# Evidence from Formal Logical Reasoning Reveals that the Language of Thought is not Natural Language

Hope Kean*[,1,2], Alexander Fung[†,1,2], Paris Jaggers[†,3], Jason Chen[1], Joshua S. Rule[1,5], Yael Benn[4], Joshua B. Tenenbaum[1], Steven T. Piantadosi[5], Rosemary A. Varley[3], Evelina Fedorenko[1,2]

[1]Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology
[2]McGovern Institute for Brain Research, Massachusetts Institute of Technology
[3]Department of Psychology and Language Sciences, University College London
[4]Department of Psychology, Manchester Metropolitan University
[5]Department of Psychology, UC Berkeley
*Corresponding author. Email: hopekean@mit.edu
[†]equal contributors

**Acknowledgements**

We would like to acknowledge the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT, including the technical team—Steve Shannon and Atsushi Takahashi. We would also like to thank Anya Ivanova, Colton Casto, Ben Lipkin, Chiebuka Ohams, and Ted Gibson for helpful comments on the manuscript. HK was supported by a Friends of McGovern graduate fellowship and a graduate fellowship from the K. Lisa Yang Integrative Computational Neuroscience (ICoN) Center. JBT was supported by research funds from the Air Force Office of Scientific Research, the Office of Naval Research Science of AI program, and a Schmidt Sciences AI2050 Fellowship. JSR and STP were supported by the Division of Research on Learning in Formal and Informal Settings (EDU/DRL) grant 2201843. RV was supported by a Leverhume Research Fellowship (RF-2023-690\10). EF and this research were partially supported by research funds from the McGovern Institute for Brain Research, the Department of Brain and Cognitive Sciences, MIT's Quest for Intelligence, a grant from the Simons Foundation to the Simons Center for the Social Brain at MIT, and a grant from the DARPA AIQ program through the DARPA CMO contract number HR00112520025.

**Keywords**
Language of Thought, Language, Logical Reasoning, Aphasia, fMRI, Cognitive Neuroscience

**Abstract**

Humans are endowed with a powerful capacity for both inductive and deductive logical thought: we easily form generalizations based on a few examples and draw conclusions from known premises. Humans also arguably have the most sophisticated communication system in the animal kingdom: natural language allows us to express complex and structured meanings. Some have therefore argued for a tight relationship between complex thought and language, postulating that reasoning, including logical reasoning, relies on linguistic representations. We systematically investigated the relationship between logical reasoning and language using two complementary approaches. First, we used non-invasive brain imaging (fMRI) to examine neural activity as healthy adults engaged in inductive and deductive logical reasoning tasks. And second, we behaviorally evaluated logical abilities in individuals with extensive lesions to the language brain areas and consequent severe linguistic impairment. Our findings reveal that the language brain network is not engaged during logical reasoning, and patients with severe aphasia exhibit intact performance on logic tasks. Instead, inductive reasoning recruits the domain-general multiple demand network implicated broadly in goal-directed behaviors, whereas deductive reasoning draws on brain regions that are distinct from both the language and the multiple demand networks. Together, these results indicate that linguistic representations are neither utilized nor required for inductive or deductive logical reasoning.

**Significance**

Which cognitive mechanisms allow humans to reason logically, to understand whether a conclusion follows from the premises? Are they the same ones that allow the assembly of words into structured representations? Scholars have debated for millennia whether logical reasoning is inextricably tied to natural language, or instead relies on a distinct "language of thought" (LOT). Using fMRI in healthy adults and evaluating logical ability in individuals with severe aphasia, we find that distinct neural systems support language processing vs. logical (inductive and deductive) reasoning. These results establish that language does not underpin logical inference and point to distinct representational systems for the logical LOT. This work contributes to our understanding of the division of cognitive labor in the human brain.

## Introduction

In 350 BC, Aristotle wrote that "writing and speech are not the same for all people, but mental acts themselves, of which words signify, are the same for all people" (1). This reflection highlights a distinction between linguistic symbols and the underlying cognitive structures they aim to express. But over the years, many philosophers and linguists have argued that natural language is, in fact, the medium of complex thought (e.g. (2–13)); for counter-arguments, see (14, 15). We here challenge this view using evidence from human neuroscience.

A key argument for language underlying thinking has to do with similarities between them. For example, one prominent proposal about thought—the "language of thought" (LOT) hypothesis (16)—has emphasized the compositional and hierarchical nature of thoughts. According to this hypothesis, thoughts are composed of smaller atomic pieces in a structured format with hierarchical relations among the component elements. A parallel to natural language is easy to draw: sentences are built out of words, which are related to one another hierarchically. But a parallel can also be drawn to computer programs, which are built out of a small collection of primitive operations, or to melodies, which are composed of hierarchically-related notes and chords. So, this similarity alone is insufficient to conclude that we think in language. Furthermore, although many thoughts can be cast into natural language (which is what makes language an effective communication system), we can also express ideas—including abstract ones—using non-linguistic symbols, mathematical expressions, visual schematics or diagrams, and so on. Therefore, a degree of isomorphism between certain thoughts and linguistic expressions need not entail that linguistic representations are used for thinking, any more than the possibility of expressing certain ideas using visual images entails that our thinking is visual in nature.

Indeed, empirical data have been accumulating that suggest that linguistic representations are neither utilized nor necessary for thinking. Patients with aphasia appear capable of diverse forms of thought, as evidenced by their intact performance on tasks requiring mathematical reasoning (17), causal reasoning (18), and Theory of Mind (19, 20); (see (21) and (22) for reviews). Converging evidence for the separability of language and reasoning comes from brain imaging studies. The brain areas that support language comprehension and production (23–25) are strongly sensitive to linguistic syntactic structure (26–32) but are not engaged during non-linguistic cognitively demanding tasks, including those that require representing and manipulating structured representations (see (33) for a review). These tasks span solving arithmetic and algebraic problems (34–36), understanding computer code (37, 38), social reasoning (39–42), and performing demanding executive function tasks (34, 43, 44), which require skills linked to fluid intelligence (45, 46).

The separability of language and thought has been questioned by some researchers based on a) correlations between the degree of linguistic impairment and impairment on cognitive tasks in aphasia (e.g., (47–50)) and b) evidence from verbal interference dual-task paradigms (for a review, see (51)). We believe that these findings have straightforward alternative explanations and thus do not present a challenge for the claim that language and thought rely on distinct resources (see Discussion). However, there is one domain of thought whose relationship with language has only received limited attention in the prior literature, despite it being a hallmark domain where the role of language has been emphasized (12, 52). This domain is *abstract logical reasoning*.

One prior fMRI study examined the relationship between language and deductive reasoning by comparing neural responses during a logic inference task (judging the validity of a conclusion given a premise) versus a linguistic inference task (judging the validity of thematic roles across two sentences) relative to lower-level control conditions (judging expression well-formedness) (53). A partial dissociation was observed: although both the logic and the language contrast elicited left-lateralized responses in the frontal, temporal, and parietal cortex, some brain areas—in the vicinity of areas traditionally implicated in language processing (54–57, 33)—showed a stronger response to linguistic inference, whereas other areas responded more strongly during logic inference (see (58) for concordant evidence from TMS). However, recent evidence suggests that the kind of linguistic paradigm used by Monti et al. (53) engages both language-processing areas but also brain areas sensitive to general cognitive effort (e.g., (59)), which makes previously observed dissociations more difficult to interpret. Another past study examined logical and mathematical reasoning and argued that neither engaged the language areas (52), but that study did not include a language task, which would be necessary to make such a claim. Thus, these prior findings are intriguing, but have not conclusively answered the question of whether logical reasoning relies on linguistic processing mechanisms.

Two other bodies of work speak to the logic-language relationship. The first comes from developmental psychology and suggests that the timelines of linguistic and logic development diverge. Inductive reasoning appears to come online very early (e.g., (60, 61)). The representations are argued to be sensorimotor, not linguistic, in nature, and the main debate concerns their domain specificity (e.g., (62, 63)) vs. generality (e.g., (64, 65)) (e.g., do infants have distinct representations for reasoning about the physical world vs. about social agents?). For deductive reasoning, the picture is less clear: some have argued that preverbal infants can already reason disjunctively (66), which implies a separation of logical and linguistic ability. However, others have reported failures in simple disjunctive inferences as late as age 2.5 years (67), and even later for more complex inferences (68). These late failures may suggest that a certain level of linguistic competence is required for the development of these capacities, although they may also reflect the slow development of executive abilities (69–71), which may be needed to deal with the task demands in complex paradigms.

The second body of work comes from artificial intelligence, where recent advances have provided additional motivation for investigating the relationship between language and logic. In particular, neural network language models (72–74) not only achieve human-level performance on diverse language tasks (75–77), but also exhibit impressive successes on certain reasoning tasks (e.g., ARC: (78); BIG-Bench: (79); LogiQA: (80); MATH: (81); WildBench: (82); Templates: (83)). These findings beg the question of whether linguistic competence inevitably leads to logical ability, although many have pointed out that reasoning in large language models lacks robustness and generalizability (e.g., (84–89)).

Thus, the relationship between language and logic—and more specifically, whether abstract logical reasoning relies on linguistic representations in humans—remains unclear. To systematically examine the role of language in human logical reasoning, we adopted a two-pronged approach. We first used fMRI in healthy adults to test whether the language network (33) responds to logical reasoning demands (Study 1). Next, we used behavioral experiments in individuals with severe aphasia resulting from extensive damage to left peri-Sylvian cortex, to test whether linguistic

ability is necessary for logical reasoning (Study 2). Note that our research does not speak to the potential role of language in the acquisition of logical concepts or in learning to reason during childhood (we briefly touch on this point in the Discussion). To foreshadow our findings, we find no evidence for the idea that abstract logical reasoning draws on linguistic resources.

**Results**

In each study, we examined two paradigmatic forms of logical reasoning: inductive and deductive reasoning. The ***inductive reasoning*** paradigm (**Figure 1**, left) was adapted from Rule et al. (90) (for related paradigms, see (91–95)). Participants were presented with an input number list and an output list (e.g., `[5, 7] → [7, 5, 7]`) and asked to infer the rule that governs the input-to-output transformation. They could then test their hypothesis on a new input list, and so on, until they guess the correct rule. The rules involve a combination of mathematical operations (e.g., $f(l) = [\, x + 2 \mid x \in l \,]$), "add 2 to every number"), list operations (e.g., $f(l) = [\, x_i \mid x_i \in l, \forall j < i, x_i \neq x_j \,]$, "leave only unique, non-repeated elements"), and structural operations (e.g., $f(l) = l[3{:}]\hat{}l[0{:}3]$, "rotate the list by three elements"), each of which can be written as a short computer program. The fMRI version also included a control condition where participants were told the rule and asked to apply it to new input lists.

