphere are involved in sentence comprehension. Temporal regions are widely thought to store the linguistic representations—syntactic and semantic—that is relevant for processing sentences. The precise role of the left frontal cortex is more debated. There is consensus that the more dorsal portions are involved in domain-general processes that support both linguistic and non-linguistic tasks. However, the more ventral/inferior portions have been linked variously to unification of syntactic and non-syntactic representations, syntax-specific operations, working memory, and cognitive control (Friederici 2002; Fiebach et al. 2005; Hagoort 2005; Thothathiri et al. 2012a; Hsu et al. 2017).

Prior literature on garden-path sentences

For garden-path sentences like (1), prior electrophysiological (ERP) studies have demonstrated a P600 signal at the point of disambiguation. Qian et al. (2018) found a P600 effect for garden-path versus non-garden-path sentences in healthy adults. Sheppard et al. (2017) contrasted garden-path sentences with plausible versus implausible noun phrases (e.g. *While the band played the song/beer . . .*) and found an N400-P600 complex for the latter compared to the former in healthy adults and patients with anomic aphasia. In ERP studies, the P600 signal is seen for a range of sentence stimuli, including those with syntactic violations, thematic role assignment errors, or even more broadly, any violation of expectations during sentence comprehension. This has led some to argue that it reflects conflict monitoring (Van De Meerendonk et al. 2010). Together, a P600 effect during the processing of sentences similar to (1) confirms that the comprehension system can utilize disambiguating syntactic or semantic information to detect conflict between interpretations and undertake structural reanalysis relatively quickly. But it does not clarify the various processes used to understand these sentences and their neural correlates.

Two previous fMRI studies have examined the neural correlates of comprehending garden-path sentences like (1). Hsu et al. (2017) compared activation for reading garden-path sentences (e.g. "While the thief hid the jewelry sparkled") and their non-garden-path counterparts (e.g. "While the thief hid, the jewelry sparkled" with the critical comma). They found that the top 100 voxels activated for this contrast in the left inferior frontal cortex were also activated for three other cognitive control tasks (Stroop, n-back with lures and recent negatives) and vice versa. This co-localization across tasks indicated that the left inferior frontal cortex was a general-purpose hub for cognitive control. At the same time, functional connectivity analyses revealed distinct connected areas for different tasks, indicating that specialization might arise at the network level rather than due to specialization within the frontal cortex per se. den Ouden et al. (2016) conducted whole-brain and ROI analyses of garden-path and non-garden-path sentences presented auditorily. They found evidence suggesting a key role for the posterior temporal cortex (cf. frontal or anterior temporal cortex) in processing syntactic ambiguity. Within the frontal cortex, more fine-grained ROI analyses detected differences between sub-regions such that Brodmann area 45 responded only to sentences where re-analysis occurred and more posterior-dorsal Brodmann area 44 responded to all sentences with ambiguity/conflict. Overall, the findings accorded with prior literature (see above) in indicating a central role of the posterior temporal cortex in syntactic processing. For the frontal cortex, the effects were subtler and limited to cases of ambiguity or reanalysis.

Neither of the previous studies was designed to fractionate difficulty during sentence comprehension. den Ouden et al. (2016) focused on the interaction between prosodic and plausibility cues during auditory sentence comprehension. Hsu et al. (2017) did conduct a secondary analysis of MD network regions to distinguish between cognitive control and broad difficulty that bears on the interpretation of our results (see Discussion). However, their emphasis was on domain-generality rather than on understanding the set of processes used to comprehend difficult garden-path sentences. The present study was designed to address this gap and answer questions about whether difficulty in the context of sentence comprehension is handled by a variety of regions that support handling different aspects of that difficulty.

The present study

We sought to fractionate the difficulty associated with hard-to-comprehend garden-path sentences by comparing activation for different kinds of difficult sentences within the same regions of interest (ROIs). We used a sentence reading plus comprehension question task with five different types of sentences, including (1, 2), reproduced from above. The first three sentence types contained a subordinate clause followed by a main clause. They were used to contrast garden-path and non-garden-path sentences (1 vs 2; 3 vs 2) and also two different variants of garden-path sentences (1 vs 3) in order to understand how the brain handles ambiguity and conflict during sentence comprehension. The last two sentence types (4 and 5) manipulated difficulty along a different dimension, namely working memory. They were included to help separate the neural substrates of processes that handle ambiguity and conflict (e.g. cognitive control for resolving conflict) from those underlying working memory. Below, we first discuss sentence types (1–3) and then (4, 5).

(1) Implausible: *As the men wrestled the rivals arrived at the gymnasium*.
(2) Control: *While the zookeeper fed the ponies the stallion stomped its hoof*.
(3) Plausible: *While the farmer steered the green tractor pulled the plow*.
(4) Long WM: *Ezekiel bragged about his circus art skills and recorded himself playing a Beethoven sonata on the piano*.
(5) Short WM: *Natalie and Margaret called themselves an overpriced cab and visited the natural history museum*.

Control sentences like (2) did not contain syntactic ambiguity or conflict. By contrast, in sentences like (1), there was syntactic ambiguity before "arrived" because "the rivals" could be either the end of the subordinate clause that started with "As the men" or the beginning of a new main clause that started with "the rivals." This should generate conflict because the parser originally constructs an analysis consistent with the former but then encounters "arrived," which is inconsistent with that analysis. The same kind of ambiguity and conflict also existed in sentences like (3). Thus, comparisons between garden-path sentences like (1, 3) versus non-garden-path sentences like (2) can identify regions involved in cognitive control for resolving conflict arising from temporary syntactic ambiguity, and syntactic analysis for reanalyzing the sentence.

The labeling of the garden-path conditions—(1) and (3)—is based on the plausibility of the original misinterpretation once the structure is disambiguated. Specifically, the key difference between (1, 3) was that the ending of the sentence in (1)—that the rivals arrived at the gymnasium—made the original misinterpretation—that the men wrestled the rivals—implausible. Therefore, these sentences were labeled "Implausible." In contrast, for (3), the ending of the sentence—that the green tractor pulled the plow—left the original interpretation—that the farmer steered the green tractor—still semantically plausible. Therefore, these sentences were labeled "Plausible." A number of previous studies have employed Implausible and Plausible sentences and shown that comprehenders are more likely to retain the original misinterpretation and say Yes to questions like "Did the farmer steer the tractor?" in the Plausible condition than to questions like "Did the men wrestle the rivals?" in the Implausible condition (see e.g. Christianson et al. 2001, 2006). Using terminology employed in this literature, Plausible sentences like (3) are more subject to "lingering misinterpretation" than Implausible sentences like (1) (Christianson et al. 2001; Slattery et al. 2013). Put the other way, comprehenders are more successful in letting go of the original misinterpretation for Implausible than Plausible sentences, suggesting that the former recruits some additional process or processes that can help reconcile incompatible interpretations.

Some of the sunglasses shown here are to the left of the lamp.

Fig. 1. Example image for the visual attention task.

Thus, differential activation for the Implausible versus Plausible condition can identify regions involved in the additional process(es).

As mentioned above, we also included two other types of sentences that were both long and could induce comprehension difficulty along a different dimension, namely working memory. These sentences (Long and Short WM) contained reflexive pronouns (herself/himself/themselves) that had to be bound to an antecedent for comprehension (***Ezekiel*** *bragged about ... and recorded* ***himself***). The distance separating the antecedent from the reflexive pronoun was longer for the long than the short WM condition. Importantly, while these sentences were expected to induce difficulty related to working memory demands, they did not contain ambiguity or conflict and were therefore not expected to recruit cognitive control or syntactic analysis regions more than the Control condition.

Additionally, we tested sentence reading in a non-linguistic context that manipulated visual attentional demands. Participants saw a sentence below a visual stimulus and were asked to indicate whether the sentence accurately described the image (Fig. 1).

(6) More/Less: *Less than half of the pigs are to the right of the lamp.*
(7) All/Some: *Some of the sunglasses shown here are to the left of the lamp.*

In the More/Less condition, the sentences contained the quantifiers "More than half" or "Less than half". In the All/Some condition, the sentences contained the quantifiers "All" or "Some". Majority quantifiers like "more than half" and "less than half" require counting or estimating the total number in the set, computing half of that number, and comparing that threshold to the number of items on a given side. In contrast, trials with logical quantifiers like "all" and "some" are easier to process because they only require assessing if all of the items are to one side or if they are present on both sides. Specifically, in the current study, on "All" trials, if the items were present on both sides of the lamp, the correct answer was No. Conversely, if they were present on only one side, the correct answer was always Yes. On "Some" trials, if the items were present on both sides, the correct answer was Yes. If they were present on only one side, the correct answer was always No. Prior behavioral evidence is consistent with the intuition that More/Less trials are harder than All/Some trials—participants are less accurate and slower to respond on the former than the latter (Olm et al. 2014).

Neurally, we predicted that more/less would involve more visual processing (in the visual cortex), number processing (in the parietal lobe), and cognitive resources related to increased task difficulty (in dorsolateral frontal cortex) (Olm et al. 2014;

Table 1. Logic of the functional profiling approach. Multiple hypothesized processes for implausible > control can be fractionated based on which other contrasts they are activated for.

Implausible > Control A-priori hypothesized process	Region activated for Implausible > Control should also be activated for:
Syntactic reanalysis	*Plausible > Control*
Language-specific cognitive control	*Plausible > Control*
Linguistic working memory	*Plausible > Control*
	Long WM > Control, Long WM > Short WM
Semantic/pragmatic processing	*None*
Domain-general cognitive resources, including non-language-specific cognitive control and working memory	*Plausible > Control*
	More/Less > All/Some

Heim et al. 2016). Importantly, the task involved a significant non-linguistic (visual) component and the sentences in the two conditions had similar linguistic properties, including sentence length, syntactic structure, and semantic processing demands. Thus, the More/Less condition was expected to be more difficult than All/Some but this difficulty is due to demands on visual attention and domain-general cognitive resources for doing harder tasks (e.g. general-purpose cognitive control or working memory supported by the MD network) rather than language-specific processes (including potentially language-specific cognitive control or working memory operations in Fedorenko and colleagues' language network. See e.g. Shain et al. 2022).

To summarize, we attempted to separate different sources of difficulty that can make comprehension challenging when sentences are not simple or straightforward. Garden-path sentences can potentially trigger syntactic reanalysis, language-specific cognitive control, linguistic working memory, semantic/pragmatic processing, and domain-general cognitive resources (including non-language-specific cognitive control and working memory) relative to syntactically unambiguous sentences. Therefore, we tested the comprehension of sentences that tap these different resources to different extents, in the same set of individuals. Our analyses examined the regions activated for the Implausible condition over the Control condition and the functional profile of those regions (i.e. which other conditions those regions were activated for. See Table 1). We reasoned that regions involved in syntactic reanalysis, language-specific cognitive control and domain-general task demands would be activated for Plausible sentences—in addition to Implausible sentences—because these sentences contain ambiguity and generate conflict between interpretations. In contrast, regions involved in semantic/pragmatic processing should be selectively activated for Implausible but not Plausible sentences relative to Control because only the former requires handling the implausibility of maintaining the original (mis) interpretation after encountering the information at the end of the sentence. If regions involved in linguistic working memory, separately from those engaged in cognitive control, are used to deal with garden-pathing, we should see increased recruitment for Implausible and Plausible over Control but also for the Long WM condition. The Long WM condition involves binding a reflexive pronoun to a distant antecedent, which is thought to require working memory (Dillon 2014). Therefore, we should observe increased activation in such regions for Long WM over Control and more stringently, for Long WM over Short WM. Finally, if there are regions that assist sentence comprehension via a broad response to task difficulty, including domain-general, non-language-specific cognitive control and working memory, we should see activation for Implausible and Plausible sentences and additionally, for More/Less > All/Some in the visual attention task in those specific regions. Together, the functional profile of recruited regions across the different sentence types can help fractionate the neural correlates into different processes that can support the comprehension of difficult-to-understand sentences.

Materials and methods

Norming study

We normed the stimuli to be used in the study prior to collecting data. Thirty-seven adults (15 Male, 15 Female, 21–36 years, Median age = 31 (Due to a technical error, demographic information was lost for 7 adults. The reported numbers are for 30 out of 37 participants.)) participated in a web-based experiment. Participants were recruited using Prolific, a web-based platform for conducting research. Participants self-reported being right- handed, native English speakers. Prolific collects demographic information independently of specific studies, minimizing concerns about misrepresentation by the participants. Stimuli were presented using Python on Pavlovia (https://pavlovia.org).

Participants completed two tasks. In the sentence reading task, they read 42 sentences in each of the five conditions (Control, Implausible, Plausible, LongWM, ShortWM) word by word and answered yes/no comprehension questions. In the visual attention task, they saw a picture and read the sentence below it to answer whether the sentence accurately described the picture. There were 42 trials in the More/Less condition (21 More, 21 Less) and 42 trials in the All/Some condition (21 All, 21 Some). For both tasks, they used the "j" key for a yes and the "k" key for a no response. Participants were assigned to one of two lists, which differed only in the order of trials (List 2 was the reverse of List1).

For the sentences with subordinate clauses (Implausible, Plausible, Control), the comprehension question tested the thematic role assigned to the noun phrase following the verb. For garden-path Implausible and Plausible sentences, the correct answer was No (e.g. Sentence: *As the men wrestled the rivals arrived at the gymnasium;* Question: *Did the men wrestle the rivals?*). This is because the sentence structure indicates that "the rivals" are the subject of arriving and not the object of wrestling. The proportion of "Yes" responses is commonly used to infer the extent to which comprehenders hold on to or let go of the original misinterpretation (see e.g. Christianson et al. 2001). To partially balance out the type of correct response, for non-garden-path Control sentences, the correct answer was Yes. In the working memory conditions, the comprehension question tested whether participants remembered various details from the content of the sentences. Correct responses were split evenly between Yes and No.

