Constructed languages are processed by the same brain mechanisms as natural languages

Saima Malik-Moraleda[1,2,3], Maya Taliaferro[1,2], Steve Shannon[2], Niharika Jhingan[1,2], Sara Swords[1,2], David J. Peterson, Paul Frommer [4], Marc Okrand, Jessie Sams, Ramsey Cardwell[5], Cassie Freeman[5], Evelina Fedorenko[1,2,3]

[1] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139
[2] McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139
[3] Program in Speech and Hearing Bioscience and Technology, Harvard University, Boston, MA 02114
[4] Marshall School of Business, University of Southern California, Los Angeles, CA 90089
[5] Duolingo, Pittsburgh, PA15206

Corresponding authors: Saima Malik-Moraleda (smalikmoraleda@g.harvard.edu) and Ev Fedorenko (evelina9@mit.edu).

CRediT statement:

Conceptualization: SMM, EF
Methodology: SMM, MT, EF
Software: SMM, MT
Investigation: SMM, MT, SSh, NJ, SSw
Data curation: SMM, MT
Formal analysis: SMM
Validation: SMM, MT
Resources: DJP, PF, MO, JS, RC, CF
Writing-original draft: SMM, EF
Writing-review and editing: all authors
Visualization: SMM, MT
Supervision and project administration: EF

Abstract

What constitutes a language? Natural languages share some features with other domains: from math, to music, to gesture. However, the brain mechanisms that process linguistic input are highly specialized, showing little or no response to diverse non-linguistic tasks. Here, we examine constructed languages (conlangs) to ask whether they draw on the same neural mechanisms as natural languages, or whether they instead pattern with domains like math and logic. Using individual-subject fMRI analyses, we show that understanding conlangs recruits the same brain areas as natural language comprehension. This result holds for Esperanto (n=19 speakers)— created to resemble natural languages—and fictional conlangs (Klingon (n=10), Na'vi (n=9), High Valyrian (n=3), and Dothraki (n=3)), created to differ from natural languages, and suggests that conlangs and natural languages share critical features and that the notable differences between conlangs and natural language are not consequential for the cognitive and neural mechanisms that they engage.

Introduction

The question of what constitutes a 'language' has long been debated. Some have discussed this question in the context of the relationship between human languages and non-human animal communication systems (e.g., Hockett, 1959; Premack, 1985; Swadesh, 2017; Call & Tomasello, 2020). Others have argued about whether different human cultural inventions—from math, to music and dance, to cooking, to comics, to programming languages—can or should be considered a kind of language (e.g., McGee, 1997; Hagendoorn, 2010; Patel, 2010; Cohn, 2013; Bers, 2019).

To answer the question of what counts as a 'language', one may attempt to provide a set of criteria that must be satisfied for a system to be considered a language. But evaluating any such definition is challenging: one can always argue that the definition should be broader or narrower in scope, include a particular criterion or not, and so on. Alternatively, this question can be answered empirically, by asking whether the cognitive and neural mechanisms that are used to process natural human languages are also used to process information from other systems, like math or music. If so, then presumably, some aspects of the knowledge representations or processing algorithms are sufficiently similar to draw on the same neural circuits.

Over the last few decades, brain imaging studies and investigations of patients with brain damage have revealed that the brain areas that process language are exquisitely selective for language. Using fMRI, researchers have found that language-responsive brain areas—a set of frontal and temporal areas in the left hemisphere—are not engaged when individuals solve math problems, listen to music, or engage in a wide variety of other tasks (e.g., Monti et al., 2009, 2012; Fedorenko et al., 2011; Amalric & Dehaene, 2019; Ivanova et al., 2021; Chen et al., 2023). Similarly, some patients with severe damage to the language system appear to be able to do math, to create and appreciate music, and to reason about diverse domains (e.g., Varley & Siegal, 2000; Varley et al., 2005; Apperly et al., 2007; see Fedorenko & Varley, 2016 for a review). Thus, in spite of conceptual parallels that can be drawn between natural languages and other cultural inventions, and in spite of the fact that these inventions are often casually talked about as a kind of 'language', the machinery that supports our linguistic abilities is not shared with information processing in other domains. Even systems that share numerous features with natural language, like programming languages (e.g., Bers, 2019; Fedorenko et al., 2019) do not appear to draw on the language-processing mechanisms (e.g., Ivanova et al., 2020; Liu et al., 2020; Srikant, Lipkin et al., 2022).