For ***deductive reasoning***, we used two paradigms: syllogistic reasoning and matrix reasoning. The first paradigm was introduced in Coetzee & Monti (96) and was only used here for the fMRI component of the study (**Figure 1**, center). Participants were presented with a classic syllogism consisting of two premises and a conclusion (e.g., i. If A is B, then A is also C. ii. A is not C. iii. Therefore, A is not B) and were asked to judge the validity of the conclusion. Half of the syllogisms used real words (e.g., "If the block is large..."), and the other half used nonwords (e.g., "If the tep is ag…"). Although deductive reasoning and language processing are required for all problems, the problems critically vary in the difficulty of the deduction. The easier problems, known as Modus Ponens, had the following forms: "If the block is large, then it is not yellow. The block is large. So, the block is not yellow." (the conclusion follows from the premises), or "If the block is large, then is it not yellow. The block is large. So, the block is yellow." (the conclusion does not follow from the premises). The more complex problems, known as Modus Tollens, had the following forms: "If the block is large, then it is not yellow. The block is yellow. So, the block is not large." (the conclusion follows from the premises), or "If the block is large then it is not yellow. The block is not yellow. So, the ball is large." (the conclusion does not follow from the premises). This is a classic paradigm in logical reasoning research (97–101) and the contrast between the Modus Tollens versus Modus Ponens problems elegantly isolates deductive demands. However, this task is not suitable for patients with aphasia because it uses verbal stimuli.

As a result, for the second fMRI paradigm, we chose a non-verbal task, which could also be performed by the patients: matrix reasoning (102) (for prior fMRI use, see (103)). Participants were presented with a matrix of four geometric patterns and asked to find an outlier (**Figure 1**, right). Patients with aphasia performed a similar version (104), except that a matrix of geometric patterns was missing a pattern, and participants had to decide which option from a set of possible answers completes the matrix (**Figure 1**, bottom right panel). This type of paradigm, commonly used in assessing the fluid reasoning capacity across diverse populations (105–114), taxes both deductive and inductive reasoning (115, 116). Inductive reasoning is needed to make guesses about the

underlying rules, and deductive reasoning—to evaluate logical constraints and eliminate incorrect options. Thus, although not as pure as the syllogistic reasoning task, this paradigm is still well-suited to test the necessity of language for deductive reasoning.

If linguistic representations underlie logical inference, then a) the language brain areas should be sensitive to inductive demands, deductive demands, or both, as measured with fMRI; and b) patients with profound aphasia should show impairments on these tasks.

**The language network is not engaged during inductive nor deductive reasoning.**

The language-processing brain areas show little or no response to the inductive and deductive reasoning contrasts. As expected based on much prior work (e.g., see (33) for a review), these areas showed a robust response to language processing, evidenced by a stronger response during the sentence compared to the nonword-list condition estimated in independent data ($p<0.001$ at the network level; **Figure 2B**, **Table 1**). Note that although a particular localizer contrast is used here to define the language areas, this contrast is highly generalizable across presentation modalities, languages, stimuli, and tasks (e.g., (117, 118, 43, 119)), and this network can also be defined without a contrast, using patterns of functional connectivity (e.g., (120–122)). Moreover, these language-responsive areas have been shown to be robustly sensitive to linguistic processing demands in both comprehension and production, and across controlled and naturalistic paradigms (review: (33)), and this ubiquity of engagement across diverse language tasks aligns with the fact that damage to these areas leads to language deficits (55, 123).

Critically, the language areas showed little or no sensitivity to the logic contrasts. The inductive contrast did elicit a reliable response at the network level ($p<0.05$), but the effect was small, with the language contrast being over four times stronger, and the critical induction condition eliciting a response at or below the level of the control condition of the language task (reading nonwords) (**SI Figure 1b**). For deductive reasoning, neither the syllogistic nor the matrix reasoning contrast elicited a significant response in the language areas ($p$s>0.1 at the network level; **Figure 2A-B**, **Table 1**; see **SI Figure 1a** and **SI Table 1** for evidence that the results are similar for the five language areas examined separately), despite significant behavioral difficulty effects in both accuracies and reaction times for both tasks (**SI Figure 4**). For the syllogistic task, the responses relative to the low-level baseline were strong to both the Modus Ponens and the Modus Tollens conditions, which is expected given that these conditions require lexical and syntactic processing; **SI Figure 1B**), and the critical result of no sensitivity to deductive reasoning difficulty generalized across words and nonwords (**SI Figure 5**). Interestingly, the conditions that use real words did not elicit a stronger response than the conditions that use nonwords in the language areas. The nonword conditions ("If the tep is ag…") are similar to the so-called Jabberwocky sentences (124), which in past fMRI studies have been shown to elicit a lower response than real sentences in the language areas (e.g., (32, 117, 125, 126)). We speculate that we do not see this effect here because the words' meanings are irrelevant to the task, and, as a result, participants treat both words and nonwords as variables and parse the sentences in similar ways, extracting basic relational structures, which can then be 'passed' to the brain system that supports this type of reasoning (see <u>Inductive and deductive reasoning elicit strong responses outside the language network</u>).

**Patients with profound aphasia can nevertheless reason logically.**

Two individuals with profound aphasia, S.A. and G.S., showed preserved logical reasoning abilities. Both individuals have sustained extensive damage to left peri-Sylvian cortex (**Figure 2C**) and, consequently, exhibited severe linguistic impairments (**Figure 2D**, **SI Table 3**). In particular, critical to the hypothesis that linguistic syntactic structures mediate logical reasoning, these patients showed severe grammatical impairments in both comprehension and production, with at or near-chance performance on multiple syntactic assessments. For example, both exhibited marked impairment in assigning thematic roles in understanding reversible sentences, such as *The diver splashed the dolphin*. Nevertheless, despite their profound aphasia, both patients showed typical-like performance on the inductive and deductive reasoning tasks. In the inductive reasoning task (**Figure 1**, left), they solved the vast majority of the rules presented to them (S.A.: 19/25; G.S.: 39/40). This performance level is comparable to a normative sample of control participants; neither patient significantly differs from the controls (Crawford-Howell single-case test *p*s>0.499). Similarly, in the matrix reasoning task, both patients solved the vast majority of the problems (25/30 and 26/30), which puts them +2.3 and +1.8 standard deviations above the mean based on the age-matched normative data available for this task (WASI-II; (104)) (these raw scores correspond to *T*-scores of 73 and 68; normative $T = 50$, SD = 10).

**Inductive and deductive reasoning elicit strong responses outside the language network.**

Given that logical reasoning demands in our paradigms did not engage the language network, we wanted to ensure that they elicit a response somewhere in the brain. We first examined neural responses in one plausible candidate system—the Multiple Demand (MD) network (46, 127, 128). This network comprises a set of bilateral frontal and parietal brain areas that show strong activity during diverse cognitively demanding tasks, such as standard executive function tasks (127, 129–132). Damage to the MD network is associated with decreases in fluid intelligence (133–135). In addition, and most relevantly here, certain forms of reasoning recruit the MD network, including mathematical reasoning (35, 36, 136) and the kind of reasoning necessary to understand computer code (37, 38, 137). Replicating much prior work (e.g., (136, 138)), the areas of the MD network showed a robust response to a spatial working memory task, including a stronger response during the more demanding condition estimated in independent data ($p<0.001$ at the network level; **SI Figure 2** and **SI Table 2**). The inductive reasoning task elicited a robust response across the MD network, with a stronger response during the rule induction condition compared to the control, rule application condition ($p<0.001$ at the network level; **SI Figure 2**, **SI Table 2**). However, the deductive reasoning task did not engage the MD network ($p>.48$ at the network level). This finding aligns with the results in Coetzee & Monti (2018), who reported a dissociation between brain areas sensitive to deductive reasoning versus to general task difficulty, the latter being a key signature of the MD network (see also (52, 139, 140)). A whole-brain search revealed several frontal and parietal areas that showed a stronger response to the Modus Tollens condition compared to the Modus Ponens condition estimated in independent data ($p<0.001$ across the set of areas; **SI Figure 2**; see (141) for a detailed investigation of these brain areas).

## Discussion

To shed light on the long-standing debate on the role of linguistic syntactic representations in formal logical reasoning, we examined the responses of the language brain areas—robustly sensitive to linguistic syntactic structure (24, 27, 31, 32)—to inductive and deductive reasoning. In a complementary approach, we tested logical reasoning abilities in individuals with severe damage to the language areas. The results converged on a clear answer: linguistic representations are not engaged nor necessary for logical reasoning, at least in the adult brain. Below, we contextualize these findings with respect to the broader literature.

***Accumulating evidence for the separability of language from other cognitive functions.*** Our results showing that the left-lateralized language network is not engaged in logical induction nor deduction add to a growing body of evidence that the language network supports specifically linguistic computations. Language is a relatively recent cultural invention (142). As a result, it is reasonable to hypothesize that when language emerged, it co-opted pre-existing brain machinery. However, although language undeniably shares similarities with diverse non-linguistic functions, hypotheses about neural overlap have not found empirical support. For example, despite the fact that language serves a social-communicative function (143–145, 22) and is similar, in some ways, to non-verbal socially-relevant signals, such as facial expressions, non-verbal vocalizations, and gestures, evidence suggests that such non-verbal signals are processed in brain areas distinct from the language network (e.g., (146–148)). Despite the fact that language shares hierarchical structure with music and other domains (149, 150), music does not engage the language areas (34, 151, 152). And despite the fact that language processing requires executive resources, such as working memory and cognitive control (153–157), executive tasks do not engage the language network (34, 43, 44), and demanding linguistic computations—for example, forming non-local syntactic dependencies—are instead carried out locally within the language network ((29, 31, 158–160) see (161) for a review).

Of greatest relevance to the current investigation, the syntactic structure of language bears similarity to structure that characterizes abstract reasoning, including formal—mathematical and logical—thinking (8, 12, 162), as well as domain-specific reasoning, such as intuitive physical reasoning (163–165) and social reasoning (166). However, as with other non-linguistic functions, the studies that had examined the relationship between language processing and these types of reasoning found no overlap: mathematical reasoning (34, 36, 167) (for evidence from aphasia, see (17)), computer code comprehension (37, 38), physical reasoning (168), and social reasoning ((39–41, 169) for evidence from aphasia, see (18, 20, 170)). Our study adds to this body of work, showing that formal logical reasoning—including both induction and deduction—does not recruit nor require linguistic representations.

***Challenges to the view that language does not support thinking?*** Some have challenged the separability of language and thought based on correlations between the degree of linguistic impairment and impairment on certain cognitive tasks (e.g., (47–50)). However, these effects can be explained by the proximity between the language areas and other cognitive areas (120, 171, 172)—and, in some cases, co-lateralization of those functions to the left hemisphere (e.g., language and arithmetic: (173))—combined with the fact that stroke damage does not respect functional

boundaries. For this reason, dissociations are generally considered more informative than associations (174).

Another potential challenge to the language-thought separability comes from dual-task studies that show that a verbal task can interfere with one's ability to perform a non-verbal cognitive task (e.g., (175–177)). However, the empirical landscape is complex (for a review, see (51)), and many studies do not meet the evidential standard necessary to argue for the role of verbal representations in non-linguistic cognition. In particular, to argue that a verbal task that interferes with some critical task shares resources with that task, it is important to include a control interference task, matched for difficulty to the verbal task (evaluated in a single-task design), and to obtain an interaction such that the verbal task affects performance on the critical task to a reliably greater extent; the non-verbal task should further be shown to reliably affect performance on some other task (see e.g., (178) for a discussion of these methodological issues). In addition to these design and statistical considerations, many verbal interference paradigms (e.g., syllable or word repetition tasks) engage lower-level speech articulation and perception areas (e.g., (179)), not the language areas. Because most proposals about the role of language in thinking concern lexical and/or syntactic representations (which are handled by the higher-level language areas; (33)), it is difficult to know what verbal interference effects—even in studies that are properly designed—would mean for these proposals.