For sentences with subordinate clauses, 42 verbs appeared once each in Implausible, Plausible and Control conditions, within the subordinate clause. The main clauses of the sentences contained 94 different verbs that appeared between 1–3 times. Within each condition, half of the sentences began with "While" and half with "As" (21 each). Sentences in the three conditions had similar length (number of characters) (Implausible Mean = 59.1, SD = 4.7; Plausible Mean = 57.6, SD = 4.5; Control Mean = 59, SD = 7.6). Twenty-two proper names (e.g. Kendra, Sam) were used once each in each condition. Implausible and Plausible sentences were further matched such that they had the same subordinate clause verbs and had overlapping lexical frequencies of the subordinate clause nouns, main clause verbs and main clause nouns (Corpus of Contemporary American English, Davies 2008. Subordinate clause Noun: Implausible Mean = 81,327, SD = 104,375; Plausible Mean = 73,553, SD = 169,493. Main clause verb: Implausible Mean = 79,536, SD = 105,548; Plausible Mean = 70,444, SD = 130,756. Main clause Noun1: Implausible Mean = 39,292, SD = 66,760; Plausible Mean = 55,844, SD = 91,346. Main clause Noun2: Implausible Mean = 91,092, SD = 157,642; Plausible Mean = 52,769, SD = 75,203).(There is debate about whether null-hypothesis significance tests of confound variables are appropriate to conduct and interpret (Sassenhagen and Alday 2016). For informational purposes, pairwise 2-tailed t-tests comparing the lengths of Implausible, Plausible and Control sentences did not reveal any significant differences (p's > 0.05). Frequency comparisons between Implausible and Plausible sentences also did not reveal any significant differences (p's > 0.05).) For the working memory stimuli, the Long and Short WM conditions contained the same content with the two coordinated clauses in alternate orders (e.g. Long WM: *Josh went to see a movie and bought himself snacks*; Short WM: *Luke bought himself snacks and went to see a movie*).

In the visual attention task, each image featured a lamp in the center of the screen and various objects to the left and/or right of the lamp. Forty-two nouns appeared once each in the More/Less and All/Some conditions. The correct answer was split evenly between Yes and No within each condition.

For the sentence reading task, we confirmed that the Plausible sentences were more subject to lingering misinterpretation than the Implausible sentences (Plausible mean accuracy = 59.5% (SD = 54.9, SE = 1.4, 95% CI = 2.7). Implausible mean accuracy = 79.6% (SD = 45.1, SE = 1.1, 95% CI = 2.2). z = −5.13, $P < 0.001$). Accuracies for Long WM and Short WM sentences were at ceiling and did not differ from one another (Long WM: Mean = 97% (SD = 19.2, SE = 0.5, 95% CI = 1.0); Short WM: Mean = 96.7% (SD = 20.1, SE = 0.5, 95% CI = 1.0). z = 0.56, $P = 0.57$). For the Control condition, accuracy was 89.5% (SD = 34.3, SE = 0.9, 95% CI = 1.7). (Here and below, we compared the question response accuracies and reaction times of closely matched conditions that had similar questions and response types. We did not compare Implausible, Plausible and other sentence types to Control because the correct responses were always Yes for Control sentences, always No for garden-path sentences, and split evenly for working memory stimuli, as explained above.) For the visual attention task, we confirmed that the More/Less condition was less accurate than the All/Some condition (More/Less mean accuracy = 85% (SD = 50.5, SE = 1.3, 95% CI = 2.5). All/Some mean accuracy = 95.9% (SD = 28.1, SE = 0.7, 95% CI = 1.4). z = −3.82, $P < 0.001$). Here and below, standard deviation and other summary variables are within-subjects estimates obtained using summary SE within (Rmisc package, version 1.5.1).

For the sentence reading task, linear mixed modeling of raw reaction times for answering the questions did not yield residuals that were normally distributed, so we modeled log-transformed reaction times. Outliers were removed using median absolute deviation (outliers_mad function in ROutliers, version 0.0.0.3. Parameters b = 1.4826, threshold = 3 for normal distribution). Reaction times were slower for the Plausible than the Implausible condition (Mean log reaction time for Plausible = 0.37 (SD = 0.55, SE = 0.02, 95% CI = 0.04); Implausible = 0.29 (SD = 0.52, SE = 0.02, 95% CI = 0.03); t(41.1) = 3.87, $P < 0.001$). For Long versus Short WM, the difference was not significant (Mean log reaction time for Long WM = 0.30 (SD = 0.40, SE = 0.01, 95% CI = 0.02); Short WM = 0.28 (SD = 0.39, SE = 0.01, 95% CI = 0.02); t(40.64) = 0.7, $P = 0.49$). For control sentences, mean log reaction time was 0.34 (SD = 0.43, SE = 0.01, 95% CI = 0.02). For the visual attention task, we log-transformed the reaction times and removed outliers using the same procedure as above. Reaction times were slower for the More/Less than the All/Some condition (Mean log reaction time for More/Less = 0.88 (SD = 0.61, SE = 0.02, 95% CI = 0.03); All/Some = 0.65 (SD = 0.63, SE = 0.02, 95% CI = 0.03); t(37.29) = 10.21, $P < 0.001$).

Overall, the norming study confirmed that the Plausible condition was more susceptible to lingering misinterpretation than the Implausible condition, yielding less accurate and slower responses to the questions. For the visual attention task, the results confirmed that the More/Less condition was more difficult than the All/Some condition, with less accurate and slower responses. We did not find significant effects for the Long vs Short WM contrast, suggesting that this manipulation was subtle and may potentially yield only weak brain activation differences.

fMRI study

Participants

For the fMRI experiment, we recruited participants through flyers and online postings on the George Washington University's Psychology research credit portal. Thirty-two participants (8 Male, 24 Female, 18–28 years, Median age = 19) participated. All were right-handed, native English speakers with normal or corrected to normal vision from the Washington, DC, area. They self-reported no history of neurological disorders or brain injury or use of neuropsychiatric medications. All participants underwent MRI safety screening and provided written informed consent under a protocol approved by the George Washington University Institutional Review Board. They received $25 or course credit for their participation. Three participants were excluded—one did not complete all the imaging runs, one had poor accuracy (<50% across all conditions), and one belatedly reported an ADHD diagnosis—leaving 29 participants in the final analyses.

Materials and procedure

Before entering the scanner, participants were familiarized with the two tasks that they would be doing during the experiment. For the sentence reading task, participants first received eight practice trials with feedback. They repeated this practice until they got at least seven out of the eight trials correct (i.e. answered the comprehension question correctly). Subsequently, they received 10 practice trials with no feedback, resembling the task structure inside the scanner. Practice sentences repeated during different phases of the practice but did not appear during the experiment. For the visual attention task, participants first received eight practice trials with feedback. They repeated this practice until they got at least seven out of the eight trials correct. Subsequently, they received four practice trials with no feedback, resembling

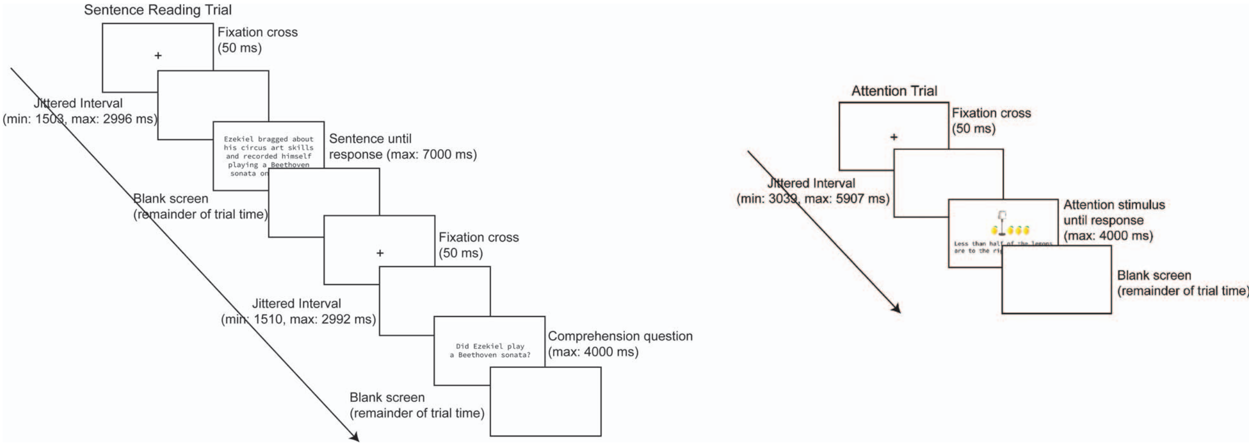

Fig. 2. Trial structure in the sentence reading task (left) and the visual attention task (right).

the task structure inside the scanner. The sentences used in the practice trials did not appear during the experiment.

Inside the scanner, participants completed two runs each of the sentence reading and visual attention tasks, respectively. The order of the runs was counterbalanced, with half the participants completing the sentence reading task during the first and the third runs and the visual attention task during the second and the fourth, and the other half doing the reverse. Stimuli were the same as those used in the norming study or contained minor changes (see below). All stimuli were presented using E-Prime 2.0.

Sentence reading task

During the sentence reading task, participants silently read whole sentences that were presented one at a time on a computer screen. After reading each sentence, they answered a comprehension question. All text appeared in black font on a white background. Responses were made using the index, middle, and ring fingers of the right hand on the left, middle, and right buttons, respectively. Participants were asked to press the middle button to indicate completion of reading. Subsequently, they answered the yes/no comprehension question using the left button for "Yes" or the right button for "No."

We used the same Implausible, Plausible, and Control sentences as in the norming study. For Long and Short WM, we modified the sentences to be longer because the norming data suggested that the working memory manipulation might be weak (e.g. *Josh went to see a movie and bought himself snacks* ➔ *Josh went to see a foreign film and bought himself a variety of snacks*). The 210 sentences were split across two runs. Each run contained 105 sentences (21 in each condition) with their corresponding comprehension questions. For Implausible and Plausible sentences, the correct answer was Yes. For Control sentences, the correct answer was No. Sentences and questions were separated by variable jitters and modeled separately (see more below). We draw inferences based on neural activation for the sentences only. Therefore, differences in correct responses to the questions do not impact interpretation of the relevant fMRI results. In the working memory conditions, the comprehension question tested whether participants remembered various details from the content of the sentences. Responses within each condition (Long or Short WM) and run were split roughly evenly between Yes and No (11 vs. 10 out of 21). See Appendix for a full list of sentences and questions. Trials were pseudorandomly ordered such that the same condition did not occur more than two times in a row and the correct answer was not the same more than three times in a row.

For sentences with subordinate clauses, 42 verbs appeared once each in Implausible, Plausible and Control conditions, within the subordinate clause, across the whole experiment. If a verb appeared in the Implausible condition in the one run, it appeared in the Plausible condition in the other run and vice versa. A verb appeared in the Control condition in the same run as the Implausible or the Plausible condition roughly half the time.

Each sentence reading trial began with a 50 ms fixation cross followed by a jittered interval (min = 1503, max = 2996, mean = 2256) and then the sentence, which was shown until the participant responded or for a maximum of 7 s. This was followed by a 50 ms fixation cross, another jittered interval (min = 1510, max = 2992, mean = 2243), and then the comprehension question, which appeared until the participant responded or for a maximum of 4 s. See Fig. 2a for an example trial.

Visual attention task

In the visual attention task, participants saw a colored image and a sentence below it and were asked to evaluate whether the sentence described the image accurately. They responded using the index and ring fingers of their right hand on the left and right buttons to indicate Yes and No, respectively. Sentences were displayed in black text on a white background.

We used the same stimuli as in the norming study with a minor modification. The word "gloves" was changed to "mittens" because the latter was a more accurate description of the image. See Appendix for a full list of sentences. The 84 stimuli were split into two runs with 42 trials each (21 More/Less, 21 All/Some). Each quantifier appeared 10–11 times within a run. The correct response was Yes or No half of the time (21 each, split roughly evenly, 10–11 times in each condition). Across runs, each condition contained "right" versus "left" an equal number of times (21 each). Trials were pseudorandomly ordered such that the same quantifier did not occur more than two times in a row, nouns did not repeat in consecutive trials, and left/right and yes/no correct answer did not occur more than three times in a row.

Each trial began with a 50 ms fixation cross followed by a jitter (min = 3,039 ms, max = 5,907 ms, mean = 4,483 ms) and then the stimulus, which was shown until the participant responded or for a maximum of 4,000 ms. See Fig. 2b for an example trial.

fMRI acquisition and analyses

Structural (TR = 1,900 ms, TE = 2.52 ms, Slice thickness = 1 mm, 176 slices) and functional images (TR = 1,400 ms, TE = 29 ms, Interleaved multi-slice mode, Slice thickness = 2.3 mm, Flip angle = 90°)

were acquired using a 3 T Prisma-Fit scanner. All scans took place at the Center for Functional and Molecular Imaging at Georgetown University. Analyses were conducted using FSL. Out-of- brain voxels were removed. Pre-processing included motion correction using MCFLIRT, interleaved slice timing correction, spatial smoothing (FWHM = 5 mm), and high-pass filtering (100 s). Images were registered to the MNI 2-mm template. All statistical models included the standard motion parameters and motion outliers (fsl_motion_outliers). For the sentence reading task, the model included 10 events—two jitters, two fixations, the five conditions (Implausible, Control, Plausible, LongWM, ShortWM), and the question. For the visual attention task, the model included four events: jitter, fixation, and the two conditions (More/Less, All/Some). Jitters were of variable durations (see Fig. 2) to facilitate separating activation for the sentence from activation for the question and/or separating activation for different trials.