Here we examine a class of stimuli that have been a part of human culture for at least four centuries (Okrent, 2010) but that, to our knowledge, have not been previously examined in cognitive neuroscience: constructed or invented languages, or conlangs. Conlangs are communication systems that, like natural languages, can be used to talk about the outer world

(the objects/entities and events) and the inner world (our feelings, thoughts, and attitudes). However, conlangs differ from natural languages in several important respects. ***First***, the earliest natural languages emerged sometime between 100,000 and 1,000,000 years ago (e.g., Perreault & Mathew, 2012; Dediu & Levinson, 2013; Barham & Everett, 2021), likely naturally and gradually, through interactions among groups of individuals (e.g., Jackendoff, 1999; Kirby et al., 2008; Senghas et al., 2014; Ergin et al., 2020). Conlangs, on the other hand, are more recent and intentional inventions (see Senghas et al., 2014 and Ergin et al., 2020 for examples of relatively recently created natural languages), typically created by a single individual (Okrent, 2010). ***Second***, although it remains debated *why* language emerged in the history of human evolution (e.g., Jackendoff, 1999, 2010; Fitch, 2017; Fedorenko et al., subm.), the pressures that led to the emergence of human language likely differ from those that lead to the creation of conlangs. In particular, some conlangs are created to facilitate communication among speakers of diverse languages (e.g., Esperanto, 1887), others – to help create an identity for fictional species (e.g., Klingon, Na'vi, or Tolkien's Elvish languages) or a fictional people (e.g., Dothraki), yet others – simply as a hobby (for a review, see Goodall, 2022). ***Third***, natural languages are strongly shaped by learning and processing pressures (e.g., Kirby, 2017; Gibson et al., 2019; Levshina, 2022; Fedorenko et al., subm.). In contrast, at least for some conlangs (especially those that are created as part of an identity for a fictional group), such pressures do not apply: if human actors have to produce utterances in the conlang, they typically simply memorize them rather than learning the words and their meanings and the rules for how the words combine into phrases and sentences. Of course, conlangs that end up being used by groups of individuals for communication do have to be learnable (at least as a second language late in life). In fact, one recent conlang (Ithkuil; Quijada, 2011), which was created to be more "logical" (whatever that may mean) and less ambiguous than natural languages, turned out to be not learnable in its original form and ended up being simplified across several iterations (Joshua, 2012). In any case, even conlangs that are used communicatively are typically (with the exception of Esperanto) spoken by small groups of individuals and often in specific contexts, which may be insufficient for the processing pressure considerations to shape the language (e.g., Lupyan & Dale, 2010).

Given all these differences between natural and constructed languages, we asked whether the brain areas that process natural languages are also engaged when individuals process conlangs that they are proficient in. One salient alternative possibility is that conlangs instead draw on a brain network that supports diverse goal-directed behaviors and has been linked to fluid intelligence and learning new skills—the so-called Multiple Demand network (e.g., Duncan, 2010, 2013; Duncan et al., 2020; Assem et al., 2020). The Multiple Demand network is also the network that supports mathematical and logical reasoning and comprehension of programming languages (e.g., Monti et al., 2009, 2012; Fedorenko et al., 2013; Amalric & Dehaene, 2019; Ivanova et al., 2020; Liu et al., 2020). We tested speakers of five different constructed languages—Esperanto and four fictional conlangs (Klingon, Na'vi, High Valyrian, and Dothraki)—which differ in the purpose for which they were created, the span of time they have

been spoken, and the size of the community that uses the language (Table 1). To foreshadow our results, we found that all five conlangs recruit the same brain areas that support the processing of diverse natural languages.