Thus, to our knowledge, no prior study—based on neuropsychological patient data or using verbal interference paradigms—provides unambiguous evidence that humans rely on lexical or syntactic representations to think, but we remain open to seeing such studies in the future, including studies that may show transient recruitment of linguistic representations, which may not be easily detectable with fMRI but may show up in studies that use intracranial recordings.

***Neural dissociation does not entail encapsulation.*** The evidence that language and thinking are dissociable in the human mind and brain is perfectly compatible with the fact that the language network has to constantly interact with our knowledge and reasoning systems (see (33, 85, 180) for discussion). First, language is only useful insofar as it allows us to share ideas with one another: both language production and comprehension therefore necessarily require our knowledge and reasoning systems in addition to our core linguistic mechanisms. Moreover, language is a powerful information compression tool, allowing for efficient representation of concepts that are useful to a given culture (181–187). In this way, language may play a facilitatory role during some cognitive tasks, but not because we use words and linguistic syntax to represent and manipulate ideas, but because using words or non-linguistic symbols (e.g., Arabic digits) or schematics (e.g., Venn diagrams) allows offloading partial or intermediate products of internal computation. Whether language is special relative to other compression tools is not known and would require careful comparisons. Finally, language can certainly facilitate learning of certain ideas, including abstract logical or relational structures (e.g., (188–192, 62, 193)). However, this role of language in concept acquisition again does not imply that we use linguistic representations to think any more than learning concepts from visual inputs implies that we use our visual system to think.

***Why do language processing and logical reasoning dissociate?*** The language-logic dissociation presumably stems from the distinct demands associated with linguistic communication vs. formal reasoning. One key difference may have to do with the kinds of meanings these systems express:

natural languages tend to express meanings related to the external and internal world, but mathematics, logic, and computer code express mostly abstract, relational meanings that do not bear a direct or necessary connection to the world (see (194) for discussion). This difference may provide a strong signal during the initial processing of the input that a particular problem should be routed to a particular system.

Another reason may be that the linguistic format is not well-suited for formal reasoning: in spite of superficial similarities in the structured nature of linguistic and logical expressions, linguistic representations are noisy and ambiguous in ways that make them unreliable for supporting formal inference. Natural language is riddled with referential vagueness, scope ambiguities, and underspecification, requiring pragmatic enrichment, all of which can obscure the logical structure of a sentence. Indeed, evidence from AI research shows that LLM performance on reasoning tasks improves when linguistic inputs are converted into first-order logic, which can then be fed into an external theorem solver (e.g., (195–200)).

***Alternatives to linguistic representations.*** If we don't think in language, what kind of representations support logical reasoning? Competing accounts postulate distinct formats of thought-mediating representations, which could translate into distinct predictions about neural implementation. For example, according to the *mental model* framework, reasoners construct analog simulations of the problem, exploiting visuo-spatial representations and working memory resources to model the possible consequences of premises (201, 202). These accounts therefore predict the engagement of brain areas that support visual/spatial imagery and working memory (203, 204). Alternatively, *script- or schema-based* approaches propose that reasoning draws on cached relational templates (structured representations of events or situations abstracted from prior experience) (205–208). These accounts suggest that reasoning is scaffolded by context- and domain-specific knowledge structures, which may draw on multiple semantic networks (e.g., (209, 210)). And the *mental logic* proposals construe abstract reasoning as the manipulation of amodal, abstract symbols according to syntactic inference rules (211–213). Perhaps the most promising candidate from this class of proposals is the probabilistic language of thought (PLOT) hypothesis (214–220). According to this hypothesis, mental algorithms are construed as symbolic programs over concepts, which encode probabilistic knowledge, including both domain-specific knowledge and abstract relational knowledge. In this way, PLOT provides a flexible medium for any type of reasoning, and captures the graded nature of mental representations. Future brain imaging studies may help determine whether logical reasoning depends on visuo-spatial, fully abstract, or domain-specific (at least during early stages of development) representations.

***Modularity of reasoning.*** Aside from the language-logic dissociation, this study highlights the fact that different kinds of reasoning dissociate from one another. Some forms of reasoning have been shown to be domain-specific, including social reasoning (221–224) and intuitive physical reasoning (225–228). Both of these systems are distinct from the Multiple Demand (MD) network, implicated in diverse goal-directed behaviors and some forms of abstract reasoning (46, 132), and from the Default network, implicated in episodic self-projection and constructing situation models (229–232), perhaps using spatial representations (233). We here show that inductive reasoning, similar to arithmetic reasoning, strongly engages the MD network, but syllogistic deductive reasoning does not, and instead recruits a distinct set of brain areas, in line with past work ((52, 96, 139, 140); see (141) for a systematic investigation of these areas). Why do these different forms

of reasoning dissociate? Is it because of the distinct content/structure of the problems? Or, perhaps, differences in the mental operations that those problems require? Given the strong integration between memory and computation in the brain (see e.g., (234) and (235) for discussions), both of these factors may drive the observed dissociations.

**Limitations and open questions.** The current study is limited in at least two ways. (#1) The aphasia component includes only two participants. This small number was dictated by the requirement of a *profound* linguistic impairment, which is rare (236). Evidence of intact logical abilities in patients with less severe aphasia would be less informative because it would always be possible that they are relying on the remaining portions of the language network to reason. Evidence of intact cognition in severely linguistically impaired individuals is precious and important, even if it comes from a handful of participants. After all, the field of cognitive neuroscience has gained some of its greatest insights from case studies (237–240). (#2) These findings should be extended to a broader array of logical reasoning tasks, including contextualized paradigms (241–244), and non-verbal versions of purely deductive tasks.

Many questions remain open. For example: Is the language-logic dissociation already present during early childhood, or does it emerge over development? Is linguistic exposure necessary to learn to reason logically? And is the dissociation between language and logic, or among different types of reasoning, an inevitability given the computational demands of these different cognitive tasks, or merely an accident of biological evolution—a question that can now be tested in neural network models (245–249).

**Conclusions.** Our results provide evidence against the hypothesis that natural language serves as the medium of abstract logical reasoning and suggest that such reasoning is underpinned by a distinct representational format.

## MATERIALS AND METHODS

***Study 1 (fMRI) Participants.*** Twenty-nine adults (15 females; mean age=27.3 years, SD=12.9 years) from MIT and the surrounding Boston, MA community participated for payment. All were native English speakers, had normal hearing and vision, no history of language impairment, and canonical left-lateralized language networks (as determined by the language localizer). All participants provided written informed consent in line with the requirements of MIT's Committee on the Use of Humans as Experimental Subjects (Protocol #2010000243). All participants completed the language localizer task and the Multiple Demand (MD) localizer task; different subsets of the 29 participants completed the logic tasks, with at least 17 participants per task (17 completed the induction and the matrix tasks; 23 completed the syllogistic task).

***Study 2 (behavior) Participants.*** Two profoundly aphasic male participants participated (S.A.: 78 years; G.S.: 50 years). Both had large lesions that had significantly damaged the left inferior frontal and left temporal brain areas, and were classified as severely agrammatic (**SI Table 3**). Both were native English speakers, did not present with any visual impairments, and were pre-morbidly right-handed. Forty age-matched neurotypical participants were included as controls (see Supp. Methods for details). Ethics approval was granted by the UCL Language and Cognition Departmental Ethics Committee (LC/2023/05). All participants provided informed consent prior to taking part in the study.

***Induction Task (fMRI, behavior).*** Participants were shown an input list of 1-5 single-digit numbers and an output list of 0-5 single-digit numbers and asked to guess the rule that transformed the input list into the output list. They were then shown another input list and asked to type the output list. They were told whether or not their guess about the underlying rule is correct and are shown another input list. In the **fMRI version**, each trial (corresponding to a rule) consisted of 8 problems. After the last problem, participants were told—via a schematic (**Figure 1**, left)—the correct rule and asked to apply this rule to two more input lists. Because participants often guessed the correct rule part-way through the initial list of 8 problems (as indicated by generating correct output lists for the last several problems), we defined the *induction* period as the subset of the eight problems before the problem after which the participant made no more errors, and the *application* period as the subsequent problems along with the two problems after the correct rule was revealed. The induction > application contrast targets neural processes involved in hypothesis generation and rule discovery beyond those required for rule implementation and response entry. Each participant completed 40 rules. For this and other tasks, all the materials are available at OSF (https://osf.io/jm9rd); additional procedural and timing details are included in Supplementary Methods.

For the **behavioral version** of the task, the problems were presented on paper, and participants provided written responses. The same 40 rules were used, except that a few problems were added in order to have 14 problems (input-output pairs) per rule, and a few problems were replaced based on feedback from the fMRI participants who found some input-output pairs confusing. The stimuli were divided into four sets, with ten rules per set. GS completed all 40 rules across several sessions (~6 rules per session); SA only completed sets 1-3 also across sessions (~4 rules per session); the variability in the number of rules per session was due to age-related health conditions and because the experiment had to be completed amongst other activities at routine check-in sessions. Control

participants each completed one set of 10 rules within a single testing session, with a brief break half-way through, and 10 participants completed each set. For both the patients and the controls, testing for a given rule was stopped after the participant answered four problems in a row correctly. Time permitting, the participants with aphasia were sometimes offered to schematically illustrate the rule they guessed (see **Figure 2** for an example).

***Deduction Task 1 – Syllogistic reasoning (fMRI).*** Participants were presented with classic syllogisms and asked to judge the validity of the third sentence (the conclusion) given the first two sentences (the premises). Each trial corresponded to a syllogism, and trials varied in difficulty between harder deduction (Modus Tollens) and easier deduction (Modus Ponens), and in whether the problems used real words or nonwords. The Modus Tollens > Modus Ponens contrast targets cognitive processes related to the complexity of logical deductive reasoning. Each participant completed between 16 and 32 trials of each condition.

***Deduction Task 2 – Matrix reasoning (fMRI, behavior).*** Participants were presented with sets of images of geometrical shapes and asked to find an outlier (fMRI version) or a missing image (behavioral version). In **fMRI**, participants were presented with four images and asked to decide which image does not fit with the others. The experiment used a blocked design; trials in the hard blocks used stimuli from Cattell (102), and trials in the easy blocks used simpler problems created by Woolgar et al. (136), who adapted this task for fMRI (e.g., three images of the same simple shape and an image of a different shape). The hard > easy contrast targets cognitive processes related to relational reasoning, hypothesis generation, and deductive inference.

For the **behavioral** version, we used the Matrix Reasoning subtest of the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II). Participants were presented with an array of images with one image missing and asked to choose the missing image from five options to complete the pattern. The test was administered on paper, following the standard WASI-II administration procedures. Testing was stopped after the participant answered three problems incorrectly.