In ROI analyses, we functionally profiled left hemisphere regions identified from the contrast of garden-path Implausible sentences versus non-garden-path Control sentences (Z threshold = 3.1, Cluster $P < 0.05$), which could include regions involved in cognitive control, syntactic analysis, semantic/pragmatic processing, working memory, and general task demands. The large left temporal cluster from this analysis was split into an anterior temporal cluster (intersection with left anterior MTG from the Harvard-Oxford atlas) and a posterior temporal cluster (all other voxels posterior to the anterior part). The left frontal cluster from the analysis was split into a left frontal (language) cluster based on intersection with the language network, and a left frontal (language/MD) cluster based on intersection with both the language and the MD networks from the probabilistic functional atlas of Fedorenko and colleagues (Lipkin et al. 2022). Using these group-based clusters, we defined subject-specific ROIs as 5 mm spherical regions around the peak Implausible activation coordinate for each subject. (We used the peaks for Implausible versus Baseline rather than Implausible versus Control here because the former is orthogonal with the planned contrasts (Plausible vs Control and Long WM vs Control) and the latter is not. Using peaks identified using the latter would bias the ROIs towards voxels with low Control activation and thereby bias in favor of finding a difference of all conditions against Control. Note however, that the peaks were determined within ROIs that were selected from the Implausible vs Control contrast. Therefore, we expect Implausible to show higher activation than Control in these ROIs by default and do not test again for that.) Thus, there were four ROIs per subject—left posterior temporal, left anterior temporal, left frontal (language) and left frontal (language/MD). Within each ROI, Implausible will have higher activation than Control by definition because the voxels were identified within clusters from the Implausible > Control group analysis. Our main interest was in determining whether these ROIs also showed higher activation for Plausible and Long WM compared to Control, and for Long vs Short WM. Accordingly, we conducted three paired t-tests for these contrasts (Bonferroni correction for $P < 0.05/3 = P < 0.0166$) within each ROI. Additionally, to determine if the regions showing increased recruitment for one or more of the three sentence types were also engaged during the visual attention task, we examined the contrast of More/Less versus All/Some within each ROI. Together, these analyses were used to functionally profile different regions based on the logic outlined in Table 1.

In addition to the ROI analyses, we conducted whole-brain analyses contrasting closely matched conditions for each type of stimulus: (i) Implausible vs Plausible garden-path; (ii) Long vs Short WM; and (iii) More/Less vs All/Some visual attention. This was intended to provide supplementary information about other potential regions engaged for comprehending sentences that were difficult along different dimensions (semantic/pragmatic processing, working memory, visual attention). In each case, we also examined the reverse contrast (e.g. Short vs Long WM) for completeness.

To summarize, planned ROI analyses in regions activated for Implausible versus Control examined whether Plausible and Long WM were also activated related to Control and whether Long WM showed more activation than Short WM in those regions. Planned whole-brain analyses between closely matched conditions were used to determine the regions engaged for Implausible vs Plausible, Long vs Short WM, and More/Less vs All/Some (and vice versa). In addition, we conducted post-hoc analyses in some cases to clarify surprising results (see below).

Results

Behavioral results: Sentence reading task

For the sentence reading task, we analyzed reading time for the sentence, and accuracy and reaction time for responding to the comprehension question. For reading times, trials where the participant did not respond within the time allowed were replaced with the maximum duration (7,000 ms). Residuals from modeling raw reading times were normally distributed, so we did not log-transform them. Reading times were adjusted for length. (Mean non-length-adjusted reading times for the different sentence conditions were as follows: Control M = 4,172 ms, SD = 1,728 ms; Implausible M = 4,163 ms, SD = 1,727 ms; Plausible M = 4,151 ms, SD = 1,727 ms; Long WM M = 4,448 ms, SD = 1,669 ms; Short WM M = 4,388 ms, SD = 1,605 ms. The fMRI models included variable event durations based on the sentence reading times. Variable epoch (cf. constant epoch or impulse) models accurately represent the cognitive differences between stimuli and conditions and show higher detection power without increasing false positives (Grinband et al. 2008). Additionally, our results show informative differences between the different ROIs, all of which were modeled using the same variable epoch durations.) There were no outliers. Mean length-adjusted reading time for the different conditions were as follows: Control = 43.9 (SD = 1717.3, SE = 49.2, 95% CI = 96.5), Implausible = 41.6 (SD = 1729.5, SE = 49.6, 95% CI = 97.2), Plausible = 38.2 (SD = 1721.8, SE = 49.3, 95% CI = 96.8), Long WM = -30.7 (SD = 1640.1, SE = 47.0, 95% CI = 92.2), Short WM = -92.9 (SD = 1590.3, SE = 45.6, 95% CI = 89.4). Compared to Control, Implausible, and Plausible sentences were not significantly different (P's > 0.9). Short WM had a significantly shorter length-adjusted reading time ($t(2406.00) = -2.93$, $P < 0.01$). For Long WM, the effect was in the same direction but not statistically significant ($t(2406.00) = -1.52$, $P = 0.13$). Thus, the working memory stimuli, although longer, were read at a quicker rate, suggesting that these sentences with coordinate clauses were easier to read than the sentences with subordinate clauses (Control, Implausible, and Plausible).

For comprehension questions during the reading task, mean accuracies in the different conditions were: Control = 88.5% (SD = 35.7, SE = 1.0, 95% CI = 2.0), Implausible = 89.8% (SD = 33.8, SE = 1.0, 95% CI = 1.9), Plausible = 78.2% (SD = 46.2, SE = 1.3, 95% CI = 2.6), Long WM = 82.8% (SD = 42.2, SE = 1.2, 95% CI = 2.4), Short WM = 89% (SD = 35.0, SE = 1.0, 95% CI = 2.0). (Comprehension accuracies for Long WM and Short WM were lower than in the norming study, in accordance with changes to the stimuli that made the sentences longer (see above). For Implausible and Plausible, accuracies were higher than in the norming study despite using the same stimuli. This was not predicted. There were multiple differences between the two studies, including the participant population (broader sample versus a university-based sample) and test setting (online versus in-person) that could have impacted the results. Importantly, both studies demonstrated a robust behavioral difference between the Plausible and Implausible conditions that was consistent with psycholinguistic predictions.) All conditions were above chance (exact binomial test P's < 0.001), indicating that participants attended to the task. We compared closely matched conditions that had similar questions (Implausible vs Plausible and Long WM vs Short WM). The Plausible condition had significantly lower accuracy than the Implausible condition ($z = -8.40$, $P < 0.001$), consistent with our predictions and previous findings on lingering misinterpretations. The Long WM condition had significantly lower accuracy than the Short WM condition ($z = -4.78$, $P < 0.001$), consistent with predictions based on working memory demands. For reaction times, trials where the participant did not respond within the time allowed were replaced with the maximum duration (4,000 ms). We log-transformed the reaction times and removed outliers (using the same procedure as for the Norming study). Mean log reaction times for the different conditions were as follows: Control = 7.03 (SD = 0.51, SE = 0.02, 95% CI = 0.03), Implausible = 6.94 (SD = 0.58, SE = 0.02, 95% CI = 0.03), Plausible = 7.0 (SD = 0.61, SE = 0.02, 95% CI = 0.04), Long WM = 7.29 (SD = 0.43, SE = 0.01, 95% CI = 0.03), and Short WM = 7.30 (SD = 0.42, SE = 0.01, 95% CI = 0.03). Comparing Implausible and Plausible conditions, the latter had a significantly longer reaction time than the former ($t(1980) = 3.58$, $P < 0.001$). Long and short WM conditions did not differ from one another ($P > 0.6$). To summarize, participants showed lower accuracy and longer reaction time to questions about the ambiguous portion of Plausible than Implausible sentences, indicating more lingering misinterpretations for the former that cannot be attributed to a speed-accuracy tradeoff. For working memory, participants showed lower accuracy and equivalent reaction time to questions about Long than Short WM sentences, tentatively indicating some difficulty with the former sentence type.

Behavioral results: Visual attention task

For the visual attention task, we analyzed accuracy and reaction time. Mean accuracies for the More/Less and All/Some conditions were 80.5% and 92.1%, respectively. Both were above chance (binomial P's < 0.001). Comparing the two conditions, participants were significantly less accurate on More/Less ($z = -8.12$, $P < 0.001$), consistent with our predictions. Reaction times were log-transformed. (Mean raw reaction times for the two conditions were as follows: More/Less M = 2,520 ms, SD = 1,017 ms; All/Some M = 2,043 ms, SD = 1,016 ms. Thus, the fMRI models, on average, comprised longer epochs for the More/Less than the All/Some condition. However, we found different relative activation patterns in different ROIs (e.g. All/Some > More/Less in posterior temporal but not in left frontal (language and MD)) that cannot be reduced to epoch duration differences. See also footnote 6.) There were no outliers identified using the same procedure as above. Log-transformed reaction times for the two conditions were as follows: More/Less = 7.75, All/Some = 7.54. There was a significant effect of condition ($t(2050) = 20$, $P < 0.001$), with responses being slower in the More/Less condition. Together, these analyses confirmed that More/Less was harder than All/Some, as intended, resulting in lower accuracy and longer reaction time.

fMRI results: Sentence reading task

Planned ROI analyses

Garden-path Implausible sentences showed more activation over non-garden-path Control sentences in left temporal and frontal areas that fell within Fedorenko and colleagues' language as well as MD networks (Fig. 3. See SI #1 for the list of clusters). Four regions—left posterior temporal, left anterior temporal, left frontal (language) and left frontal (language and MD)—were extracted from these clusters (see Methods). The left posterior temporal region was the largest (1,169 voxels) compared to the other regions (left anterior temporal: 276, left frontal (language): 249, and left frontal (language and MD): 166). Subject-specific ROIs were defined around the peak activation for Implausible sentences for each subject within the four regions (see insets in Fig. 4). These subject-specific ROIs varied in their degree of inter-subject overlap for the four regions. For the left posterior temporal cortex, out of all the voxels across all subject's ROIs, 62% were unique to a single subject (overlap = 1) followed by 19%, 9% and 6% for overlap values of 2, 3, and 4 respectively. The corresponding numbers for the other three regions showed less unique and more shared voxels (overlap 1–4 for left anterior temporal: 44%, 24%, 10%, 6%; left frontal (language): 38%, 33%, 16%, 9%, left frontal (language and MD): 29%, 16%, 18%, 19% for overlap values 1 to 4, respectively). Thus, as can be seen visually in the insets in Fig. 4, the left posterior temporal cortex, by virtue of being the largest identified region, allowed for more variation in the peak coordinates for different participants. We return to this in Discussion.

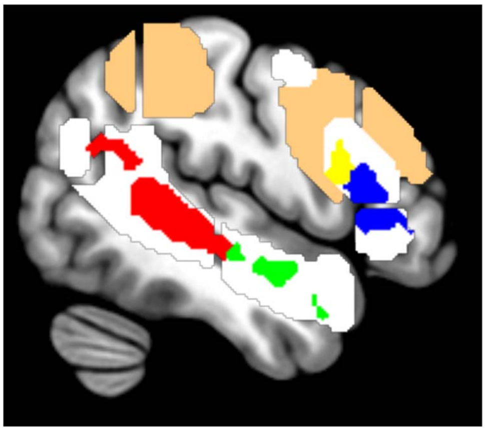

Fig. 3. Implausible versus control contrast revealed activation in left posterior temporal (red), left anterior temporal (green), left frontal (language) (blue), and left frontal (language and MD) (yellow) regions. Clusters are superimposed on Fedorenko and colleagues' language (white) and MD (tan) networks.