Conlang	**Inception**	**Number of users**	**Purpose of its creation**
Esperanto	1887	~60,000	Facilitate international communication
Klingon	1985	~2,000	Fiction
Na'vi	2009	~100	Fiction
High Valyrian	2011	~50	Fiction
Dothraki	2011	~10	Fiction

Table 1: Properties of the conlangs that are included in the present study.

Results

Behavioral:

Participants reported a range of proficiency levels in their conlang(s), with most participants self-reporting moderate to high proficiency (73.17% of the participants reported proficiency of 3 and above on the 1-6 CEFR scale; M=3.7, range: 1.2-5.8; **Figure 1a**). Higher proficiency levels were reported for Esperanto (*M*=4.0, range: 1.4-5.8), Klingon (*M*=3.3, range: 1.6-5.4), and Na'vi (*M*=4.2, range: 2.6-5.2), compared to High Valyrian (*M*=1.9, range: 1.2-3.2) and Dothraki (*M*=2.8, range: 1.2-4.0). For the open-ended written production task, participants produced 61.68 words on average and 15.83 words per minute on average (Esperanto: *M*=19.26, range: 4.17-44.29; Klingon: *M*=9.26, range: 1.60-23.37; Na'vi: *M*=15.82, range: 2.60-29.78*;* Dothraki: *M*=9.50, range: 4.73-15.75; High Valyrian was excluded from these analyses because participants in this group did not consider themselves proficient enough to perform the task). And for the lexical decision task, the average accuracy was 0.90 for Esperanto speakers and 0.90 for Klingon speakers (both reliably greater than chance; ts>7, ps<0.001). Furthermore, across the full set of participants, the self-reported proficiency scores were predictive of both the words per minute scores on the production task (r=0.53; t(36)=3.73, p<0.001) and the lexical decision accuracies (r=0.50; t(23)=2.76, p=0.01; **Figure 1b**).