***Critical Analyses - fMRI.*** (For the details of fMRI data acquisition, preprocessing, and modeling, see Supplementary Methods.) To test whether the language areas respond during logical reasoning tasks, we used an extensively validated localizer (117) to identify regions of interest functionally in individual participants. Participants were asked to attentively read sentences and sequences of nonwords (for details, see Supp. Methods). The responses in the language fROIs (and control fROIs, as explained below) were statistically evaluated using linear mixed-effects models. For each contrast, we fit separate models at both the network level and for individual fROIs.

The **network-level model**: $BOLD \sim Condition + (1 \mid Participant) + (1 \mid fROI)$.
The **individual fROI-level model**: $BOLD \sim Condition + (1 \mid Participant)$.

To ensure that our logical reasoning tasks elicit a strong response somewhere in the brain, we performed two analyses. *First*, we used a spatial working memory task (keeping track of 8 vs. 4 squares in a 3x4 grid) as a localizer for the Multiple Demand network (e.g., (138)) (for details, see Supp. Methods). And *second*, because the syllogistic reasoning task did not elicit a response in the MD network, we additionally performed a whole-brain search for areas that respond to the Modus Tollens > Modus Ponens contrast (see Supp. Methods). (For information on the behavioral performance of participants in the critical fMRI tasks, see Supp. Methods.)

***Critical Analyses – Behavior.*** For the induction task, we used the Crawford and Howell (250) test to compare each patient's performance to the controls. The Matrix Reasoning WASI-II data were scored according to the standard guidelines provided in the official manual.

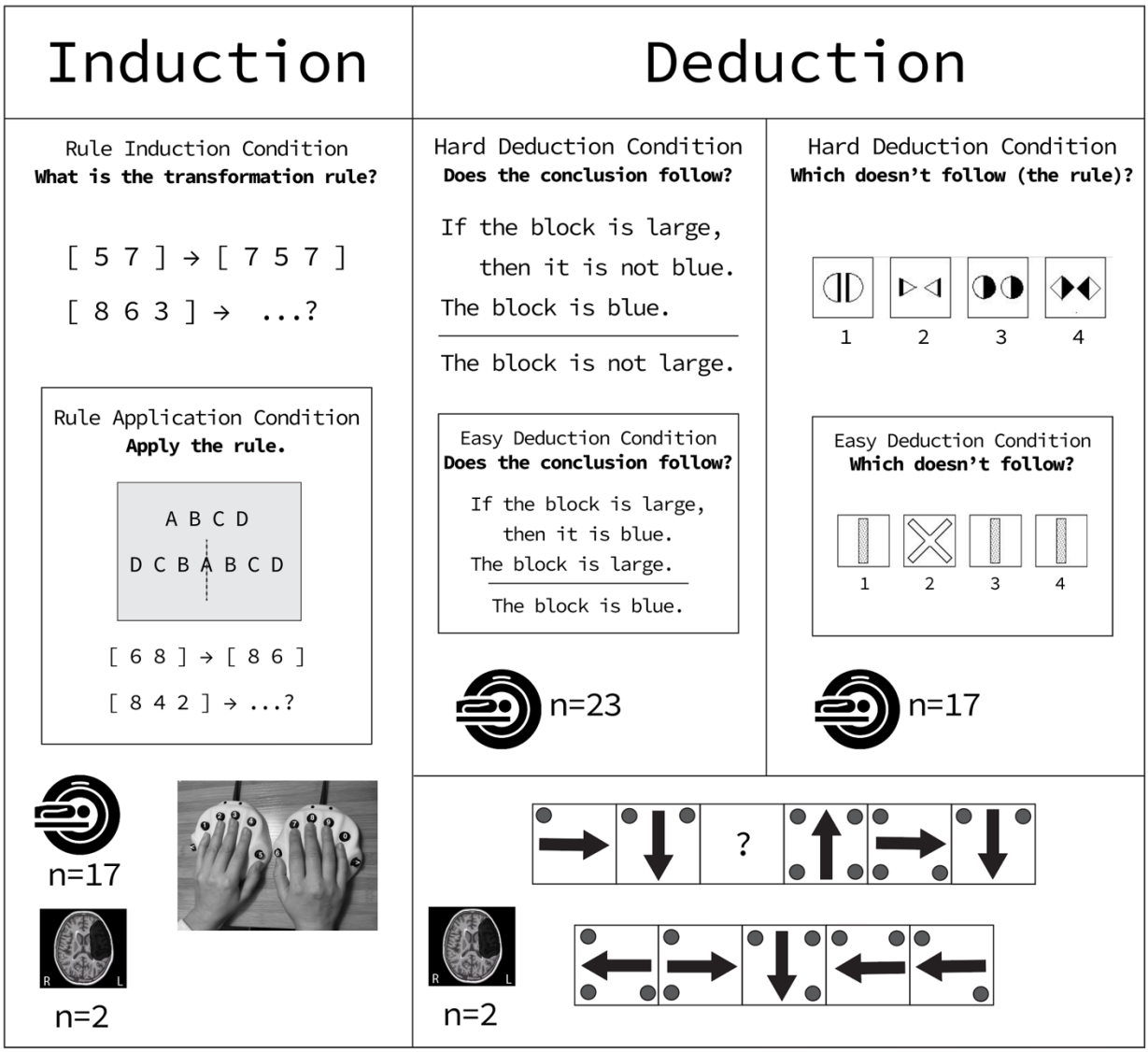


**Figure 1. Inductive and deductive logical reasoning paradigms**. For ***inductive reasoning*** (left), participants are shown an input list of 1-5 single-digit numbers and an output list of 0-5 single-digit numbers and asked to guess the rule that transformed the input list into the output list (see Methods for details). In the fMRI version, after the eighth problem, participants were told—via a schematic— the correct rule and were asked to apply this rule to two more input lists. For ***deductive reasoning*** (right), in the syllogistic reasoning task, participants are presented with classic three-sentence syllogisms and asked to judge the validity of the third sentence (the conclusion) given the first two sentences (the premises). Trials vary in difficulty between harder deduction (Modus Tollens) and easier deduction (Modus Ponens). In the matrix reasoning task, participants are presented with images of geometrical shapes and asked to figure out the rule. In the fMRI version, they are asked to decide which image does not fit with the others, and in the behavioral version, they have to identify the missing image (see Methods for details).

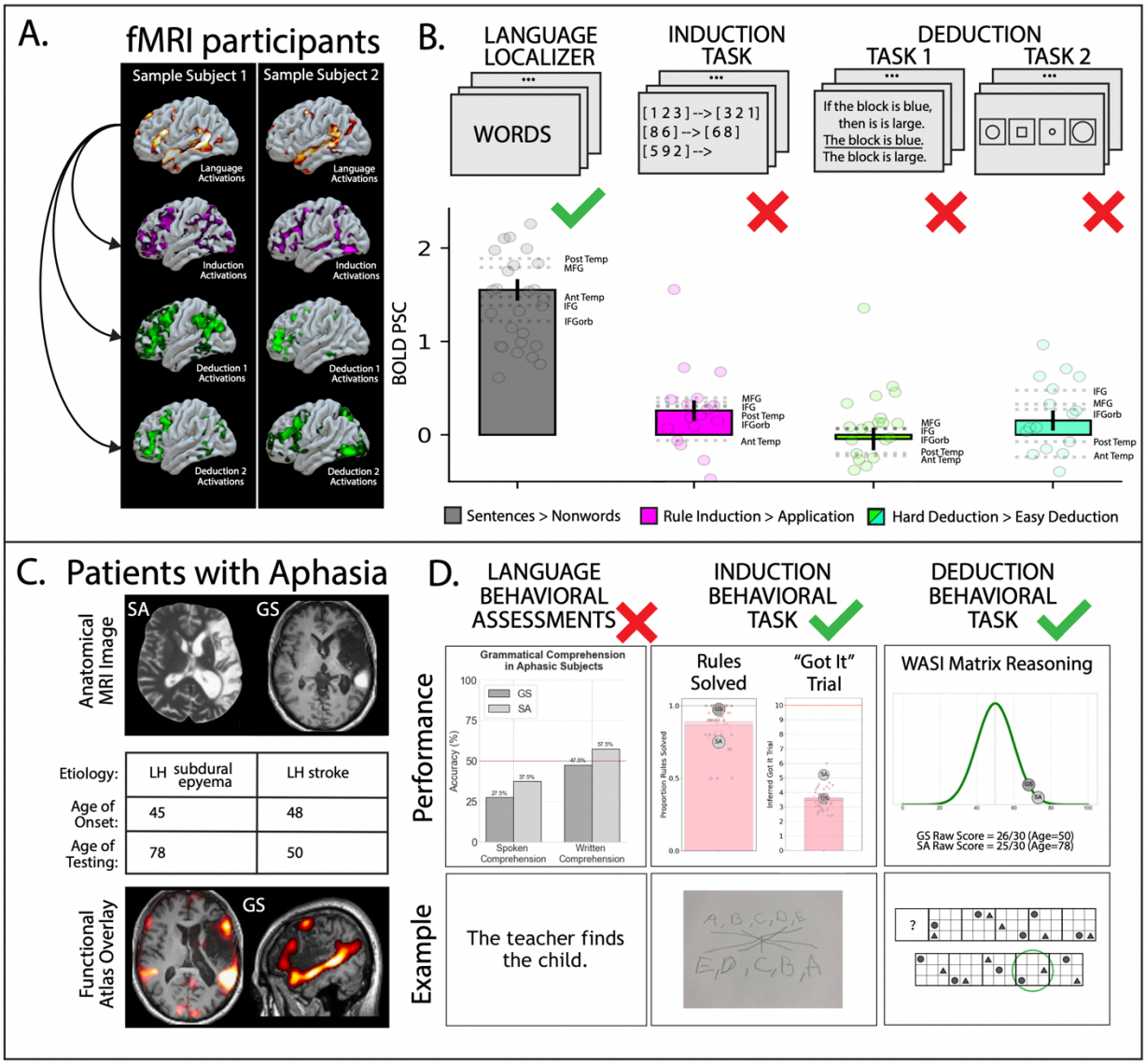


**Figure 2. The language network does not support logical reasoning. A.** Activation maps for the language contrast (1st row, yellow) and three logical reasoning contrasts (2nd-4th rows, purple and green) in two individual sample participants. **B.** Responses in the language network (individually defined in each participant; see Methods), in percent BOLD signal change, to the language contrast (sentences > nonwords; estimated in left-out data; grey), the induction contrast (induction > application; magenta), and the two deduction contrasts (hard deduction > easy deduction; green shades). The bars show the across-network averages and the error bars correspond to the standard error of the mean by participant; the light dots correspond to individual participants' responses; the means for each of the 5 language fROIs (IFGorb, IFG, MFG, AntTemp, and PostTemp) are shown with horizontal dashed lines. **C.** The anatomical scans of the two patients with severe aphasia (S.A. and G.S.), brief info on the patients, and a projection of the probabilistic atlas for the language network (from (251)) into the MRI image of G.S. **D.** Behavioral performance of the patients with aphasia and—for the critical tasks—control participants. The first column shows performance on the language evaluation tasks (spoken and written comprehension; (252)) for S.A. (light grey bars) and G.S. (dark grey bars). Both participants show at or below chance performance. In the bottom row, we show sample items. (See Methods for additional details of the linguistic evaluation of the patients.) The second and third columns show performance on the induction and

deduction tasks (S.A.: light grey dots; G.S.: dark grey dots). For the induction task, the control data are shown with a pink bar. In the bottom row, we show a sample rule representation generated by G.S., where he illustrates the “reverse the numbers” rule. For the deduction task, the control data are shown with the normal age-matched distribution, and in the bottom row, we show a sample item.

| Predictors | Estimate | p-value |
|---|---|---|
| *Language Task* | | |
| Intercept (nonwords) | 0.7407 | 0.0666 |
| Critical Effect (sentences) | **1.5518** | **<0.0001** |
| *Induction Task* | | |
| Intercept (application/post-solution) | 0.1408 | 0.7414 |
| Critical Effect (induction/pre-solution) | 0.2601 | 0.0224 |
| *Deduction Task 1 (syllogistic reasoning)* | | |
| Intercept (easy/Modus Ponens) | 1.9361 | 0.0065 |
| Critical Effect (hard/Modus Tollens) | -0.0444 | 0.7102 |
| *Deduction Task 2 (matrix reasoning)* | | |
| Intercept (easy deduction) | -0.2234 | 0.3527 |
| Critical Effect (hard deduction) | 0.1533 | 0.1185 |

**Table 1. The responses of the language network to the language and logical reasoning contrasts.** The results from the linear mixed-effects models (see Methods) at the network level (see **SI Table 1** for the results broken down by fROI).