Figure 4 shows activation for the different conditions and the planned contrasts for the sentence reading task within the left posterior temporal, left anterior temporal, left frontal (language), and left frontal (language/MD) ROIs. Paired-sample t-tests, which correspond to the contrasts shown in the figure, revealed that in the left posterior temporal ROI, Plausible showed increased activation over Control ($t(28) = 4.81$, $P < 0.001$. Fig. 4a). Long WM did not show a significant difference from Control ($t(28) = 0.57$, $P = 0.58$) or Short WM ($t(28) = 1.17$, $P = 0.25$). The same pattern was observed in the left frontal (language) ROI (Fig. 4c. Plausible: $t(28) = 3.97$, $P < 0.001$; Long WM vs Control: $t(28) = -.36$, $P = 0.72$; Long WM vs Short WM: $t(28) = 0.91$, $P = 0.37$). In the left anterior temporal ROI, none of the three comparisons revealed significant effects (Fig. 4b. All P's > 0.2). Finally, in the left frontal (language/MD) ROI (Fig. 4d), Plausible showed increased activation compared to Control ($t(28) = 3.50$, $P = 0.002$). Long WM showed decreased activation relative to Control ($t(28) = -2.58$, $P = 0.015$) and was not different from Short WM ($t(28) = 0.07$, $P = 0.95$). To summarize, a pattern consistent with involvement

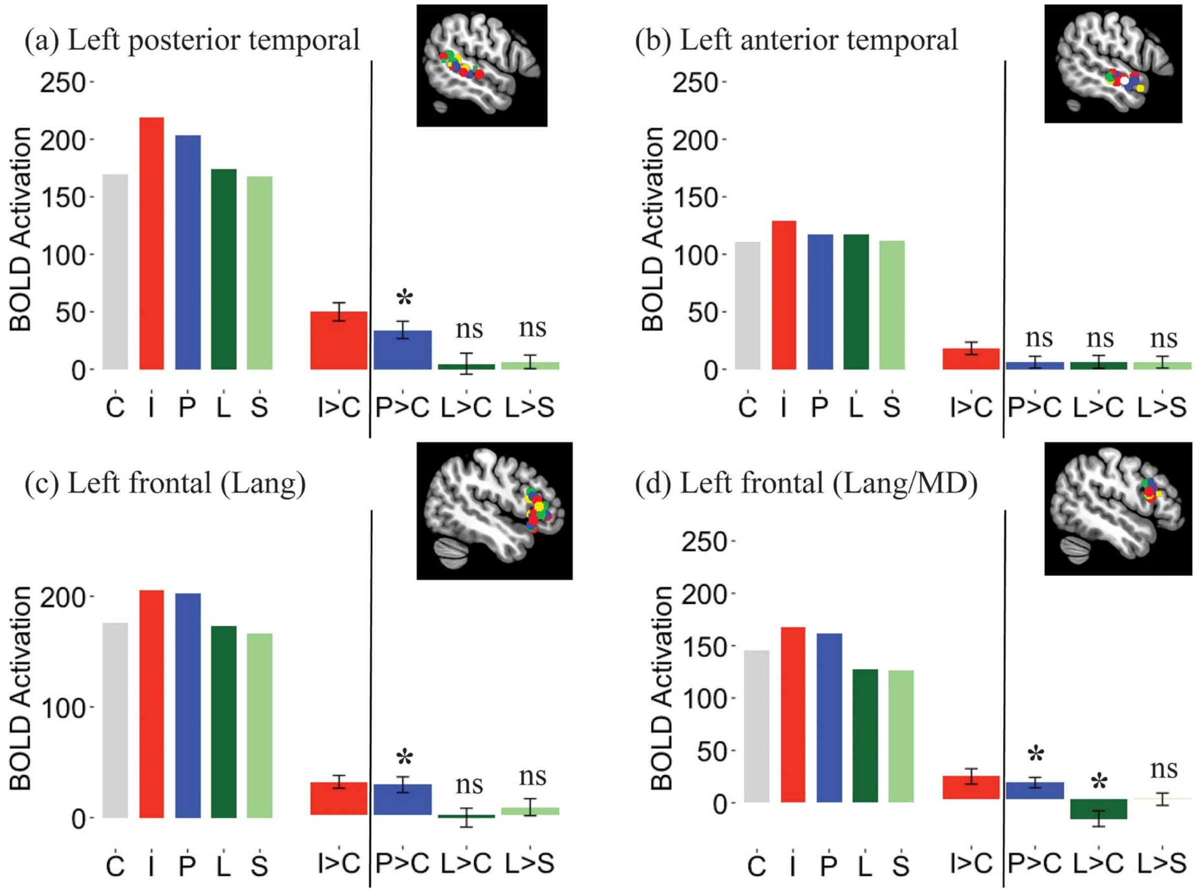

Fig. 4. Results for the sentence reading task in (a) left posterior temporal (b) left anterior temporal (c) left frontal (language) and (d) left frontal (language/MD) ROIs. C = control, I = implausible, P = plausible, L = long WM, S = short WM. ROIs were selected for showing activation for implausible. Subject-specific spherical ROIs were defined around the peak implausible > baseline activation within group-level clusters showing implausible > control. The analyses examined which other planned contrasts showed significant effects (information to the right of the vertical line in each panel). * = significant at Bonferroni-corrected threshold, ns = not significant. Brain image inset in each panel shows the spherical ROIs for different participants.

in cognitive control or syntactic reanalysis was found in the left frontal (language) and the left posterior temporal ROIs, which showed increased activation for Plausible over Control (in addition to Implausible over Control). The left frontal (language/MD) ROI was subtly but importantly different from the left frontal (language) ROI in how it responded to the Long WM condition (see more in Discussion). The left anterior temporal ROI was the only ROI that was activated for the Implausible but not the Plausible condition relative to Control (see more below).

Planned whole-brain analyses

Whole-brain analysis for Implausible versus Plausible revealed a significant effect in left anterior temporal cortex only (Fig. 5), consistent with the ROI results. For Long versus Short WM, there were no significant clusters. In both cases, the reverse contrast (Plausible > Implausible, Short WM > Long WM) revealed no significant effects.

Post-hoc analyses

The planned ROI and whole-brain analyses did not detect any significant differences between the Long WM and Short WM conditions. The ROI analyses also did not reveal any differences between Long WM and Control. To determine whether any regions showed an effect of the Long WM manipulation, we conducted a post-hoc whole-brain analysis for Long WM > Control. This revealed several occipital, temporal and medial frontal clusters

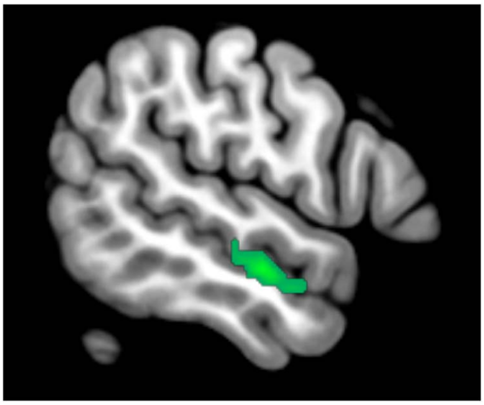

Fig. 5. Whole-brain analysis results implausible vs. plausible sentences. Single cluster. Peak coordinate at MNI (−56, −12, −14).

(see SI #2), suggesting that we could detect activation related to the WM manipulation and that the null results in the left frontal cortex were not solely due to low power. In the left perisylvian cortex, there were two temporal clusters. The pattern of activation within the larger of these clusters is shown in Fig. 6. As before, we defined spherical ROIs around subject-specific peaks for Long WM (versus Baseline) within the cluster. Long WM showed more activation over Short WM, indicating sensitivity to the working memory manipulation ($t(28) = 2.12$, $P < 0.05$). Further, Long WM showed more activation than the hardest condition, namely the

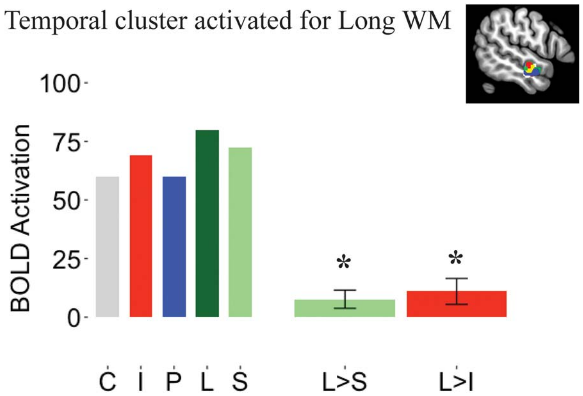

Fig. 6. Results for the sentence reading task in subject-specific ROIs defined around the peak long WM > baseline activation within a group-level left temporal cluster obtained from the post-hoc contrast of LongWM > control. C = control, I = implausible, P = plausible, L = long WM, S = short WM. * = $P < 0.05$.

Implausible condition, suggesting that the effect was specifically about working memory related to binding rather than general-purpose difficulty ($t(28) = 2.15$, $P < 0.05$). These results should be interpreted with caution because the analysis was post-hoc and we looked within ROIs defined for showing high Long WM activation. However, they provide additional context for interpreting the patterns shown in Fig. 4 (see Discussion).

fMRI results: Visual attention task

Planned ROI analyses

Looking in the ROIs examined for the sentence reading task (Fig. 7), All/Some showed more activation than More/Less in both temporal regions (left posterior temporal: $t(28) = 3.72$, $P < 0.001$; left anterior temporal: $t(28) = 2.28$, $P = 0.03$). The difference was marginally significant in the left frontal (language) ROI ($t(28) = 1.87$, $P = 0.07$) and not significant in the left frontal (language/MD) ROI ($t(28) = 0.28$, $P = 0.78$). Thus, contrary to our expectations (Table 1), All/Some showed increased activation over More/Less rather than vice versa. We conducted post-hoc analyses to better understand these surprising results (see below). Similar to the results for the sentence reading task, the two frontal ROIs were subtly but importantly different, suggesting a gradient for language-specificity (see Discussion).

Planned whole-brain analysis

Whole-brain analysis of More/Less versus All/Some found a significant effect in the visual cortex, consistent with the hypothesized need for greater visual processing in this condition (Fig. 8a). The reverse contrast (All/Some versus More/Less) revealed significant effects in several bilateral anterior and posterior regions (Fig. 8b), consistent with the ROI results, suggesting that All/Some recruited additional—including linguistic—processing that was different from the visual processing used for More/Less.

Post-hoc analyses

A priori, we had predicted that the harder of the two conditions—More/Less—would show more activation than All/Some in regions associated with visual processing and general-purpose task demands. We did not expect to find increased activation for All/Some relative to More/Less in any brain regions. Given the surprising reverse findings, we investigated whether "All" and "Some" differed individually relative to More/Less. Specifically, because "Some" has been theorized to require additional processing to derive an implicature (see e.g. Huang and Snedeker 2018), we explored whether the reverse pattern was primarily driven by "Some." The results revealed similar results for All and Some in three of the four ROIs. Both showed more activation than More/Less in the left posterior temporal region (All: $t(28) = 2.41$, $P = 0.02$; Some: $t(28) = 3.08$, $P < 0.01$). Neither showed increased activation in the left frontal (Lang/MD) ROI (both P's > 0.4). The effects were marginal or not significant in the left frontal (Lang) ROI (All: $t(28) = 1.70$, $P > 0.1$; Some: $t(28) = 1.77$, $P = 0.09$). The only region showing a differential effect was the left anterior temporal ROI, where was a significant effect for All versus More/Less ($t(28) = 3.45$, $P < 0.01$) but not Some vs More/Less ($t(28) = 0.76$, $P = 0.45$). We discuss this further below.

Discussion

The goal of this study was to fractionate difficulty during sentence comprehension. Difficult-to-comprehend sentences are often difficult in more than one way. Therefore, we used a functional profiling approach to isolate different processes that can contribute to the interpretation of such sentences. Our results suggest that sentence comprehension recruits regions that lie on a spectrum of specificity to generality with respect to the representations and processes that are involved. On the highly specific end, hard-to-understand garden-path sentences recruited left temporal regions that are widely thought to support language-selective processing of linguistic representations—left posterior temporal cortex for syntactic analysis and left anterior temporal cortex for semantic/pragmatic processing. On the highly general end, these sentences recruited dorsal left frontal cortex, which is linked to domain-general processing of a variety of representations (e.g. linguistic, spatial, mathematical) as part of the MD network. In between these extremes, these sentences also recruited ventral left frontal cortex, whose role is debated. This region could be involved in language-selective processing of linguistic representations (akin to the left temporal cortex. See e.g. Shain et al. 2022), a computationally general process of integration or cognitive control operating over linguistic representations (e.g. Hagoort 2005; Thothathiri et al. 2012c), or a computationally general process of cognitive control operating over a variety of representations (e.g. Hsu et al. 2017). Below, we discuss each region in turn.

The left posterior temporal cortex showed a functional profile that we had hypothesized a priori as reflecting syntactic processing (see Table 1). Specifically, this region showed increased recruitment for garden-path Implausible and Plausible sentences that tend to be misinterpreted initially, therefore requiring additional syntactic analysis. Prior evidence has shown that damage to this region impacts comprehension, especially when sentences have to be interpreted using syntax and cannot be understood using semantic cues alone. Neuroimaging studies in neurotypical adults routinely report activation in this region when contrasting sentences that differ in their syntactic processing demands. Our results corroborate this evidence and show that one aspect of difficulty for garden-path sentence comprehension is syntactic analysis. An unexpected pattern found in our results was activation in this region for All and Some over More/Less trials in the visual attention task. We did not predict this pattern and can only speculate about the explanation. The All/Some sentences contained the reduced relative clause "shown here" in order to match their sentence lengths to their More than half/Less than half counterparts. This added syntactic complexity could explain why the posterior temporal cortex was more activated for the former than the latter.

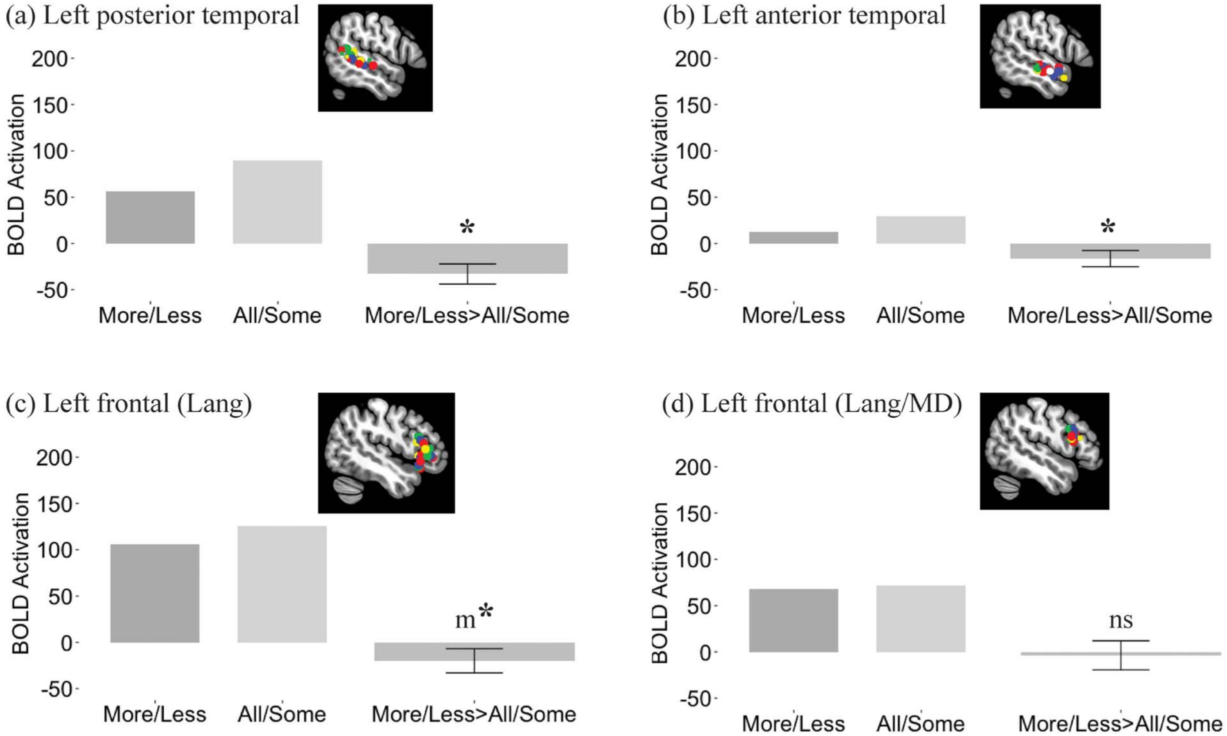

Fig. 7. Results for the visual attention task in (a) left posterior temporal (b) left anterior temporal (c) left frontal (language) and (d) left frontal (language/MD) ROIs (same as in Fig. 4). * = significant, m* = marginally significant, ns = not significant.