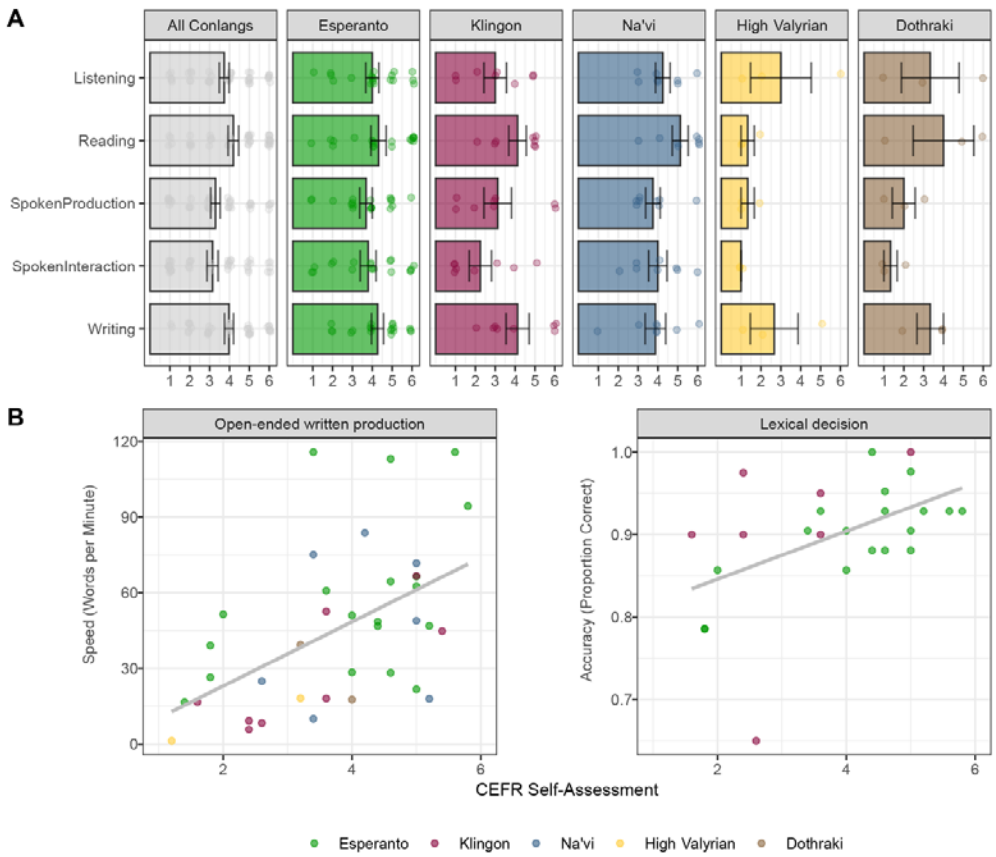

Figure 1: Self-reported proficiencies are in the moderate to high range for most participants and are predictive of open-ended written production ability and lexical decision task accuracies. A) The average self-report CEFR scores for each conlang and across conlangs in the five domains (listening, reading, spoken interaction, spoken production, and writing). Dots represent individual participant CEFR scores. **B)** Correlation of overall CEFR scores (averaged across the five domains) to the open-ended written production task and the lexical decision task. Dots represent individual participants and are colored according to the conlang.

fMRI:

In line with past results, the language fROIs (defined by the Sentences > Control contrast from the reading-based English localizer) responded more strongly during the *Sentences* condition than the *Control* (nonwords) condition (2.16% vs 0.77% BOLD signal change relative to the fixation baseline; β=1.40, $p<0.001$; see **Supp. Table 5** for responses in each conlang group separately). The language fROIs also responded more strongly during the *Sentences* than the *Control* condition in the listening-based localizer (2.02 vs. 0.59% BOLD signal change relative to fixation; β=1.43, $p<0.001$; **Supp. Table 5**).

Critically, the language fROIs showed a robust response during the comprehension of constructed languages (**Figures 2** and **3**; see **Supp. Figure 2** for responses broken down by fROI). The *Sentences* condition elicited a stronger response than the *Control* condition (1.96 vs. 0.75% BOLD signal change relative to fixation; β=1.22, $p<0.001$). This critical result held for

each of the five conlangs tested (**Figure 1**): Esperanto (2.16 vs. 0.89; β=1.27, p<0.001), Klingon (1.36 vs. 0.41; β=0.96, p<0.001), Na'vi (2.67 vs 1.07; β=1.61, p<0.001), High Valyrian (1.37 vs. 0.45; β=0.92, p<0.001) and Dothraki (1.86 vs. 0.73, β=1.13, p<0.001).