**References**


1. Aristotle, *Categories and De Interpretatione* (Harvard University Press, 1938).

2. L. Wittgenstein, *Tractatus Logico Philosophicus* (1921).

3. D. Davidson, Truth and Meaning. *Synthese* **17**, 304–323 (1967).

4. D. Davidson, "Thought and Talk" in *Mind and Language*, S. D. Guttenplan, Ed. (Clarendon Press, 1975), pp. 1975–7.

5. D. C. Dennett, *Consciousness explained* (Little, Brown and Co, 1991).

6. D. C. Dennett, *Kinds of minds: Toward an understanding of consciousness* (Basic Books, 1996).

7. D. C. Dennett, *From bacteria to Bach and back: the evolution of minds*, First edition (W.W. Norton & Company, 2017).

8. N. Chomsky, *Syntactic structures* (Mouton, 1957).

9. N. Chomsky, *Aspects of the theory of syntax* (M.I.T. Press, 1965).

10. N. Chomsky, "A Minimalist Program for Linguistic Theory" in *The View From Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, K. L. Hale, S. J. Keyser, Eds. (MIT Press, 1993).

11. N. Chomsky, *The Minimalist program* (The MIT Press, 1995).

12. P. Carruthers, The cognitive functions of language. *Behav Brain Sci* **25**, 657–674 (2002).

13. W. Munroe, Why are you talking to yourself? The epistemic role of inner speech in reasoning. *Noûs* **56**, 841–866 (2022).

14. S. Pinker, *The language instinct* (William Morrow & Co, 1994).

15. J. A. Fodor, *Concepts: Where Cognitive Science Went Wrong*, 1st Ed. (Oxford University PressOxford, 1998).

16. J. A. Fodor, *The Language of Thought* (Harvard University Press, 1975).

17. R. A. Varley, N. J. C. Klessinger, C. A. J. Romanowski, M. Siegal, Agrammatic but numerate. *Proceedings of the National Academy of Sciences* **102**, 3519–3524 (2005).

18. R. Varley, M. Siegal, Evidence for cognition without grammar from causal reasoning and 'theory of mind' in an agrammatic aphasic patient. *Current Biology* **10**, 723–726 (2000).

19. I. A. Apperly, D. Samson, C. Chiavarino, G. W. Humphreys, Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. *J Cogn Neurosci* **16**, 1773–1784 (2004).

20. I. A. Apperly, D. Samson, N. Carroll, S. Hussain, G. Humphreys, Intact first- and second-order false belief reasoning in a patient with severely impaired grammar. *Social Neuroscience* **1**, 334–348 (2006).

21. E. Fedorenko, R. Varley, Language and thought are not the same thing: evidence from neuroimaging and neurological patients: Language versus thought. *Ann. N.Y. Acad. Sci.* **1369**, 132–153 (2016).

22. E. Fedorenko, S. T. Piantadosi, E. A. F. Gibson, Language is primarily a tool for communication rather than thought. *Nature* **630**, 575–586 (2024).

23. L. Menenti, S. M. E. Gierhan, K. Segaert, P. Hagoort, Shared language: overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychol Sci* **22**, 1173–1182 (2011).

24. L. Giglio, M. Ostarek, K. Weber, P. Hagoort, Commonalities and Asymmetries in the Neurobiological Infrastructure for Language Production and Comprehension. *Cereb Cortex* **32**, 1405–1418 (2022).

25. J. Hu, *et al.*, Precision fMRI reveals that the language-selective network supports both phrase-structure building and lexical access during language production. *Cerebral Cortex* **33**, 4384–4404 (2023).

26. M. Ben-Shachar, T. Hendler, I. Kahn, D. Ben-Bashat, Y. Grodzinsky, The Neural Reality of Syntactic Transformations: Evidence From Functional Magnetic Resonance Imaging. *Psychol Sci* **14**, 433–440 (2003).

27. C. Pallier, A.-D. Devauchelle, S. Dehaene, Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences* **108**, 2522–2527 (2011).

28. L. K. Tyler, *et al.*, Left inferior frontal cortex and syntax: function, structure and behaviour in patients with left hemisphere damage. *Brain* **134**, 415–431 (2011).

29. I. Blank, Z. Balewski, K. Mahowald, E. Fedorenko, Syntactic processing is distributed across the language system. *NeuroImage* **127**, 307–323 (2016).

30. N. Ding, L. Melloni, H. Zhang, X. Tian, D. Poeppel, Cortical tracking of hierarchical linguistic structures in connected speech. *Nat Neurosci* **19**, 158–164 (2016).

31. C. Shain, I. A. Blank, E. Fedorenko, E. Gibson, W. Schuler, Robust Effects of Working Memory Demand during Naturalistic Language Comprehension in Language-Selective Cortex. *J. Neurosci.* **42**, 7412–7430 (2022).

32. C. Shain, *et al.*, Distributed Sensitivity to Syntax and Semantics throughout the Language Network. *Journal of Cognitive Neuroscience* **36**, 1427–1471 (2024).

33. E. Fedorenko, A. A. Ivanova, T. I. Regev, The language network as a natural kind within the broader landscape of the human brain. *Nat. Rev. Neurosci.* **25**, 289–312 (2024).

34. E. Fedorenko, M. K. Behr, N. Kanwisher, Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences* **108**, 16428–16433 (2011).

35. M. M. Monti, L. M. Parsons, D. N. Osherson, Thought Beyond Language: Neural Dissociation of Algebra and Natural Language. *Psychol Sci* **23**, 914–922 (2012).

36. M. Amalric, S. Dehaene, A distinct cortical network for mathematical knowledge in the human brain. *Neuroimage* **189**, 19–31 (2019).

37. A. A. Ivanova, *et al.*, Comprehension of computer code relies primarily on domain-general executive brain regions. *Elife* **9**, e58906 (2020).

38. Y.-F. Liu, J. Kim, C. Wilson, M. Bedny, Computer code comprehension shares neural resources with formal logical inference in the fronto-parietal network. *eLife* **9**, e59340 (2020).

39. A. M. Paunov, I. A. Blank, E. Fedorenko, Functionally distinct language and Theory of Mind networks are synchronized at rest and during language comprehension. *J Neurophysiol* **121**, 1244–1265 (2019).

40. A. M. Paunov, *et al.*, Differential Tracking of Linguistic vs. Mental State Content in Naturalistic Stimuli by Language and Theory of Mind (ToM) Brain Networks. *Neurobiol Lang (Camb)* **3**, 413–440 (2022).

41. C. Shain, A. Paunov, X. Chen, B. Lipkin, E. Fedorenko, No evidence of theory of mind reasoning in the human language network. *Cerebral Cortex* **33**, 6299–6319 (2023).

42. L. M. DiNicola, R. M. Braga, R. L. Buckner, Parallel distributed networks dissociate episodic and social functions within the individual. *J Neurophysiol* **123**, 1144–1179 (2020).

43. S. Malik-Moraleda, *et al.*, An investigation across 45 languages and 12 language families reveals a universal language network. *Nat Neurosci* **25**, 1014–1019 (2022).

44. K. J. Hiersche, E. Schettini, J. Li, Z. M. Saygin, Functional dissociation of the language network and other cognition in early childhood. *Hum Brain Mapp* **45**, e26757 (2024).

45. R. W. Engle, S. W. Tuholski, J. E. Laughlin, A. R. A. Conway, Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General* **128**, 309–331 (1999).

46. J. Duncan, The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences* **14**, 172–179 (2010).

47. J. V. Baldo, *et al.*, Is problem solving dependent on language? *Brain Lang* **92**, 240–250 (2005).

48. J. V. Baldo, S. A. Bunge, S. M. Wilson, N. F. Dronkers, Is relational reasoning dependent on language? A voxel-based lesion symptom mapping study. *Brain Lang* **113**, 59–64 (2010).

49. J. V. Baldo, S. R. Paulraj, B. C. Curran, N. F. Dronkers, Impaired reasoning and problem-solving in individuals with language impairment due to aphasia or language delay. *Front Psychol* **6**, 1523 (2015).

50. E. Duricy, C. Durisko, J. A. Fiez, Lesion-symptom mapping reveals differential relationships between language and precise versus approximate numeracy. *Cortex* **193**, 117–133 (2025).

51. J. S. K. Nedergaard, M. Wallentin, G. Lupyan, Verbal interference paradigms: A systematic review investigating the role of language in cognition. *Psychon Bull Rev* **30**, 464–488 (2023).

52. J. K. Kroger, L. E. Nystrom, J. D. Cohen, P. N. Johnson-Laird, Distinct neural substrates for deductive and mathematical processing. *Brain Research* **1243**, 86–103 (2008).

53. M. M. Monti, L. M. Parsons, D. N. Osherson, The boundaries of language and thought in deductive inference. *Proceedings of the National Academy of Sciences* **106**, 12554–12559 (2009).

54. N. Geschwind, The organization of language and the brain. *Science* **170**, 940–944 (1970).

55. H. Goodglass, *Understanding aphasia* (Academic Press, 1993).

56. A. D. Friederici, Towards a neural basis of auditory sentence processing. *Trends Cogn Sci* **6**, 78–84 (2002).

57. P. Hagoort, The neurobiology of language beyond single-word processing. *Science* **366**, 55–58 (2019).

58. J. P. Coetzee, *et al.*, Dissociating Language and Thought in Human Reasoning. *Brain Sciences* **13**, 67 (2023).

59. E. Diachek, I. Blank, M. Siegelman, J. Affourtit, E. Fedorenko, The Domain-General Multiple Demand (MD) Network Does Not Support Core Aspects of Language Comprehension: A Large-Scale fMRI Investigation. *J. Neurosci.* **40**, 4536–4550 (2020).