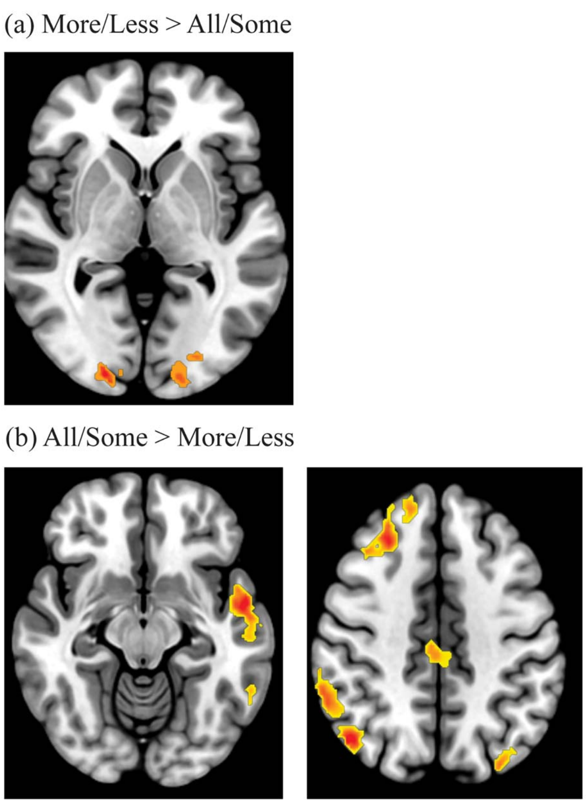

Fig. 8. Whole-brain analysis results for the visual attention task.

As described in Results, the left posterior temporal cortex was the largest of the four regions activated for Implausible versus Control. The location of the subject-specific peaks for the Implausible condition—and therefore the location of the subject-specific ROIs—was more widespread for this than the other three regions. Hypothetically, this greater variation in location could have weakened the results if different parts within the posterior temporal lobe performed different functions. The findings revealed robust activation for Plausible over Control (and for All/Some over More/Less), however. This suggests that even though the peak locations were anatomically variable across individuals, the functional operations at those locations were relatively similar.

The profile for the left anterior temporal cortex differed from the posterior temporal ROI. This region showed increased activation for Implausible but not Plausible sentences, consistent with our hypothesis about semantic/pragmatic processing regions. Notably, this was the only region that showed this pattern in the ROI analyses, out of the two temporal and two frontal ROIs. The whole brain contrast of Implausible > Plausible also identified only this region. Across multiple studies, Implausible sentences show higher comprehension accuracy than Plausible sentences (corroborated by the behavioral accuracy results here) and stronger activation in language processing regions (see e.g. Christianson et al. 2001; den Ouden et al. 2016). Thus, Implausible sentences appear to involve some additional processing that enables the comprehender to successfully let go of the original misinterpretation. We had hypothesized that this processing might be semantic/pragmatic in nature, for assessing the compatibility of the original and the later interpretations of the sentence given our knowledge of the world (e.g. Can the wrestlers be wrestling the rivals if the rivals just arrived?). The recruitment of the left anterior temporal cortex for such semantic/pragmatic processing is consistent with prior neuropsychological evidence from semantic dementia and neuroimaging evidence from neurotypical individuals.

Two other unexpected result patterns would have to be reconciled with the above interpretation, however. In the visual attention task, "All" trials but not "Some" trials activated this region more than "More/Less" trials. Prior evidence suggests that conceptual combination, including for words denoting quantities, engages the anterior temporal lobe (Blanco-Elorrieta and Pylkkänen 2016). However, it is unclear why "All of the <noun>" should be more likely to engage such combinatorial processing than the other quantifiers used in this study. The second surprising pattern was that post-hoc analysis of LongWM versus Control sentences identified an ROI that was quite close to the anterior temporal ROI (compare the brain insets in Fig. 4b and Fig. 6). As designed, the LongWM sentences were not subject to any misinterpretation that should require deeper semantic/pragmatic processing and it is therefore unclear why this condition recruited the anterior temporal lobe. One previous study reported that the anterior temporal lobe was sensitive to the type of antecedent (person versus thing) in sentences containing reflexive pronouns, with greater recruitment for person than thing antecedents (Hammer et al. 2011). This study was conducted in German, where syntactic gender (which applies to nouns referring to both persons and things) can be distinguished from semantic gender (which applies to nouns referring to persons only). The authors concluded that the activation pattern (person > thing) was consistent with greater semantic gender processing for the former. Taken together, the results are broadly consistent with the idea that implausible garden-path sentences recruited additional semantic processing of some kind within the left anterior temporal lobe. But the precise nature of that processing—spanning across implausible garden-path sentences, specific quantifiers, and sentences that involve binding between reflexive pronouns and animate antecedents in nearby regions within the left anterior temporal lobe—needs to be clarified by future studies. (A different possibility suggested by a reviewer is that left anterior temporal lobe activation could reflect greater surprisal for Implausible than Plausible sentences, given that the comprehension system is forced to recognize and handle the ambiguity/misinterpretation more often for the former than the latter. Under this account, it is unclear why the left anterior temporal lobe rather than left frontal or left posterior temporal cortex should be especially sensitive to surprisal (given the extensive evidence for those other regions' involvement in sentence processing). It is also unclear why unambiguous Long WM sentences would show surprisal-related activation. Overall, there are puzzling patterns for either the semantic processing or the surprisal account. But given the extensive literature on semantic processing within the left anterior temporal lobe, we favor an interpretation based on semantics.)

On the domain-general end of the spectrum, dorsal left frontal cortex (the left frontal (language/MD) ROI) revealed a profile consistent with general task-difficulty-based activation, consistent with previous evidence on the MD network. We observed maximal activation in this region for the difficult sentences containing ambiguity/conflict, namely Implausible and Plausible. But we also saw more activation here for Control over LongWM. Thus, increased activation of this ROI was not linked to ambiguity/conflict alone but seemed to track general difficulty. Sentences with the highest length-adjusted reading times (Implausible, Plausible, Control) showed the most activation. Further, this ROI differed from the temporal ROIs in its pattern for the visual attention task. In the temporal ROIs, All/Some showed significantly more activation than More/Less, consistent with higher linguistic demands (see above). By contrast, the left frontal (language/MD) ROI showed equivalent activation between More/Less and All/Some. We had predicted more activation for More/Less in this region, in accordance with the greater task difficulty that was confirmed by lower accuracies and longer reaction times. The observed pattern instead suggests that this region was sensitive to difficulty in a way that encompasses both linguistic and non-linguistic task demands. The pattern within left frontal (language/MD) also differed subtly from the adjacent left frontal (language) ROI, congruent with prior evidence that suggests more domain-generality as we move more dorsally (see below. Hagoort 2005; Fedorenko et al. 2013).

Last but not least, we observed increased activation for Implausible and Plausible garden-path sentences over Control sentences within ventral left frontal cortex (the left frontal (language) ROI), consistent with a role for this region especially in cases of ambiguity and conflict. This is consistent with either cognitive control or syntactic processing (see Table 1). Hsu et al. (2017) demonstrated that several conflict resolution tasks that do not involve syntax (Stroop, N-back with lures, and Recent Negatives) showed activation within the same subject-specific voxels within ventral left frontal cortex as a syntactic ambiguity task. This suggests that this region is involved in a computationally general process, namely cognitive control, that can resolve conflict during different tasks (not just sentence processing). The authors suggested that functional specificity for particular representations might arise at the network level, wherein ventral left frontal cortex interacts with different brain regions for different kinds of representations. By comparison, Fedorenko and colleagues have argued that ventral frontal (and other) regions within their language-selective network are selectively activated for linguistic tasks. At the representational level, this can be successfully reconciled with Hsu et al. (2017) because all tasks in that study involved some kind of linguistic representation (letters, words, sentences). The results for the present study too are consistent with language-selectivity in this sense because all of our stimuli included linguistic content. The question remains however about the type of *process* that is supported by this region. While Fedorenko and colleagues have suggested, either implicitly or explicitly, that this region is engaged in a language-specific process, others have pointed to co-localization among very different tasks (e.g. Stroop, sentence comprehension) to argue that the process is a more general one like integration or conflict resolution (Hagoort 2005; Thothathiri et al. 2012c; Hsu et al. 2017). The present results are consistent with both the "more general" and the "more specific" perspectives. They show increased recruitment especially for sentences containing conflict, consistent with a general process of cognitive control within this region. However, we cannot rule out the possibility that this increased recruitment reflects something more language-specific like syntactic processing. In this context, it is worth noting that in the visual attention task, there was a weak tendency towards All/Some > More/Less in the left frontal (language) ROI, which fell in between the pattern in the left temporal ROIs (which uncontroversially deal with linguistic representations) and the pattern in the dorsal left frontal ROI (which is linked to the domain-general MD network). This raises challenges for both types of accounts. On the one hand, the "cognitive control over linguistic representations" account would have predicted a null effect because neither condition involved linguistic conflict. Conversely, the "language-specific processing" account would have predicted a significant All/Some > More/Less effect akin to what was observed in the temporal ROIs, which are thought to be a part of the language network. Because we observed neither pattern, future research is needed to evaluate which account is better supported by the collective evidence across multiple studies and/or reconcile the two theories.

Could activation for the garden-path sentences within the left frontal (language) ROI reflect syntactic working memory, as suggested by some previous studies (e.g. Shain et al. 2022)? We did not find increased activation for Long WM over Short WM or Control sentences in this ROI, which is inconsistent with this account. We designed the Long and Short WM conditions based on multiple previous studies that used the distance between items in a syntactic dependency to tap into working memory (e.g. Fiebach et al. 2005; Makuuchi et al. 2009). However, our behavioral and planned fMRI analyses suggested that this manipulation might have been weak, which could potentially explain the lack of WM-related effects within this ROI. Therefore, we conducted post-hoc whole-brain and ROI analyses, which showed increased activation for the Long WM condition in left temporal cortex and other non-lateral-frontal regions. This suggests that there was detectable activation for operations related to binding and working memory within this study and that these effects were more reliable in the left temporal than frontal cortex. Overall, the evidence tentatively suggests that the left frontal (language) ROI is not the most sensitive part of the language network as pertains to working memory, which stands in contrast to its robust sensitivity to garden-path sentence processing demands. Therefore, we favor an interpretation of the results in terms of cognitive control or some other language-specific operation, as discussed above. Future studies with stronger working memory manipulations are needed to tease apart whether working memory plays a role during garden-path sentence processing that is independent of cognitive control.

The issue of difficulty during sentence comprehension and its relationship to cognitive control has come to the forefront in the past two decades of neuroimaging research because of growing evidence on two different fronts. On the one hand, a number of studies have now shown that frontal regions engaged during cognitive control tasks are also engaged during sentence comprehension and more causally, that triggering cognitive control facilitates the comprehension of sentences containing conflict. However, these demonstrations have raised questions about whether the evidence indicates a role for conflict resolution or cognitive control specifically, or whether the evidence can be interpreted more broadly in terms of "difficult" situations and the brain regions that support difficulty. Thus, whether cognitive control, in particular, is useful for sentence comprehension has been called into question. On a different front, other evidence has indicated a distinction between brain networks that support language processing and those that support broader demands like general-purpose difficulty. This evidence is usually interpreted as showing that language processing is mostly supported by language-selective processes and that any difficulty-related processing occurs in distinct brain areas (the MD network) and is only used under some circumstances. Thus, this strand of research too has questioned the relevance of cognitive control for sentence comprehension, relegating it to a cluster of general-purpose difficulty processing within the MD network.

Hsu et al. (2017) addressed this concern by examining whether the contrast between conflict and no-conflict trials in different tasks (Stroop, syntactic ambiguity, etc.) activated ventral left frontal cortex or the MD network. They found that four different conflict tasks converged in activating the same voxels within ventral left frontal cortex, supporting a role for this region in conflict resolution. In contrast, they did not observe such co-localization within the MD network. Therefore, the authors argued that general-purpose difficulty cannot account for their findings. The present study offers new complementary evidence on this debate. Like Hsu and colleagues, we found increased activation within ventral left frontal cortex for conflict versus no-conflict sentences. We also found activation for such sentences within more dorsal MD-related parts of the left frontal cortex, consistent with other researchers' claims about general-purpose task demands and difficulty. However, through examining different kinds of sentences within the same participants, our results demonstrate the distinct functional profiles of ventral versus dorsal left frontal cortex. Both regions are engaged for sentences containing ambiguity and conflict but they show different recruitment patterns for other kinds of sentences that did not contain conflict and varied in difficulty. Thus, our results suggest that conflict resolution and the handling of general difficulty are not the same (recruiting different regions) and that they are both relevant during the comprehension of garden-path sentences.