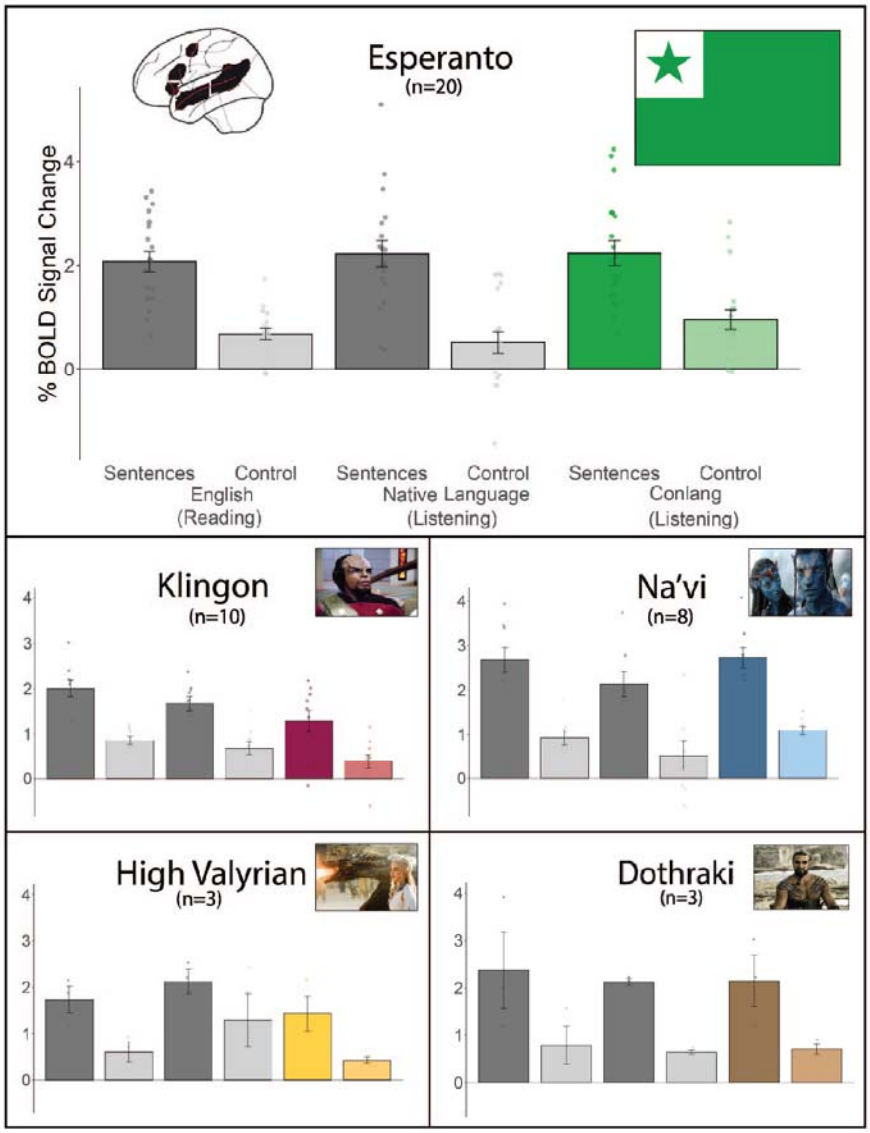

Figure 2. The language areas show a strong response during the processing of both natural and constructed languages. Percent BOLD signal change across the five language fROIs (the results are similar in each fROI separately; **Supp. Figure 2**) for the reading-based and the listening-based versions of the language localizer (dark grey bars = sentences, light grey bars = control condition; the reading-based version used English materials, in which all participants were native or proficient, and the listening-based version used the participants' native language materials) and the critical conlang comprehension task, where participants listened to sentences in the constructed language (colored bars) and to control, acoustically degraded versions of those sentences (lighter versions of the colored bars). The language network is defined by the Sentences>Control condition of the English reading task in the language-dominant hemisphere (left hemisphere for 36 of the 38 participants), but the results are similar when the language network is defined by the Sentences>Control contrast of the listening language localizer

task (**Supp. Figure 1**) or when the left-hemisphere fROIs are used for all participants (**Supp. Figure 3**). Dots represent individual participants; error bars represent standard errors of the mean by participant.

To further characterize the similarity of brain responses to natural vs. constructed languages, we created binarized versions of the activation maps for the Sentences>Control contrast for the native language and the relevant conlang where—restricting the analysis to the LH language parcels—voxels that belonged to the set of 10% of most responsive voxels were assigned a value of 1 and voxels that did not were assigned a value of 0 (**Figure 3**). We then computed the Dice coefficient (Rombouts et al., 1997), which ranges from 0 (no overlapping voxels) to 1 (all voxels overlap) between a) the native language map and the conlang map within each participant (n=38 pairs), and b) the native language maps between each pair of participants (n=703 pairs). We found that the sets of most responsive voxels are more similar (show greater overlap) within an individual between a natural and a constructed language (0.45) than between individuals for a natural language (0.31; β=0.23, $p<0.001$; see **Figure 3** for sample maps). This result further supports the claim that conlangs are represented and processed in a similar brain network to a native language within individual speakers.