60. A. Gopnik, Words and plans: early language and the development of intelligent action. *J Child Lang* **9**, 303–318 (1982).

61. M. K. Goddu, A. Gopnik, The Development of Human Causal Learning and Reasoning. *Nature Reviews Psychology* **3**, 319–339 (2024).

62. S. Carey, *The origin of concepts* (Oxford University Press, 2009).

63. S. A. Gelman, *The essential child: Origins of essentialism in everyday thought* (Oxford University Press, 2003).

64. A. Gopnik, *et al.*, A Theory of Causal Learning in Children: Causal Maps and Bayes Nets. *Psychological Review* **111**, 3–32 (2004).

65. A. Gopnik, H. M. Wellman, Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin* **138**, 1085–1108 (2012).

66. N. Cesana-Arlotti, *et al.*, Precursors of logical reasoning in preverbal human infants. *Science* **359**, 1263–1266 (2018).

67. S. Mody, S. Carey, The emergence of reasoning by the disjunctive syllogism in early childhood. *Cognition* **154**, 40–48 (2016).

68. S. Gautam, T. Suddendorf, J. Redshaw, When can young children reason about an exclusive disjunction? A follow up to Mody and Carey (2016). *Cognition* **207**, 104507 (2021).

69. J. R. Best, P. H. Miller, A developmental perspective on executive function. *Child Dev* **81**, 1641–1660 (2010).

70. A. Diamond, Executive functions. *Annu Rev Psychol* **64**, 135–168 (2013).

71. P. D. Zelazo, S. M. Carlson, The neurodevelopment of executive function skills: Implications for academic achievement gaps. *Psychology & Neuroscience* **13**, 273–298 (2020).

72. J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. [Preprint] (2019). Available at: http://arxiv.org/abs/1810.04805 [Accessed 24 July 2025].

73. A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, Improving Language Understanding by Generative Pre-Training. (2018).

74. T. B. Brown, *et al.*, Language Models are Few-Shot Learners. [Preprint] (2020). Available at: http://arxiv.org/abs/2005.14165 [Accessed 24 July 2025].

75. D. Hendrycks, *et al.*, Measuring Massive Multitask Language Understanding. [Preprint] (2021). Available at: http://arxiv.org/abs/2009.03300 [Accessed 24 July 2025].

76. OpenAI, *et al.*, GPT-4 Technical Report. [Preprint] (2024). Available at: http://arxiv.org/abs/2303.08774 [Accessed 24 July 2025].

77. D. Rein, *et al.*, GPQA: A Graduate-Level Google-Proof Q&A Benchmark. [Preprint] (2023). Available at: http://arxiv.org/abs/2311.12022 [Accessed 24 July 2025].

78. F. Chollet, M. Knoop, G. Kamradt, B. Landers, ARC Prize 2024: Technical Report. (2024).

79. A. Srivastava, *et al.*, Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models. [Preprint] (2023). Available at: http://arxiv.org/abs/2206.04615 [Accessed 24 July 2025].

80. J. Liu, *et al.*, LogiQA: A Challenge Dataset for Machine Reading Comprehension with Logical Reasoning. [Preprint] (2020). Available at: http://arxiv.org/abs/2007.08124 [Accessed 24 July 2025].

81. D. Hendrycks, *et al.*, Measuring Mathematical Problem Solving With the MATH Dataset. [Preprint] (2021). Available at: http://arxiv.org/abs/2103.03874 [Accessed 24 July 2025].

82. B. Y. Lin, *et al.*, WildBench: Benchmarking LLMs with Challenging Tasks from Real Users in the Wild. [Preprint] (2024). Available at: http://arxiv.org/abs/2406.04770 [Accessed 24 July 2025].

83. E. Boix-Adsera, S. Bengio, O. Saremi, E. Littwin, WHEN CAN TRANSFORMERS REASON WITH ABSTRACT SYMBOLS? (2024).

84. S. Ryb, M. Giulianelli, A. Sinclair, R. Fernández, AnaLog: Testing Analytical and Deductive Logic Learnability in Language Models in *Proceedings of the 11th Joint Conference on Lexical and Computational Semantics*, V. Nastase, E. Pavlick, M. T. Pilehvar, J. Camacho-Collados, A. Raganato, Eds. (Association for Computational Linguistics, 2022), pp. 55–68.

85. K. Mahowald, *et al.*, Dissociating language and thought in large language models. *Trends Cogn Sci* **28**, 517–540 (2024).

86. L. Qiu, *et al.*, Phenomenal Yet Puzzling: Testing Inductive Reasoning Capabilities of Language Models with Hypothesis Refinement. [Preprint] (2023). Available at: http://arxiv.org/abs/2310.08559 [Accessed 19 October 2023].

87. P. Shojaee*†, *et al.*, The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity. (2025).

88. R. T. McCoy, S. Yao, D. Friedman, M. Hardy, T. L. Griffiths, Embers of Autoregression: Understanding Large Language Models Through the Problem They are Trained to Solve. [Preprint] (2023). Available at: http://arxiv.org/abs/2309.13638 [Accessed 24 July 2025].

89. M. Nezhurina, L. Cipolina-Kun, M. Cherti, J. Jitsev, Alice in Wonderland: Simple Tasks Showing Complete Reasoning Breakdown in State-Of-the-Art Large Language Models. [Preprint] (2025). Available at: http://arxiv.org/abs/2406.02061 [Accessed 24 July 2025].

90. J. S. Rule, *et al.*, Symbolic metaprogram search improves learning efficiency and explains rule learning in humans. *Nat Commun* **15**, 6847 (2024).

91. J. S. Bruner, J. J. Goodnow, G. A. Austin, *A study of thinking* (John Wiley and Sons, 1956).

92. M. M. Bongard, *Problema Uznavaniya* (Nauka Press, 1967).

93. D. Gentner, Structure-mapping: A theoretical framework for analogy. *Cognitive Science* **7**, 155–170 (1983).

94. D. R. Hofstadter, M. Mitchell, "The Copycat project: A model of mental fluidity and analogy-making" in *Analogical Connections*, Advances in connectionist and neural computation theory, Vol. 2., (Ablex Publishing, 1994), pp. 31–112.

95. J. B. (Joshua B. Tenenbaum, "A Bayesian framework for concept learning," Massachusetts Institute of Technology. (1999).

96. J. P. Coetzee, M. M. Monti, At the core of reasoning: Dissociating deductive and non-deductive load. *Hum Brain Mapp* **39**, 1850–1861 (2018).

97. P. C. Wason, "Reasoning" in *New Horizons in Psychology*, B. Foss, Ed. (Harmondsworth: Penguin Books, 1966), pp. 135–151.

98. P. N. Johnson-Laird, P. C. Wason, A theoretical analysis of insight into a reasoning task. *Cognitive Psychology* 134–148 (1970).

99. P. C. Wason, P. N. Johnson-Laird, *Thinking and reasoning: selected readings* (Penguin Books, 1972).

100. P. N. Johnson-Laird, "Models of Deduction" in *Reasoning: Representation and Process*, (Psychology Press, 1975).

101. L. J. Rips, Cognitive Processes in Propositional Reasoning. *Psychological Review* **90**, 38–71 (1983).

102. R. B. Cattell, The dimensions of culture patterns by factorization of national characters. *The Journal of Abnormal and Social Psychology* **44**, 443–469 (1949).

103. A. Woolgar, D. Bor, J. Duncan, Global Increase in Task-related Fronto-parietal Activity after Focal Frontal Lobe Lesion. *Journal of Cognitive Neuroscience* **25**, 1542–1552 (2013).

104. D. Wechsler, WASI-II - Wechsler Abbreviated Scale of Intelligence | Second Edition | Pearson Assessments US. (2011). Available at: https://www.pearsonassessments.com/en-us/Store/Professional-Assessments/Cognition-%26-Neuro/Wechsler-Abbreviated-Scale-of-Intelligence-%7C-Second-Edition/p/100000593?srsltid=Afm-BOoqQrcPvO66dGHj7cXmbX3RGDp5uGvNKBVmgrpqcP_wP_Fwyw8vJ [Accessed 24 July 2025].

105. R. B. Cattell, A culture-free intelligence test. I. *Journal of Educational Psychology* **31**, 161–179 (1940).

106. J. Raven, The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology* **41**, 1–48 (2000).

107. D. Wechsler, *Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV)* (Pearson, 2008).

108. R. Primi, M. E. Ferrão, L. S. Almeida, Fluid intelligence as a predictor of learning: A longitudinal multilevel approach applied to math. *Learning and Individual Differences* **20**, 446–451 (2010).

109. D. P. Flanagan, P. L. Harrison, Eds., *Contemporary intellectual assessment: theories, tests, and issues*, 3rd ed (Guilford Press, 2012).

110. B. Roth, *et al.*, Intelligence and school grades: A meta-analysis. *Intelligence* **53**, 118–137 (2015).

111. L. G. Weiss, D. H. Saklofske, J. A. Holdnack, A. Prifitera, *WISC-V assessment and interpretation: scientist-practitioner perspectives* (Elsevier/AP, Academic Press is an imprint of Elsevier, 2016).

112. B. Bediou, *et al.*, Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin* **144**, 77–110 (2018).

113. B. Bediou, *et al.*, Effects of Action Video Game Play on Cognitive Skills: A Meta-Analysis. *Technology, Mind, and Behavior* **4** (2023).

114. S. Zorowitz, G. Chierchia, S.-J. Blakemore, N. D. Daw, An item response theory analysis of the matrix reasoning item bank (MaRs-IB). *Behav Res* **56**, 1104–1122 (2024).

115. P. A. Carpenter, M. A. Just, P. Shell, What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review* **97**, 404–431 (1990).

116. R. Primi, Complexity of geometric inductive reasoning tasks: Contribution to the understanding of fluid intelligence. *Intelligence* **30**, 41–70 (2001).

117. E. Fedorenko, P.-J. Hsieh, A. Nieto-Castañón, S. Whitfield-Gabrieli, N. Kanwisher, New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *J Neurophysiol* **104**, 1177–1194 (2010).

118. T. L. Scott, J. Gallée, E. Fedorenko, A new fun and robust version of an fMRI localizer for the frontotemporal language system. *Cognitive Neuroscience* **8**, 167–176 (2017).

119. C. Gao, *et al.*, Increasing alignment of large language models with language processing in the human brain. *Nat Comput Sci* **5**, 1080–1090 (2025).

120. R. M. Braga, L. M. DiNicola, H. C. Becker, R. L. Buckner, Situating the left-lateralized language network in the broader organization of multiple specialized large-scale distributed networks. *Journal of Neurophysiology* **124**, 1415–1448 (2020).

121. J. Du, *et al.*, Within-individual precision mapping of brain networks exclusively using task data. *Neuron* **113**, 4069-4083.e6 (2025).

122. C. Shain, E. Fedorenko, A language network in the individualized functional connectomes of over 1,000 human brains doing arbitrary tasks. *bioRxiv* 2025.03.29.646067 (2025). https://doi.org/10.1101/2025.03.29.646067.