More broadly, the dichotomization of language-selective versus MD regions in recent literature can be a useful empirical tool (as the present study too demonstrates), but we would suggest that this should not obscure the fact that the human brain can use multiple tools at its disposal to handle any given instance of sentence comprehension, be they domain-general or language-specific. Put another way, when a listener or reader encounters hiccups during sentence comprehension, both general-purpose processes that enable adjustment broadly to a difficult situation and more specific processes that adapt in a language-specific way can be helpful (see Sharer and Thothathiri 2020 for additional evidence and discussion).

To summarize, we found that a variety of brain regions support comprehension when sentences are difficult to understand. Specifically, regions linked to syntactic analysis (left posterior temporal cortex), semantic/pragmatic processing (left anterior temporal cortex), linguistic cognitive control or some other language-specific processing (ventral left frontal cortex), and general task demands (dorsal left frontal cortex) were all recruited when participants read sentences that triggered conflicting interpretations, which must be reconciled for accurate comprehension. Thus, difficulty during sentence comprehension is multi-faceted and different demands are handled by different brain regions.

Acknowledgments

A huge thanks to Luca Cavan and Phoebe Jeng for assistance with stimulus design, norming, and data collection. The corresponding author (MT) can be contacted at malathi@gwu.edu.

Author contributions

Malathi Thothathiri (Conceptualization, Formal analysis, Funding acquisition, Project administration, Writing—original draft, Writing—review & editing), Jana Basnakova (Conceptualization, Writing—review & editing), Ashley Lewis (Conceptualization, Writing—review & editing), Josephine M. Briand (Data curation, Visualization, Writing—original draft).

Supplementary material

Supplementary material is available at *Cerebral Cortex* online.

Funding

This work was supported by a National Institutes of Health grant (R01DC017138) awarded to MT and Gravitation Grant of the Language in Interaction Consortium from the Netherlands Organization for Scientific Research (024.001.006) to AGL.

Conflict of interest statement: None declared.

Appendix

CCWM Stimuli

Condition	Sentence	Question
Implausible	While Harry steered the wooden canoe was moored at the dock.	Did Harry steer the canoe?
Implausible	While Henry stirred the sauce boiled on another burner.	Did Henry stir the sauce?
Implausible	As the girls raced the competitive opponents watched from the sideline.	Did the girls race the opponents?
Implausible	As the artist painted the self-portrait hung in the famous museum.	Did the artist paint the portrait?
Implausible	While the woman ordered the brown chicken escaped the farm.	Did the woman order the chicken?
Implausible	While Tom wrote the note rested in the trashcan with the waste.	Did Tom write the note?
Implausible	As the purring kitten sniffed the catnip was missing from the bag.	Did the kitten sniff the catnip?
Implausible	As the maid dusted the grand piano was left untouched and silent.	Did the maid dust the piano?
Implausible	As Rick explored the job opportunity escaped his notice entirely.	Did Rick explore the job opportunity?
Implausible	As the cameraman filmed the crew was nowhere to be found.	Did the cameraman film the crew?
Implausible	While the assistant typed the transcript came in the mail.	Did the assistant type the transcript?
Implausible	As Lola drank the white wine perched unopened in the wine cellar.	Did Lola drink the wine?
Implausible	While Janet walked the rambunctious pug played at home.	Did Janet walk the pug?
Implausible	As the soldiers fought the enemies never appeared for battle.	Did the soldiers fight the enemies?
Implausible	While Jerry drew the complicated blueprint was shipped overseas.	Did Jerry draw the blueprint?
Implausible	As Caroline cleaned the pan stayed in the oven with the vegetables.	Did Caroline clean the pan?
Implausible	While Angela grilled the chicken kebabs cooled in the fridge.	Did Angela grill the kebabs?
Implausible	While Jack paddled the kayak leaned in the shed untouched.	Did Jack paddle the kayak?
Implausible	As Sam smoked the hookah sat unused on the highest shelf.	Did Sam smoke the hookah?
Implausible	As the eagle ate the sparrow sheltered safe and sound.	Did the eagle eat the sparrow?
Implausible	As the teacher lectured the parents missed the meeting unknowingly.	Did the teacher lecture the parents?
Implausible	As the boys counted the small dogs were nowhere to be seen.	Did the boys count the dogs?
Implausible	While the cowboy rode the bull stood by the fence unfazed.	Did the cowboy ride the bull?
Implausible	While Kendra juggled the torches dangled high on the wall.	Did Kendra juggle the torches?
Implausible	While Felicia whittled the pencil lay unsharpened in the box.	Did Felicia whittle the pencil?
Implausible	While Juan drove the monster truck remained in the empty lot.	Did Juan drive the truck?
Implausible	While the student read the book burned in the fireplace.	Did the student read the book?
Implausible	As the police investigated the heist occurred undetected at the store.	Did the police investigate the heist?
Implausible	As the hooting owl swallowed the rat hid in the deep hole.	Did the owl swallow the rat?
Implausible	As the men wrestled the rivals arrived at the gymnasium.	Did the men wrestle the rivals?
Implausible	While the traveler sailed the dinghy stayed by the shore.	Did the traveler sail the dinghy?
Implausible	While the dancer performed the ballet was canceled suddenly.	Did the dancer perform the ballet?
Implausible	While Bill hunted the lion paced in the zoo quickly.	Did Bill hunt the lion?
Implausible	As Mark ate the plump turkey roamed in its spacious cage.	Did Mark eat the turkey?
Implausible	While the intern photographed the workers hid from her.	Did the intern photograph the workers?
Implausible	While the brown cow fed the calf slept in the barn.	Did the brown cow feed the calf?
Implausible	While Susan studied the graphs resided unexamined on the board.	Did Susan study the graphs?
Implausible	As Klaus parked the sports car sped quickly away.	Did Klaus park the car?
Implausible	As Maria chewed the white bread laid uneaten in the food pantry.	Did Maria chew the bread?
Implausible	As Jamal baked the apple pie waited be made soon.	Did Jamal bake the pie?

(Continued)

CCWM Stimuli

Condition	Sentence	Question
Implausible	As Mike played the violin was placed in its case carefully.	Did Mike play the violin?
Implausible	While Amir vacuumed the dirty carpet spun in the dryer rapidly.	Did Amir vacuum the carpet?
Control	While Bill stirred the stew the soup boiled on the stove.	Did Bill stir the stew?
Control	While Jerry juggled the balls the props rested in the box.	Did Jerry juggle the balls?
Control	As the musician played the trumpet the oboe balanced on the stand.	Did the musician play the trumpet?
Control	While the zookeeper fed the ponies the stallion stomped its hoof.	Did the zookeeper feed the ponies?
Control	As Jamal cleaned the floor the dust blew away.	Did Jamal clean the floor?
Control	As the men parked the mopeds the tractor went in for repairs.	Did the men park the mopeds?
Control	As the teacher baked the cookies the cake was decorated.	Did the teacher bake the cookies?
Control	As the troops fought the invaders the civilians escaped to the shelter.	Did the troops fight the invaders?
Control	While Rick drove the limo the truck pulled over.	Did Rick drive the limo?
Control	As the woman wrestled the opponent the onlookers started yelling.	Did the woman wrestle the opponent?
Control	As Mark filmed the scene the actress took a break.	Did Mark film the scene?
Control	While the woman walked the poodle the pug slept in the house.	Did the woman walk the poodle?
Control	As the speaker lectured the audience the visitors returned home.	Did the speaker lecture the audience?
Control	While Janet wrote the comments the letter sat on the porch.	Did Janet write the comments?
Control	As Tom investigated the attack the suspect stayed in jail.	Did Tom investigate the attack?
Control	As the horse sniffed the hay the weeds swayed in the wind.	Did the horse sniff the hay?
Control	While Harry photographed the astronaut the probe landed on Mars.	Did Harry photograph the astronaut?
Control	As Sam painted the portrait the sculptor left for lunch.	Did Sam paint the portrait?
Control	As Mike drank the soda the tea brewed in the pot.	Did Mike drink the soda?
Control	While Angela sailed the schooner the jetski lingered at the dock.	Did Angela sail the schooner?
Control	While the women studied the diagram the report dangled on the bulletin board.	Did the women study the diagram?
Control	While Susan whittled the wood the twigs burned in the campfire.	Did Susan whittle the wood?
Control	While the students drew the landscape the vase held the flowers.	Did the students draw the landscape?
Control	While Amir steered the baler the cows chewed the grass.	Did Amir steer the baler?
Control	While the travelers rode the mules the camels drank water.	Did the travelers ride the mules?
Control	While the guests ordered the chicken the fish swam in the tank.	Did the guests order the chicken?
Control	As Henry counted the cars the girls formed a group.	Did Henry count the cars?
Control	As Klaus dusted the bookshelves the collectibles were clean.	Did Klaus dust the bookshelves?
Control	While Lola paddled the raft the canoe floated nearby.	Did Lola paddle the raft?
Control	As Kendra smoked the cigarette the cigar remained unlit.	Did Kendra smoke the cigarette?
Control	While Felicia grilled the steaks the ribs waited on the table.	Did Felicia grill the steaks?
Control	As Maria raced the competitor the team cheered in the stands.	Did Maria race the competitor?
Control	While the assistant typed the email the schedule remained the same.	Did the assistant type the email?
Control	As the workers ate the fruit the bread rose in the oven.	Did the workers eat the fruit?
Control	While Jack hunted the turkey the squirrel dashed into the bush.	Did Jack hunt the turkey?
Control	As Juan chewed the gum the meat thawed on the counter.	Did Juan chew the gum?
Control	As the hornets attacked the crowd the children ran to safety.	Did the hornets attack the crowd?
Control	While the boys vacuumed the stairs the rug lay on the couch.	Did the boys vacuum the stairs?
Control	While Caroline read the newspaper the documents lay in the drawer.	Did Caroline read the newspaper?
Control	While the Broadway stars performed the number the show impressed the patrons.	Did the Broadway stars perform the number?
Control	As the cat swallowed the lizard the birds flew away.	Did the cat swallow the lizard?

(*Continued*)

CCWM Stimuli

Condition	Sentence	Question
Control	As the hikers explored the forest the meadows were blooming with wildflowers.	Did the hikers explore the forest?
Plausible	While Rick drove the pick-up truck plunged into a ditch.	Did Rick drive the truck?
Plausible	As the athlete wrestled the tough opponent shouted insults.	Did the athlete wrestle the opponent?
Plausible	While the clown juggled the green apples dropped to the ground.	Did the clown juggle the apples?
Plausible	While Maria read the newspaper got wet from her hot coffee.	Did Maria read the newspaper?
Plausible	While the skipper sailed the cargo ship veered off course.	Did the skipper sail the cargo ship?
Plausible	While the rowdy cowboy rode the anxious horse neighed loudly.	Did the cowboy ride the horse?
Plausible	While the orchestra performed the symphony sounded beautiful.	Did the orchestra perform the symphony?
Plausible	While Klaus studied the chemistry textbook remained on the desk.	Did Klaus study the textbook?
Plausible	As Bill ate the ham lay on the dining room table.	Did Bill eat the ham?
Plausible	While the big man hunted the deer ran into the dark woods.	Did the big man hunt the deer?
Plausible	As the snake swallowed the white mouse kicked vigorously.	Did the snake swallow the mouse?
Plausible	As Harry chewed the ripe banana fell to the carpeted floor.	Did Harry chew the banana?
Plausible	As Janet baked the sourdough bread rose in the convection oven.	Did Janet bake the bread?
Plausible	While Mark vacuumed the drapes hung in the large window.	Did Mark vacuum the drapes?
Plausible	As the detective investigated the gruesome murder caused panic.	Did the detective investigate the murder?
Plausible	As Kendra parked the school bus bumped into the cement curb.	Did Kendra park the bus?
Plausible	While the reporter photographed the rocket sat on the launch pad.	Did the reporter photograph the rocket?
Plausible	While the old mare fed the small colt made happy noises.	Did the mare feed the colt?
Plausible	While Henry whittled the mahogany wood stick broke in half.	Did Henry whittle the stick?
Plausible	As Sam counted the little children boarded the long train.	Did Sam count the children?
Plausible	As Jerry played the grand piano echoed in the living room.	Did Jerry play the piano?
Plausible	As the tired sailor smoked the Cuban cigar glowed brightly.	Did the sailor smoke the cigar?
Plausible	As the vicious lion attacked the baboon screamed in terror.	Did the lion attack the baboon?
Plausible	As Lola dusted the wooden coffee table tipped over.	Did Lola dust the coffee table?
Plausible	While the farmer steered the green tractor pulled the plow.	Did the farmer steer the tractor?
Plausible	As Angela cleaned the restless dog stood in the backyard.	Did Angela clean the dog?
Plausible	As Caroline painted the model posed in the leather armchair.	Did Caroline paint the model?
Plausible	While Jamal walked the collie barked loudly.	Did Jamal walk the collie?
Plausible	While Susan wrote the love letter fluttered in the billowing wind.	Did Susan write the letter?
Plausible	As the puppy sniffed the new-born baby perched on the sofa.	Did the puppy sniff the baby?
Plausible	As the boy raced the challenger stumbled over the hurdle.	Did the boy race the challenger?
Plausible	While Amir paddled the small canoe headed towards a huge waterfall.	Did Amir paddle the canoe?
Plausible	As the cameraman filmed the actor recited the monologue.	Did the cameraman film the actor?
Plausible	While Mike stirred the hot tomato soup boiled over.	Did Mike stir the soup?
Plausible	As the distinguished scientists explored the deep cave grew cold.	Did the scientists explore the cave?
Plausible	While Jack ordered the white fish cooked in a large pot.	Did Jack order the fish?
Plausible	As the professor lectured Juan took detailed chemistry notes.	Did the professor lecture Juan?
Plausible	As the courageous warrior fought the evil enemy retreated from the battle.	Did the warrior fight the enemy?
Plausible	While Tom grilled the veggie burger burned to a crisp.	Did Tom grill the burger?
Plausible	While Felicia typed the urgent memo was almost fully complete.	Did Felicia type the memo?
Plausible	While the caricaturist drew the girl lingered on the sidewalk.	Did the caricaturist draw the girl?