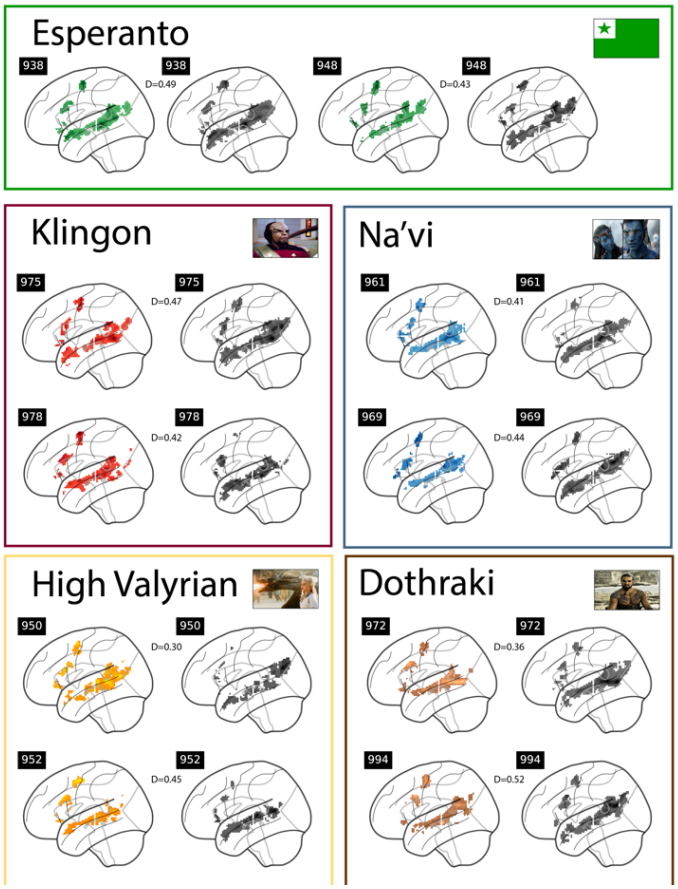

Figure 3. The topography of the language network's response is similar between constructed languages (different colors, left map for each participant) and natural languages (grey, right map for each participant).

For each conlang, we show two sample participants (participant ID—according to a lab-internal numbering system—is shown in small black boxes; these can be cross-referenced with the data tables on OSF). Within each parcel of the language network and for each language, the 10% of most language-responsive voxels are shown. The D values correspond to the Dice coefficient value (computed across the brain, as described in the text) for the comparison between the activations in response to the participant's native language vs. their conlang. All the individual maps are available at https://osf.io/pe74v/.

As discussed in the Introduction, language areas have been shown to be selective for language processing relative to diverse cognitive tasks (e.g., Fedorenko et al., 2011). Here, we replicate this finding in this new set of participants with respect to two non-linguistic tasks and directly compare the magnitude of the response for conlang comprehension to these non-linguistic conditions. As can be seen in **Figure 4**, the response in the language fROIs to the four non-linguistic conditions (a hard and an easy condition of a spatial working memory task and a mental and a physical condition of a naturalistic Theory of Mind paradigm) falls below the fixation baseline (in line with e.g., Mineroff, Blank et al., 2018; Malik-Moraleda, Ayyash et al., 2022; Shain, Paunov, Chen et al., 2022). The response to the sentence condition in the conlang experiment is comparable to the magnitude of response to the sentence condition in the two language localizer paradigms and is reliably higher than each of the four non-linguistic conditions (ps<0.001; see **Supp. Table 5** for the results broken down by conlang; see **Supp. Figure 4** for evidence that the non-linguistic paradigms elicit the expected response in the Multiple Demand and Theory of Mind areas; see **Supp. Figure 5** for the whole-brain overlap analysis between conlang activations and the language localizer activations vs. non-linguistic task activations).