123. S. M. Wilson, J. Fridriksson, Aphasia and aphasia recovery. *The cognitive neurosciences* (2020).

124. L. Carroll, *Through the Looking-Glass, and What Alice Found There* (Macmillan and Co., 1871).

125. C. Humphries, J. R. Binder, D. A. Medler, E. Liebenthal, Syntactic and Semantic Modulation of Neural Activity during Auditory Sentence Comprehension. *J Cogn Neurosci* **18**, 665–679 (2006).

126. M. Bedny, A. Pascual-Leone, D. Dodell-Feder, E. Fedorenko, R. Saxe, Language processing in the occipital cortex of congenitally blind adults. *Proceedings of the National Academy of Sciences* **108**, 4429–4434 (2011).

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

128. J. Duncan, M. Assem, S. Shashidhara, Integrated Intelligence from Distributed Brain Activity. *Trends in Cognitive Sciences* **24**, 838–852 (2020).

129. T. A. Niendam, *et al.*, Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cogn Affect Behav Neurosci* **12**, 241–268 (2012).

130. K. Hugdahl, M. E. Raichle, A. Mitra, K. Specht, On the existence of a generalized non-specific task-dependent network. *Front. Hum. Neurosci.* **9** (2015).

131. S. Shashidhara, D. J. Mitchell, Y. Erez, J. Duncan, Progressive Recruitment of the Frontoparietal Multiple-demand System with Increased Task Complexity, Time Pressure, and Reward. *Journal of Cognitive Neuroscience* **31**, 1617–1630 (2019).

132. M. Assem, M. F. Glasser, D. C. Van Essen, J. Duncan, A Domain-General Cognitive Core Defined in Multimodally Parcellated Human Cortex. *Cerebral Cortex* **30**, 4361–4380 (2020).

133. J. Gläscher, *et al.*, Distributed neural system for general intelligence revealed by lesion mapping. *Proceedings of the National Academy of Sciences* **107**, 4705–4709 (2010).

134. A. Woolgar, *et al.*, Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. *Proceedings of the National Academy of Sciences* **107**, 14899–14902 (2010).

135. A. Woolgar, J. Duncan, F. Manes, E. Fedorenko, Fluid intelligence is supported by the multiple-demand system not the language system. *Nat Hum Behav* **2**, 200–204 (2018).

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

137. Y.-F. Liu (劉耘非), M. Bedny, Learning to Program "Recycles" Preexisting Frontoparietal Population Codes of Logical Algorithms. *J. Neurosci.* **45** (2025).

138. M. Assem, I. A. Blank, Z. Mineroff, A. Ademoglu, E. Fedorenko, Activity in the Fronto-Parietal Multiple-Demand Network is Robustly Associated with Individual Differences in Working Memory and Fluid Intelligence. *Cortex; a journal devoted to the study of the nervous system and behavior* **131**, 1 (2020).

139. V. Goel, R. J. Dolan, Functional neuroanatomy of three-term relational reasoning. *Neuropsychologia* **39**, 901–909 (2001).

140. C. Reverberi, T. Shallice, S. D'Agostini, M. Skrap, L. L. Bonatti, Cortical bases of elementary deductive reasoning: inference, memory, and metadeduction. *Neuropsychologia* **47**, 1107–1116 (2009).

141. H. Kean, *et al.*, A human brain network specialized for abstract formal reasoning. [Preprint] (2025). Available at: https://www.biorxiv.org/content/10.1101/2025.10.21.683445v1 [Accessed 10 February 2026].

142. L. Barham, D. Everett, Semiotics and the Origin of Language in the Lower Palaeolithic. *J Archaeol Method Theory* **28**, 535–579 (2021).

143. H. P. Grice, Utterer's Meaning, Sentence-Meaning, and Word-Meaning. *Foundations of Language* **4**, 225–242 (1968).

144. H. P. Grice, "Logic and Conversation" in *The Logic of Grammar*, D. Davidson, Ed. (Dickenson Pub. Co., 1975), pp. 64–75.

145. D. Sperber, D. Wilson, *Relevance: Communication and Cognition* (Blackwell, 1986).

146. B. Deen, K. Koldewyn, N. Kanwisher, R. Saxe, Functional Organization of Social Perception and Cognition in the Superior Temporal Sulcus. *Cereb Cortex* **25**, 4596–4609 (2015).

147. B. L. Pritchett, C. Hoeflin, K. Koldewyn, E. Dechter, E. Fedorenko, High-level language processing regions are not engaged in action observation or imitation. *Journal of Neurophysiology* **120**, 2555–2570 (2018).

148. O. Jouravlev, *et al.*, Speech-accompanying gestures are not processed by the language-processing mechanisms. *Neuropsychologia* **132**, 107132 (2019).

149. A. D. Patel, Language, music, syntax and the brain. *Nat Neurosci* **6**, 674–681 (2003).

150. W. T. Fitch, M. D. Martins, Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences* **1316**, 87–104 (2014).

151. C. Rogalsky, F. Rong, K. Saberi, G. Hickok, Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *J Neurosci* **31**, 3843–3852 (2011).

152. X. Chen, *et al.*, The human language system, including its inferior frontal component in "Broca's area," does not support music perception. *Cereb Cortex* **33**, 7904–7929 (2023).

153. S. L. Thompson-Schill, "Dissecting the Language Organ: A New Look at the Role of Broca's Area in Language Processing" in *Twenty-First Century Psycholinguistics: Four Cornerstones*, (Lawrence Erlbaum Associates Publishers, 2005), pp. 173–189.

154. E. Kaan, T. Y. Swaab, The brain circuitry of syntactic comprehension. *Trends in Cognitive Sciences* **6**, 350–356 (2002).

155. J. M. Novick, J. C. Trueswell, S. L. Thompson-Schill, Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. *Cogn Affect Behav Neurosci* **5**, 263–281 (2005).

156. J. M. Novick, E. Hussey, S. Teubner-Rhodes, J. I. Harbison, M. F. Bunting, Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience* **29**, 186–217 (2014).

157. D. January, J. C. Trueswell, S. L. Thompson-Schill, Co-localization of Stroop and Syntactic Ambiguity Resolution in Broca's Area. *J Cogn Neurosci* **21**, 2434–2444 (2009).

158. C. Shain, I. A. Blank, M. van Schijndel, W. Schuler, E. Fedorenko, fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia* **138**, 107307 (2020).

159. I. A. Quillen, M. Yen, S. M. Wilson, Distinct Neural Correlates of Linguistic and Non-Linguistic Demand. *Neurobiology of Language* **2**, 202–225 (2021).

160. L. Wehbe, *et al.*, Incremental Language Comprehension Difficulty Predicts Activity in the Language Network but Not the Multiple Demand Network. *Cereb Cortex* **31**, 4006–4023 (2021).

161. E. Fedorenko, C. Shain, Similarity of computations across domains does not imply shared implementation: The case of language comprehension. *Curr Dir Psychol Sci* **30**, 526–534 (2021).

162. G. F. Marcus, *The algebraic mind: Integrating connectionism and cognitive science* (The MIT Press, 2001).

163. J. McCarthy, P. J. Hayes, Some Philosophical Problems From the Standpoint of Artificial Intelligence. *Machine Intelligence* **4**, 463–502 (1969).

164. R. Kowalski, M. Sergot, A logic-based calculus of events. *New Gener Comput* **4**, 67–95 (1986).

165. J. Pinto, R. Reiter, Temporal Reasoning in Logic Programming: A Case for the Situation Calculus. (1993). https://doi.org/10.7551/mitpress/4305.003.0023.

166. J. G. de Villiers, P. A. de Villiers, "Linguistic determinism and the understanding of false beliefs" in *Children's Reasoning and the Mind*, (Psychology Press/Taylor & Francis (UK), 2000), pp. 191–228.

167. M. Amalric, S. Dehaene, Origins of the brain networks for advanced mathematics in expert mathematicians. *Proceedings of the National Academy of Sciences* **113**, 4909–4917 (2016).

168. H. H. Kean, *et al.*, Intuitive physical reasoning is not mediated by linguistic nor exclusively domain-general abstract representations. *Neuropsychologia* **213**, 109125 (2025).

169. L. M. DiNicola, R. L. Buckner, Adjacent Dorsolateral Prefrontal Cortex (DLPFC) Regions Participate in Distinct Large-Scale Networks Differentially Recruited for Social and Cognitive Control Functions. [Preprint] (2025). Available at: https://www.biorxiv.org/content/10.64898/2025.12.22.695973v1 [Accessed 10 February 2026].

170. R. M. Willems, Y. Benn, P. Hagoort, I. Toni, R. Varley, Communicating without a functioning language system: Implications for the role of language in mentalizing. *Neuropsychologia* **49**, 3130–3135 (2011).

171. E. Fedorenko, I. A. Blank, Broca's Area Is Not a Natural Kind. *Trends Cogn Sci* **24**, 270–284 (2020).

172. J. Du, *et al.*, Organization of the human cerebral cortex estimated within individuals: networks, global topography, and function. *Journal of Neurophysiology* **131**, 1014–1082 (2024).

173. P. Pinel, S. Dehaene, Beyond hemispheric dominance: brain regions underlying the joint lateralization of language and arithmetic to the left hemisphere. *J Cogn Neurosci* **22**, 48–66 (2010).

174. A. Caramazza, M. Coltheart, Cognitive Neuropsychology twenty years on. *Cognitive Neuropsychology* **23**, 3–12 (2006).

175. L. Hermer-Vazquez, E. S. Spelke, A. S. Katsnelson, Sources of flexibility in human cognition: dual-task studies of space and language. *Cogn Psychol* **39**, 3–36 (1999).

176. A. M. Newton, J. G. de Villiers, Thinking while talking: Adults fail nonverbal false-belief reasoning. *Psychological Science* **18**, 574–579 (2007).

177. G. Lupyan, Extracommunicative functions of language: Verbal interference causes selective categorization impairments. *Psychonomic Bulletin & Review* **16**, 711–718 (2009).

178. J. Dungan, R. Saxe, Matched false-belief performance during verbal and nonverbal interference. *Cogn Sci* **36**, 1148–1156 (2012).

179. T. L. Scott, T. K. Perrachione, Common cortical architectures for phonological working memory identified in individual brains. *Neuroimage* **202**, 116096 (2019).

180. C. Casto, A. Ivanova, E. Fedorenko, N. Kanwisher, What does it mean to understand language? *arXiv.org* (2025). Available at: https://arxiv.org/abs/2511.19757v1 [Accessed 10 February 2026].

181. S. C. Levinson, "Language and mind: Let's get the issues straight!" in *Language in Mind: Advances in the Study of Language and Thought*, (Boston Review, 2003), pp. 25–46.

182. M. C. Frank, D. L. Everett, E. Fedorenko, E. Gibson, Number as a cognitive technology: evidence from Pirahã language and cognition. *Cognition* **108**, 819–824 (2008).

183. E. Gibson, *et al.*, Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences* **114**, 10785–10790 (2017).

184. A. Majid, *et al.*, Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences* **115**, 11369–11376 (2018).

185. B. Pitt, *et al.*, Spatial concepts of number, size, and time in an indigenous culture. *Science Advances* **7**, eabg4141 (2021).

186. B. Pitt, A. Carstensen, I. Boni, S. T. Piantadosi, E. Gibson, Different reference frames on different axes: Space and language in indigenous Amazonians. *Science Advances* **8**, eabp9814 (2022).