(*Continued*)

CCWM Stimuli

Condition	Sentence	Question
Plausible	As the woman drank the steaming coffee spilled on the carpet.	Did the woman drink the coffee?
LongWM	Jean organized the extremely confidential files and clocked himself out for his lunch break.	Did Jean clock out for his lunch break?
LongWM	Last week, Bessie bought tickets for a Mediterranean cruise and educated herself about lifeboats.	Did Bessie sell tickets for a Mediterranean cruise?
LongWM	Sierra wore a bright orange silk scarf and introduced herself to the impatient audience.	Did Sierra wear a bright orange silk scarf?
LongWM	Susan got an oat milk latte and forced herself to study organic chemistry for an hour.	Did Susan study organic chemistry?
LongWM	Sarah went to run numerous errands and then washed herself in the jacuzzi bathtub.	Did Sarah run numerous errands?
LongWM	Josh went to see a foreign film and bought himself a variety of snacks.	Did Josh buy a variety of snacks?
LongWM	Jennifer moved into a brand new house and bought herself a bouquet of pink roses.	Did Jennifer buy a bouquet of pink lilies?
LongWM	James and Liam hurried not to miss the horse race and got themselves some ice cold water.	Did James and Liam sell ice cold water?
LongWM	Andrew grew a big bushy beard and bought himself a new linen suit.	Did Andrew buy a new wool suit?
LongWM	Randall and Alex played an intense game of beach volleyball and treated themselves to fresh squeezed orange juice.	Did Randall and Alex play an intense game of beach volleyball?
LongWM	Lottie ate a funnel cake with whipped cream and won herself a carnival prize.	Did Lottie eat a pound cake with whipped cream?
LongWM	Alexis swam in chilly Lake Tahoe and also taught herself how to play golf.	Did Alexis swim in chilly Lake Tahoe?
LongWM	Nicole scrolled on social media apps and poured herself a can of chicken noodle soup.	Did Nicole pour a can of broccoli cheddar soup?
LongWM	Gabe and Mel sat in the freshly painted kitchen together and made themselves ham and Swiss sandwiches.	Did Gabe and Mel make turkey and Swiss sandwiches?
LongWM	Lester and Anna took a long hike in the woods and also entertained themselves with a heated game of chess.	Did Lester and Anna take a long run in the woods?
LongWM	Araceli toured the community college campus and added herself to the electronic mailing list.	Did Araceli drive by the community college campus?
LongWM	Norris and Jessie edited their romance screenplays and taped themselves performing a skit.	Did Norris and Jessie edit their comedy screenplays?
LongWM	Alexi and Lena visited the natural history museum and called themselves an overpriced cab.	Did Alexi and Lena call an overpriced cab?
LongWM	Florence and Clare loaded up the red convertible and covered themselves in sunscreen lotion.	Did Florence and Clare cover themselves in tanning lotion?
LongWM	Johnny traveled to the southern coast of Spain and taught himself about Spanish culture.	Did Johnny move to the southern coast of Spain?
LongWM	Trisha packed a carry-on suitcase and checked herself in for her business class flight.	Did Trisha check into her business class flight?
LongWM	Ezekiel bragged about his circus art skills and recorded himself playing a Beethoven sonata on the piano.	Did Ezekiel play a Beethoven sonata?
LongWM	Jim put on a sleek groomsman outfit and admired himself in the vintage full-length mirror.	Did Jim purchase a vintage full-length mirror?
LongWM	Christopher folded the large load of laundry and drew himself a warm bubble bath.	Did Christopher draw a warm bubble bath?
LongWM	Shefali napped in the dimly lit bedroom and read Pride and Prejudice to herself very quietly.	Did Shefali nap in the brightly lit living room?
LongWM	On Saturday, Jayson went kayaking on the choppy river and later built himself a cedar wood bookshelf.	Did Jayson build a cedar wood bookshelf?
LongWM	Emilia and Rosie ordered chocolate peanut butter protein shakes and pushed themselves at the crossfit gym.	Did Emilia and Rosie order strawberry protein shakes?
LongWM	Curtis signed the six-month lease and moved himself into a new apartment complex.	Did Curtis move into an old apartment complex?
LongWM	Yusef got ready for bed and set himself a very early alarm for the morning.	Did Yusef get ready for bed?
LongWM	Leandro played at the outdoor water park and warmed himself in the scorching sun.	Did Leandro play the indoor water park?
LongWM	David ate at the Michelin star restaurant and also recommended himself as the DJ for the high school dance on Sunday.	Did David work at the Michelin star restaurant?
LongWM	Sammy wanted a vegetarian pizza and drove himself to the neighboring suburb.	Did Sammy want a vegetarian pizza?
LongWM	On Friday, Anthony and Carl shopped at the thrift store and enjoyed themselves at the engagement party.	Did Anthony and Carl shop at the department store?
LongWM	Brian went to get premium gas for his car and drove himself to an old industrial town.	Did Brian get premium gas for his car?
LongWM	Stacy worked out her quadriceps all day and hired herself a new personal trainer.	Did Stacy work out her biceps?
LongWM	Gayle wrapped Christmas gifts for her family and baked herself a loaf of banana nut bread.	Did Gayle wrap Christmas gifts?
LongWM	Sally watched the mourning doves in the backyard and knit herself a warm cashmere sweater.	Did Sally knit a warm cashmere scarf?
LongWM	Frank and Joey grilled plant-based hotdogs and made themselves raspberry lemonade.	Did Frank and Joey make raspberry lemonade?
LongWM	Dave walked in the empty park and made himself a lovely picnic.	Did Dave walk in the empty zoo?
LongWM	Olivia and Mia ate ice cream sundaes and bought themselves rock concert tickets.	Did Olivia and Mia buy concert tickets?
LongWM	Mindy carved a gigantic white pumpkin and dressed herself up in a superhero costume.	Did Mindy dress in a vampire costume?

(Continued)

CCWM Stimuli

Condition	Sentence	Question
LongWM	Idina joined the adult church choir and sang a lullaby to herself in the evening.	Did Idina sing a lullaby?
ShortWM	Grace hired herself a new personal trainer and worked out her quadriceps all day.	Did Grace hire a new personal trainer?
ShortWM	Nadine baked herself a loaf of banana nut bread and wrapped Christmas gifts for her family.	Did Nadine bake a loaf of zucchini bread?
ShortWM	Bradley set himself a very early alarm for the morning and got ready for bed.	Did Bradley set a very early alarm?
ShortWM	Colin and Nick made themselves raspberry lemonade and grilled plant-based hotdogs.	Did Colin and Nick buy raspberry lemonade?
ShortWM	Phil drove himself to an old industrial town and went to get premium gas for his car.	Did Phil take the train to an old industrial town?
ShortWM	Gerard drew himself a warm bubble bath and folded the large load of laundry.	Did Gerard draw a scalding hot bath?
ShortWM	Nancy read Pride and Prejudice to herself very quietly and napped in the dimly lit bedroom.	Did Nancy nap in the dimly lit bedroom?
ShortWM	Tina and Tracy bought themselves rock concert tickets and ate ice cream sundaes.	Did Tina and Tracy eat ice cream cake?
ShortWM	Eliza sang a lullaby to herself in the evening and joined the adult church choir.	Did Eliza join the adult church choir?
ShortWM	On Friday, Deshaun and Mark enjoyed themselves at the engagement party and shopped at the thrift store.	Did Deshaun and Mark work at the thrift store?
ShortWM	Ivan drove himself to the neighboring suburb and wanted pepperoni and mushroom pizza.	Did Ivan drive to a neighboring city?
ShortWM	Helen and Elise pushed themselves at the crossfit gym and ordered chocolate peanut butter protein shakes.	Did Helen and Elise order chocolate peanut butter protein shakes?
ShortWM	Markus recorded himself playing a Beethoven sonata on the piano and bragged about his circus arts skills.	Did Markus brag about his circus arts skills?
ShortWM	On Saturday, Eduardo built himself a cedar wood bookshelf and later went kayaking on the choppy river.	Did Eduardo build a cedar wood bookshelf?
ShortWM	Yuta admired himself in the vintage full-length mirror and put on a sleek groomsman outfit	Did Yuta put on a sleek groomsman outfit?
ShortWM	Maria knit herself a warm cashmere sweater and watched the mourning doves in the backyard.	Did Maria photograph the mourning doves?
ShortWM	Patrick moved himself into a new apartment complex and signed the six-month lease.	Did Patrick sign the six-month lease?
ShortWM	Matthew warmed himself in the scorching sun and played at the outdoor water park.	Did Matthew play at the outdoor amusement park?
ShortWM	Jack made himself a lovely picnic and walked in the empty park.	Did Jack make a lovely picnic?
ShortWM	Robyn dressed herself up in a superhero costume and carved a gigantic white pumpkin.	Did Robyn carry a gigantic white pumpkin?
ShortWM	José recommended himself as the DJ for the high school dance on Sunday and also ate at the Michelin Star restaurant.	Did José eat at the Michelin star restaurant?
ShortWM	Jeremy and Byron taped themselves performing a skit and edited their romance screenplays.	Did Jeremy and Byron tape a skit?
ShortWM	Keith clocked himself out for his lunch break and organized the extremely confidential files.	Did Keith clock out to go home?
ShortWM	Tim and Bruce made themselves ham and Swiss sandwiches and sat in the freshly painted kitchen together.	Did Tim and Bruce make bologna and cheese sandwiches?
ShortWM	Fiona won herself a carnival prize and ate a funnel cake with whipped cream.	Did Fiona win a carnival prize?
ShortWM	Natalie and Margaret called themselves an overpriced cab and visited the natural history museum.	Did Natalie and Margaret call an overpriced cab?
ShortWM	Tyler bought himself new linen suit and grew a big bushy beard.	Did Tyler buy a new linen suit?
ShortWM	Finneas and Pete treated themselves to fresh squeezed orange juice and played an intense game of beach volleyball.	Did Finneas and Pete make fresh squeezed lemonade?
ShortWM	Sandy and Steve entertained themselves with a heated game of chess and also took a long hike in the woods.	Did Sandy and Steve play a friendly game of chess?
ShortWM	Rosalina bought herself a bouquet of pink roses and moved into a brand new house.	Did Rosalina buy a bouquet of purple tulips?
ShortWM	Emily poured herself a can of chicken noodle soup and scrolled on social media apps.	Did Emily scroll on social media apps?
ShortWM	Jamie taught herself how to play golf and also swam in chilly Lake Tahoe.	Did Jamie sail in chilly Lake Tahoe?
ShortWM	Lucy added herself to the electronic mailing list and toured the community college campus.	Did Lucy tour the community college campus?
ShortWM	Timmy taught himself about Spanish culture and traveled to the southern coast of Spain.	Did Timmy travel to the northern coast of Spain?
ShortWM	Violet washed herself in the jacuzzi bathtub and then went to run numerous errands.	Did Violet wash up in the jacuzzi bathtub?
ShortWM	Cynthia forced herself to study organic chemistry for an hour and got an oat milk latte.	Did Cynthia make an oat milk latte?
ShortWM	Last week, Sherri educated herself about lifeboats and bought tickets for a Mediterranean cruise.	Did Sherri buy tickets for a Caribbean cruise?
ShortWM	Louis and Rick got themselves some ice cold water and hurried not to miss the horse race.	Did Louis and Rick get themselves ice cold water?
ShortWM	Luke bought himself a variety of snacks and went to see a foreign film.	Did Luke see a Broadway musical?
ShortWM	Tony and Franklin covered themselves in sunscreen lotion and loaded up the red convertible.	Did Tony and Franklin load up the red convertible?
ShortWM	Dorothy introduced herself to the impatient audience and wore a bright orange silk scarf.	Did Dorothy wear a thick green wool scarf?
ShortWM	Bridget checked herself in for her business class flight and packed a carry-on suitcase.	Did Bridget check into her business class flight?