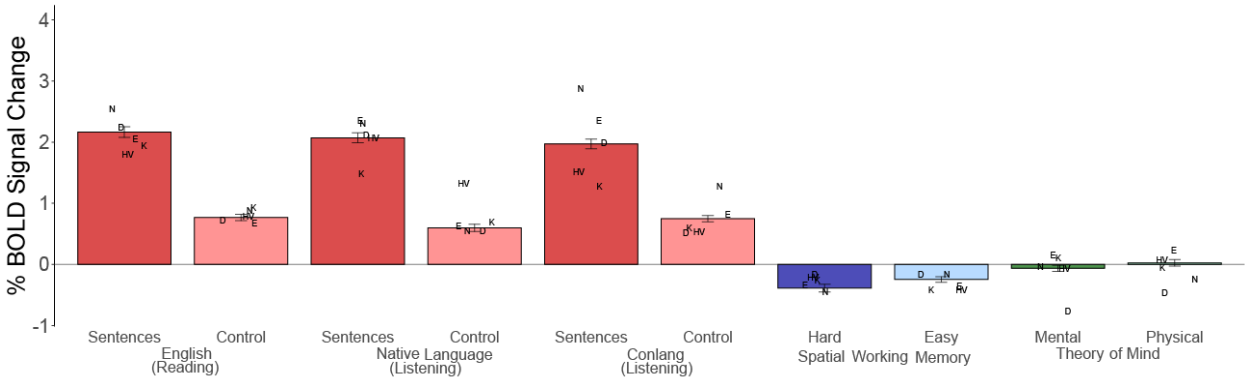

Figure 4. The language areas of conlang speakers are selective for language: the response to conlangs is similar in magnitude to the response to natural languages and reliably higher than the response to non-linguistic tasks. The letters correspond to the average response for each constructed language (E: Esperanto; K: Klingon; N: Na'vi; HV: High Valyrian; D: Dothraki).

Discussion

In an fMRI investigation of neural responses to constructed languages (conlangs), we found that conlangs recruit the same brain areas—frontal and temporal areas in the language-dominant hemisphere— as those that support the representation and processing of natural languages. This result held for both Esperanto and fictional languages. Esperanto is a conlang that was created over 130 years ago to facilitate international communication, was modeled on natural languages, and has a substantial number of speakers. In contrast, the four fictional conlangs (Klingon, Na'vi, Dothraki, and High Valyrian) were created within the last 12-40 years (**Table 1**), were constructed so as to differ from natural languages (two of them were created for fictional humanoid alien species), and have relatively small communities of speakers. Although the five conlangs differ in the strength of the response they elicit in the language areas (with Esperanto and Na'vi eliciting higher responses than Klingon, High Valyrian, and Dothraki), these differences are likely due to differences in the proficiency levels among the speakers (e.g., see Malik-Moraleda, Jouravlev et al., 2023). Below we discuss some implications of these findings for understanding the human language system.

We started this paper with a discussion of how to determine whether some information system can/should be considered a *language*. We advocated an empirical approach: testing whether some system of interest engages the same neural machinery as natural languages. If so, then we can infer that the system in question shares some aspects of representation and/or computation with natural languages. Past work has shown that the brain areas that respond during the comprehension and production of one's native language are similar across speakers of typologically diverse languages (Malik-Moraleda, Ayyash et al., 2022) and support the processing of multiple languages in bilinguals and multilinguals/polyglots (e.g., Illes et al., 1999; Chee et al., 1999, 2000; Roux & Trétmoulet, 2002; Klein et al., 2006; Sulpizio et al., 2020; Malik-Moraleda et al., 2023). However, diverse non-linguistic stimuli that share features with natural languages engage non-overlapping neural systems (e.g., Fedorenko et al., 2011; Monti et al., 2012; Chen et al., 2023). Here, we found that comprehension of constructed languages recruits the same brain areas as comprehension of natural languages. This finding suggests that the language network's engagement is robust to the features that differ between constructed and natural languages. In particular, to elicit a robust response in the language-processing areas, a language apparently does not need to have been around for very long (e.g., Dothraki and High Valyrian have only existed for a ~decade), does not need to have emerged in a community (and could instead have been created by a single individual, including for relatively esoteric purposes, like creating an identity for a fictional alien species), and does not need to be used by an extensive community of speakers. It also does not need to have been shaped by learning pressures, at least not by the kinds of pressures that act on first language acquisition.