187. D. M. O'Shaughnessy, E. Gibson, S. T. Piantadosi, The cultural origins of symbolic number. *Psychological Review* **129**, 1442–1456 (2022).

188. S. R. Waxman, "Linking object categorization and naming: Early expectations and the shaping role of language" in *The Psychology of Learning and Motivation: Advances in Research and Theory, Vol. 38*, (Academic Press, 1998), pp. 249–291.

189. B. Ferguson, S. Waxman, Linking language and categorization in infancy. *J Child Lang* **44**, 527–552 (2017).

190. S. Christie, D. Gentner, Development Of Language And Thought. *Cambridge Handbook Of Psycholinguistics* 653–673 (2012).

191. M. Casasola, When Less Is More: How Infants Learn to Form an Abstract Categorical Representation of Support. *Child Development* **76**, 279–290 (2005).

192. M. Bowerman, S. C. Levinson, *Language Acquisition and Conceptual Development* (Cambridge University Press, 2001).

193. M. Socher, E. Ingebrand, M. Wass, B. Lyxell, The relationship between reasoning and language ability: comparing children with cochlear implants and children with typical hearing. *Logopedics Phoniatrics Vocology* **47**, 73–83 (2022).

194. S. Malik-Moraleda, *et al.*, Constructed languages are processed by the same brain mechanisms as natural languages. *Proc Natl Acad Sci U S A* **122**, e2313473122 (2025).

195. M. Nye, *et al.*, Show Your Work: Scratchpads for Intermediate Computation with Language Models. [Preprint] (2021). Available at: http://arxiv.org/abs/2112.00114 [Accessed 28 October 2022].

196. T. Olausson, *et al.*, LINC: A Neurosymbolic Approach for Logical Reasoning by Combining Language Models with First-Order Logic Provers in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, H. Bouamor, J. Pino, K. Bali, Eds. (Association for Computational Linguistics, 2023), pp. 5153–5176.

197. X. Ye, Q. Chen, I. Dillig, G. Durrett, SatLM: Satisfiability-Aided Language Models Using Declarative Prompting. *Advances in Neural Information Processing Systems* **36**, 45548–45580 (2023).

198. G. Poesia, K. Gandhi, E. Zelikman, N. D. Goodman, Certified Deductive Reasoning with Language Models. [Preprint] (2023). Available at: http://arxiv.org/abs/2306.04031 [Accessed 24 July 2025].

199. L. Pan, A. Albalak, X. Wang, W. Wang, Logic-LM: Empowering Large Language Models with Symbolic Solvers for Faithful Logical Reasoning in *Findings of the Association for Computational Linguistics: EMNLP 2023*, H. Bouamor, J. Pino, K. Bali, Eds. (Association for Computational Linguistics, 2023), pp. 3806–3824.

200. N. Borazjanizadeh, S. T. Piantadosi, Reliable Reasoning Beyond Natural Language. [Preprint] (2024). Available at: http://arxiv.org/abs/2407.11373 [Accessed 24 July 2025].

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

202. M. Knauff, *Space to reason: A spatial theory of human thought* (The MIT Press, 2013).

203. M. Knauff, T. Mulack, J. Kassubek, H. R. Salih, M. W. Greenlee, Spatial imagery in deductive reasoning: a functional MRI study. *Cognitive Brain Research* **13**, 203–212 (2002).

204. K. L. Alfred, A. C. Connolly, J. S. Cetron, D. J. M. Kraemer, Mental models use common neural spatial structure for spatial and abstract content. *Commun Biol* **3**, 17 (2020).

205. J. Piaget, *The Language and Thought of the Child* (Kegan Paul, Trench, Trubner & Co., 1926).

206. F. C. Bartlett, *Remembering: A study in experimental and social psychology* (Cambridge University Press, 1932).

207. R. C. Schank, R. P. Abelson, *Scripts, plans, goals and understanding: An inquiry into human knowledge structures* (Lawrence Erlbaum, 1977).

208. G. H. Bower, J. B. Black, T. J. Turner, Scripts in memory for text. *Cognitive Psychology* **11**, 177–220 (1979).

209. A. Gilboa, H. Marlatte, Neurobiology of Schemas and Schema-Mediated Memory. *Trends Cogn Sci* **21**, 618–631 (2017).

210. R. Masís-Obando, K. A. Norman, C. Baldassano, Schema representations in distinct brain networks support narrative memory during encoding and retrieval. *Elife* **11**, e70445 (2022).

211. L. J. Rips, "Deduction" in *The Psychology of Human Thought*, (Cambridge University Press, 1988), pp. 116–152.

212. L. J. Rips, *The Psychology of Proof: Deductive Reasoning in Human Thinking* (A Bradford Book, 1994).

213. L. J. Rips, "Deduction and cognition" in *Thinking: An Invitation to Cognitive Science, Vol. 3, 2nd Ed*, An invitation to cognitive science (2nd ed.)., (The MIT Press, 1995), pp. 297–343.

214. S. T. Piantadosi, J. B. Tenenbaum, N. D. Goodman, Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition* **123**, 199–217 (2012).

215. S. T. Piantadosi, J. B. Tenenbaum, N. D. Goodman, The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological Review* **123**, 392–424 (2016).

216. N. D. Goodman, J. B. Tenenbaum, T. Gerstenberg, Concepts in a Probabilistic Language of Thought. 25 (2014).

217. K. Ellis, "Algorithms for learning to induce programs," Massachusetts Institute of Technology. (2020).

218. J. S. Rule, The child as hacker: Building more human-like models of learning. (2020).

219. S. Planton, *et al.*, A theory of memory for binary sequences: Evidence for a mental compression algorithm in humans. *PLOS Computational Biology* **17**, e1008598 (2021).

220. S. Dehaene, F. Al Roumi, Y. Lakretz, S. Planton, M. Sablé-Meyer, Symbols and mental programs: a hypothesis about human singularity. *Trends in Cognitive Sciences* **26**, 751–766 (2022).

221. P. C. Fletcher, *et al.*, Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition* **57**, 109–128 (1995).

222. H. L. Gallagher, *et al.*, Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia* **38**, 11–21 (2000).

223. R. Saxe, N. Kanwisher, People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind." *Neuroimage* **19**, 1835–1842 (2003).

224. R. Saxe, L. J. Powell, It's the Thought That Counts: Specific Brain Regions for One Component of Theory of Mind. *Psychol Sci* **17**, 692–699 (2006).

225. J. Fischer, J. G. Mikhael, J. B. Tenenbaum, N. Kanwisher, Functional neuroanatomy of intuitive physical inference. *Proceedings of the National Academy of Sciences* **113**, E5072–E5081 (2016).

226. S. Schwettmann, J. B. Tenenbaum, N. Kanwisher, Invariant representations of mass in the human brain. *eLife* **8**, e46619 (2019).

227. R. Pramod, M. A. Cohen, J. B. Tenenbaum, N. Kanwisher, Invariant representation of physical stability in the human brain. *eLife* **11**, e71736 (2022).

228. R. T. Pramod, E. Mieczkowski, C. X. Fang, J. B. Tenenbaum, N. Kanwisher, Decoding predicted future states from the brain's "physics engine." *Sci Adv* **11**, eadr7429 (2025).

229. D. Hassabis, E. A. Maguire, Deconstructing episodic memory with construction. *Trends Cogn Sci* **11**, 299–306 (2007).

230. R. N. Spreng, R. A. Mar, A. S. N. Kim, The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *J Cogn Neurosci* **21**, 489–510 (2009).

231. C. Baldassano, U. Hasson, K. A. Norman, Representation of Real-World Event Schemas during Narrative Perception. *J Neurosci* **38**, 9689–9699 (2018).

232. R. L. Buckner, L. M. DiNicola, The brain's default network: Updated anatomy, physiology and evolving insights. *Nature Reviews Neuroscience* **20**, 593–608 (2019).

233. B. Deen, W. A. Freiwald, Parallel systems for social and spatial cognition reaching the cortical apex. *Proceedings of the National Academy of Sciences* **122**, e2520067122 (2025).

234. U. Hasson, J. Chen, C. J. Honey, Hierarchical process memory: memory as an integral component of information processing. *Trends Cogn Sci* **19**, 304–313 (2015).

235. I. Dasgupta, S. J. Gershman, Memory as a Computational Resource. *Trends Cogn Sci* **25**, 240–251 (2021).

236. S. M. Wilson, *et al.*, Recovery from aphasia in the first year after stroke. *Brain* **146**, 1021–1039 (2022).

237. J. Harlow, Passage of an Iron Rod through the Head. *The Boston Medical and Surgical Journal* **39**, 389–393 (1848).

238. J. M. (John M. Harlow, Recovery from the passage of an iron bar through the head. *Massachusetts Medical Society* (1869).

239. M. P. Broca, Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). *Bulletin et mémoires de la Société Anatomique de Paris* **6**, 330–357 (1861).

240. W. B. Scoville, B. Milner, Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery & Psychiatry* **20**, 11–21 (1957).

241. J. R. Cox, R. A. Griggs, The effects of experience on performance in Wason's selection task. *Memory & Cognition* **10**, 496–502 (1982).

242. R. A. Griggs, J. R. Cox, The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology* **73**, 407–420 (1982).

243. L. Cosmides, The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* **31**, 187–276 (1989).

244. G. Gigerenzer, K. Hug, Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition* **43**, 127–171 (1992).

245. G. R. Yang, M. R. Joglekar, H. F. Song, W. T. Newsome, X.-J. Wang, Task representations in neural networks trained to perform many cognitive tasks. *Nat Neurosci* **22**, 297–306 (2019).

246. R. Csordás, S. van Steenkiste, J. Schmidhuber, Are Neural Nets Modular? Inspecting Functional Modularity Through Differentiable Weight Masks. [Preprint] (2021). Available at: http://arxiv.org/abs/2010.02066 [Accessed 24 July 2025].

247. M. A. Lepori, E. Pavlick, T. Serre, NeuroSurgeon: A Toolkit for Subnetwork Analysis. [Preprint] (2023). Available at: http://arxiv.org/abs/2309.00244 [Accessed 24 July 2025].

248. M. Lepori, T. Serre, E. Pavlick, Break it down: Evidence for structural compositionality in neural networks. *Advances in Neural Information Processing Systems* **36**, 42623–42660 (2023).

249. B. AlKhamissi, C. N. D. Sabbata, Z. Chen, M. Schrimpf, A. Bosselut, Mixture of Cognitive Reasoners: Modular Reasoning with Brain-Like Specialization. [Preprint] (2025). Available at: http://arxiv.org/abs/2506.13331 [Accessed 24 July 2025].

250. J. R. Crawford, D. C. Howell, Comparing an Individual's Test Score Against Norms Derived from Small Samples. *The Clinical Neuropsychologist* **12**, 482–486 (1998).

251. B. Lipkin, *et al.*, Probabilistic atlas for the language network based on precision fMRI data from >800 individuals. *Sci Data* **9**, 529 (2022).

252. V. C. Zimmerer, P. E. Cowell, R. A. Varley, Artificial grammar learning in individuals with severe aphasia. *Neuropsychologia* **53**, 25–38 (2014).