Attention Stimuli

Condition	Sentence
More/Less	More than half of the apples are to the left of the lamp.
More/Less	More than half of the balls are to the right of the lamp.
More/Less	More than half of the bananas are to the left of the lamp.
More/Less	More than half of the bears are to the right of the lamp.
More/Less	More than half of the bells are to the left of the lamp.
More/Less	More than half of the birds are to the right of the lamp.
More/Less	More than half of the candles are to the right of the lamp.
More/Less	More than half of the caterpillars are to the left of the lamp.
More/Less	More than half of the cats are to the right of the lamp.
More/Less	More than half of the dogs are to the left of the lamp.
More/Less	More than half of the footballs are to the right of the lamp.
More/Less	More than half of the giraffes are to the right of the lamp
More/Less	More than half of the glasses are to the right of the lamp.
More/Less	More than half of the hats are to the left of the lamp.
More/Less	More than half of the hearts are to the left of the lamp.
More/Less	More than half of the mugs are to the left of the lamp.
More/Less	More than half of the pencils are to the right of the lamp.
More/Less	More than half of the penguins are to the right of the lamp.
More/Less	More than half of the shoes are to the left of the lamp.
More/Less	More than half of the sunglasses are to the left of the lamp.
More/Less	More than half of the trees are to the right of the lamp.
More/Less	Less than half of the bikes are to the left of the lamp.
More/Less	Less than half of the books are to the left of the lamp
More/Less	Less than half of the butterflies are to the right of the lamp.
More/Less	Less than half of the carrots are to the left of the lamp.
More/Less	Less than half of the cars are to the left of the lamp.
More/Less	Less than half of the chairs are to the right of the lamp.
More/Less	Less than half of the cupcakes are to the left of the lamp.
More/Less	Less than half of the dolls are to the left of the lamp.
More/Less	Less than half of the flowers are to the right of the lamp.
More/Less	Less than half of the ladybugs are to the right of the lamp.
More/Less	Less than half of the leaves are to the right of the lamp.
More/Less	Less than half of the mittens are to the right of the lamp.
More/Less	Less than half of the pears are to the left of the lamp.
More/Less	Less than half of the phones are to the right of the lamp.
More/Less	Less than half of the pies are to the left of the lamp.
More/Less	Less than half of the pigs are to the right of the lamp.
More/Less	Less than half of the pillows are to the left of the lamp.
More/Less	Less than half of the planes are to the left of the lamp.
More/Less	Less than half of the rabbits are to the left of the lamp.

(Continued)

Attention Stimuli

Condition	Sentence
More/Less	Less than half of the sheep are to the right of the lamp.
More/Less	Less than half of the shirts are to the right of the lamp.
All/Some	All of the balls shown here are to the left of the lamp
All/Some	All of the bells shown here are to the right of the lamp.
All/Some	All of the birds shown here are to the left of the lamp
All/Some	All of the candles shown here are to the left of the lamp.
All/Some	All of the caterpillars shown here are to the right of the lamp.
All/Some	All of the chairs shown here are to the left of the lamp.
All/Some	All of the cupcakes shown here are to the right of the lamp.
All/Some	All of the dogs shown here are to the left of the lamp.
All/Some	All of the footballs shown here are to the right of the lamp.
All/Some	All of the giraffes shown here are to the left of the lamp.
All/Some	All of the ladybugs shown here are to the left of the lamp.
All/Some	All of the mugs shown here are to the left of the lamp.
All/Some	All of the pencils shown here are to the left of the lamp.
All/Some	All of the penguins shown here are to the right of the lamp
All/Some	All of the phones shown here are to the right of the lamp.
All/Some	All of the planes shown here are to the right of the lamp.
All/Some	All of the rabbits shown here are to the right of the lamp.
All/Some	All of the sheep shown here are to the right of the lamp.
All/Some	All of the shirts shown here are to the right of the lamp.
All/Some	All of the shoes shown here are to the right of the lamp.
All/Some	All of the trees shown here are to the left of the lamp.
All/Some	Some of the apples shown here are to the right of the lamp.
All/Some	Some of the bananas shown here are to the right of the lamp.
All/Some	Some of the bears shown here are to the left of the lamp.
All/Some	Some of the bikes shown here are to the right of the lamp.
All/Some	Some of the books shown here are to the right of the lamp.
All/Some	Some of the butterflies shown here are to the left of the lamp.
All/Some	Some of the carrots shown here are to the left of the lamp.
All/Some	Some of the cars shown here are to the right of the lamp.
All/Some	Some of the cats shown here are to the right of the lamp.
All/Some	Some of the dolls shown here are to the left of the lamp.
All/Some	Some of the flowers shown here are to the left of the lamp.
All/Some	Some of the glasses shown here are to the right of the lamp.
All/Some	Some of the hats shown here are to the left of the lamp.
All/Some	Some of the hearts shown here are to the right of the lamp.
All/Some	Some of the leaves shown here are to the right of the lamp.
All/Some	Some of the mittens shown here are to the left of the lamp.
All/Some	Some of the pears shown here are to the left of the lamp.
All/Some	Some of the pies shown here are to the left of the lamp.
All/Some	Some of the pigs shown here are to the left of the lamp.
All/Some	Some of the pillows shown here are to the right of the lamp.
All/Some	Some of the sunglasses shown here are to the left of the lamp.

References

Altmann GT, Kamide Y. Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*. 1999:73(3): 247–264.

Blanco-Elorrieta E, Pylkkänen L. Composition of complex numbers: delineating the computational role of the left anterior temporal lobe. *NeuroImage*. 2016:124(Pt A):194–203.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. Conflict monitoring and cognitive control. *Psychol Rev*. 2001:108(3): 624–652.

Caplan D, Vijayan S, Kuperberg G, West C, Waters G, Greve D, Dale AM. Vascular responses to syntactic processing: event-related fMRI study of relative clauses. *Hum Brain Mapp*. 2002:15(1): 26–38.

Christianson K, Hollingworth A, Halliwell JF, Ferreira F. Thematic roles assigned along the garden path linger. *Cogn Psychol*. 2001: 42(4):368–407.

Christianson K, Williams CC, Zacks RT, Ferreira F. Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Process*. 2006:42(2):205–238.

Dapretto M, Bookheimer SY. Form and content: dissociating syntax and semantics in sentence comprehension. *Neuron*. 1999:24(2): 427–432.

Davies M. *The Corpus of Contemporary American English* (COCA). 2008. Available online at https://www.english-corpora.org/coca/.

den Ouden DB, Dickey MW, Anderson C, Christianson K. Neural correlates of early-closure garden-path processing: effects of prosody and plausibility. *Q J Exp Psychol*. 2016:69(5):926–949.

Diachek E, Blank I, Siegelman M, Affourtit J, Fedorenko E. The domain-general multiple demand (MD) network does not support core aspects of language comprehension: a large-scale fMRI investigation. *J Neurosci*. 2020:40(23):4536–4550.

Dillon B. Syntactic memory in the comprehension of reflexive dependencies: an overview. *Language Linguistics Compass*. 2014:8(5): 171–187.

Embick D, Marantz A, Miyashita Y, O'Neil W, Sakai KL. A syntactic specialization for Broca's area. *Proc Natl Acad Sci*. 2000:97(11): 6150–6154.

Fedorenko E, Duncan J, Kanwisher N. Language-selective and domain-general regions lie side by side within Broca's area. *Curr Biol*. 2012:22(21):2059–2062.

Fedorenko E, Duncan J, Kanwisher N. Broad domain generality in focal regions of frontal and parietal cortex. *Proc Natl Acad Sci*. 2013:110(41):16616–16621.

Ferreira F. The misinterpretation of noncanonical sentences. *Cogn Psychol*. 2003:47(2):164–203.

Fiebach CJ, Schlesewsky M, Lohmann G, Von Cramon DY, Friederici AD. Revisiting the role of Broca's area in sentence processing: syntactic integration versus syntactic working memory. *Hum Brain Mapp*. 2005:24(2):79–91.

Fridriksson J, den Ouden DB, Hillis AE, Hickok G, Rorden C, Basilakos A, et al. Anatomy of aphasia revisited. *Brain*. 2018:141(3):848–862.

Friederici AD. Towards a neural basis of auditory sentence processing. *Trends Cogn Sci*. 2002:6(2):78–84.

Friedman NP, Miyake A. Unity and diversity of executive functions: individual differences as a window on cognitive structure. *Cortex*. 2017:86:186–204.

Grinband J, Wager TD, Lindquist M, Ferrera VP, Hirsch J. Detection of time-varying signals in event-related fMRI designs. *Neuroimage*. 2008:43(3):509–520.

Hagoort P. On Broca, brain, and binding: a new framework. *Trends Cogn Sci*. 2005:9(9):416–423.

Hammer A, Jansma BM, Tempelmann C, Münte TF. Neural mechanisms of anaphoric reference revealed by fMRI. *Front Psychol*. 2011:2:32.

Heim S, McMillan CT, Clark R, Baehr L, Ternes K, Olm C, et al. How the brain learns how few are "many": an fMRI study of the flexibility of quantifier semantics. *NeuroImage*. 2016:125:45–52.

Horne A, Zahn R, Najera OI, Martin RC. Semantic working memory predicts sentence comprehension performance: a case series approach. *Front Psychol*. 2022:13:887586.

Hsu NS, Jaeggi SM, Novick JM. A common neural hub resolves syntactic and non-syntactic conflict through cooperation with task-specific networks. *Brain Lang*. 2017:166:63–77.

Huang YT, Snedeker J. Some inferences still take time: prosody, predictability, and the speed of scalar implicatures. *Cogn Psychol*. 2018:102:105–126.

Just MA, Carpenter PA. A capacity theory of comprehension: individual differences in working memory. *Psychol Rev*. 1992:99(1): 122–149.

Kim A, Osterhout L. The independence of combinatory semantic processing: evidence from event-related potentials. *J Mem Lang*. 2005:52(2):205–225.

Lambon Ralph MA, Sage K, Jones RW, Mayberry EJ. Coherent concepts are computed in the anterior temporal lobes. *Proc Natl Acad Sci*. 2010:107(6):2717–2722.

Lewis RL, Vasishth S, Van Dyke JA. Computational principles of working memory in sentence comprehension. *Trends Cogn Sci*. 2006:10(10):447–454.

Lipkin B, Tuckute G, Affourtit J, Small H, Mineroff Z, Kean H, Jouravlev O, Rakocevic L, Pritchett B, Siegelman M, Hoeflin C. Probabilistic atlas for the language network based on precision fMRI data from >800 individuals. *Scientific Data*. 2022:9(1):529.

Makuuchi M, Bahlmann J, Anwander A, Friederici AD. Segregating the core computational faculty of human language from working memory. *Proc Natl Acad Sci*. 2009:106(20):8362–8367.

Matchin W, Basilakos A, Stark BC, den Ouden DB, Fridriksson J, Hickok G. Agrammatism and paragrammatism: a cortical double dissociation revealed by lesion-symptom mapping. *Neurobiol Lang*. 2020:1(2):208–225.

Novick JM, Trueswell JC, Thompson-Schill SL. Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affect Behav Neurosci*. 2005:5(3):263–281.

Olm CA, McMillan CT, Spotorno N, Clark R, Grossman M. The relative contributions of frontal and parietal cortex for generalized quantifier comprehension. *Front Hum Neurosci*. 2014:8:610.

Peelle J, Wingfield A. How our brains make sense of noisy speech. *Acoustics Today*. 2022:18(3):40–48.

Qian Z, Garnsey S, Christianson KA. Comparison of online and offline measures of good-enough processing in garden-path sentences. *Language, Cognition and Neuroscience*. 2018:33(2):227–254.

Sassenhagen J, Alday PM. A common misapplication of statistical inference: nuisance control with null-hypothesis significance tests. *Brain Lang*. 2016:162:42–45.

Shain C, Blank IA, Fedorenko E, Gibson E, Schuler W. Robust effects of working memory demand during naturalistic language comprehension in language-selective cortex. *J Neurosci*. 2022:42(39): 7412–7430.

Sharer K, Thothathiri M. Neural mechanisms underlying the dynamic updating of native language. *Neurobiology of Language*. 2020:1(4):492–522.

Sheppard SM, Love T, Midgley KJ, Holcomb PJ, Shapiro LP. Electrophysiology of prosodic and lexical-semantic processing during sentence comprehension in aphasia. *Neuropsychologia*. 2017:107: 9–24.

Slattery TJ, Sturt P, Christianson K, Yoshida M, Ferreira F. Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *J Mem Lang*. 2013:69(2):104–120.

Thothathiri M, Kim A, Trueswell JC, Thompson-Schill SL. Parametric effects of syntactic–semantic conflict in Broca's area during sentence processing. *Brain Lang*. 2012a:120(3):259–264.

Thothathiri M, Kimberg DY, Schwartz MF. The neural basis of reversible sentence comprehension: evidence from voxel-based lesion symptom mapping in aphasia. *J Cogn Neurosci*. 2012b:24(1): 212–222.

Thothathiri M, Gagliardi M, Schwartz MF. Subdivision of frontal cortex mechanisms for language production in aphasia. *Neuropsychologia*. 2012c:50(14):3284–3294.

Thothathiri M, Asaro CT, Hsu NS, Novick JM. Who did what? A causal role for cognitive control in thematic role assignment during sentence comprehension. *Cognition*. 2018:178:162–177.

Van De Meerendonk N, Kolk HH, Vissers CTW, Chwilla DH. Monitoring in language perception: mild and strong conflicts elicit different ERP patterns. *J Cogn Neurosci*. 2010:22(1):67–82.

Vuong LC, Martin RC. The role of LIFG-based executive control in sentence comprehension. *Cognitive Neuropsychology*. 2015:32(5): 243–265.

Walenski M, Europa E, Caplan D, Thompson CK. Neural networks for sentence comprehension and production: an ALE-based meta-analysis of neuroimaging studies. *Hum Brain Mapp*. 2019:40(8): 2275–2304.

Wartenburger I, Heekeren HR, Burchert F, Heinemann S, De Bleser R, Villringer A. Neural correlates of syntactic transformations. *Hum Brain Mapp*. 2004:22(1):72–81.

Westerlund M, Pylkkänen L. The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia*. 2014:57:59–70.

Ye Z, Zhou X. Conflict control during sentence comprehension: fMRI evidence. *NeuroImage*. 2009:48(1):280–290.

Yokoyama S, Watanabe J, Iwata K, Ikuta N, Haji T, Usui N, et al. Is Broca's area involved in the processing of passive sentences? An event-related fMRI study. *Neuropsychologia*. 2007:45(5): 989–996.