Of course, it is important to remember that conlangs are created by proficient users of natural language and are therefore typically modeled, in some way, on natural languages, even when trying to create languages with unusual properties. In fact, conlang creators are often individuals with heightened linguistic awareness, obtained through acquisition of multiple natural languages and/or formal linguistic training (Schor, 2016). Knowledge of, or at least, familiarity with, multiple natural languages allows conlang creators to borrow and combine in creative ways features of different languages to create a novel language. In this way, conlangs would be expected to share numerous features with natural languages.

So what are the critical features of natural languages that are shared with constructed languages and differentiate both types of languages from other systems, like math, music, or computer code? Both natural and constructed languages are systems of symbolic form–meaning mappings that can be used to express an open-ended range of complex ideas by virtue of properties like duality of patterning (the existence of combinatorial structure at two levels: meaningless sounds can be combined into words, and words can be combined into phrases and sentences; Hockett, 1959) and compositionality (e.g., Krifka, 2001). But math and programming languages allow for the expression of an open-ended range of complex meanings as well, so why doesn't the processing of mathematical expressions or pieces of computer programs elicit a response in the language brain areas? One salient difference has to do with the *nature and scope of meanings* that these systems can express. Natural and constructed languages express meanings that relate to the physical and social worlds, including the objects and agents, their properties, the actions they can undertake, the interactions among them, and the agents' (including our own) internal emotional and mental states. In this way, language provides a mechanism for sharing our knowledge and beliefs about the world, as well as our feelings and emotions with each other. If we construe our language knowledge as a set of pointers to particular parts of our semantic space, it becomes clear that the language system needs strong connections with i) perception (through which we learn and process information about the world), ii) action (through which we interact physically with the world), and iii) cognition (our systems of knowledge and reasoning about the world). In contrast, systems like math and programming languages are much more restricted in the kinds of meanings they can express and bear little relationship to our experiences in the world, including interactions with conspecifics. Such systems instead express highly abstract, often relational, concepts and are designed to fulfill a particular cultural need.

Our finding that conlangs, including those created in very recent history and only used by a small community of speakers, recruit the same brain mechanisms as those that support natural language processing aligns well with the body of work by Kirby and colleagues on iterated learning in the domain of languages (e.g., Smith et al., 2003; Kirby et al., 2008, 2014, 2015; see Kirby, 2017 for an overview). Iterated learning is a particular kind of cultural transmission, where an individual learns a new behavior by observing others' behaviors (which were, in turn, learned by observing others' behaviors). Iterated learning experiments and simulations have

shown that language-like communication systems can emerge quickly (within a couple of 'generations' of learners) and that such systems exhibit core features of natural languages, including compositionality. Of course, whether or not compositionality is *required* for a communication system to engage the language brain areas remains to be established. To the best of our knowledge, no work has been carried out on the processing of holistic languages (e.g., Wray, 2000)—languages made up of random pairings of signals and meanings—in cognitive neuroscience.

More generally, numerous questions remain about the necessary and sufficient conditions for an information transmission system to be represented and processed by the mechanisms that process natural languages. For example, what about conlangs that deviate substantially from natural languages? One relatively recent example is Ithkuil (Quijada, 2004-2019)—a conlang whose goal was to allow for the expression of complex ideas briefly but clearly and precisely. To allow for concise and unambiguous expressions, the original version of Ithkuil made use of 65 consonants and 17 vowels (a large sound inventory allows any given word to be short without introducing ambiguity; Piantadosi et al., 2012), 96 grammatical cases, and a highly complex morphophonemic writing system. Following complaints from those trying to learn the language, Quijada has been simplifying the language across multiple versions (Wikipedia, 2023). To the extent that any individual is able to learn Ithkuil (or at least to achieve moderate proficiency), would the processing of such a conlang draw on the language system, or would it instead pattern with math and computer languages?

Another question concerns the scope of meanings that a communication system spans. All natural languages and all the conlangs included in the current study can be used to express any, or almost any, meaning. However, it is possible to imagine a communication system (including one with properties like compositionality) that can only be used to talk about a particular class of objects or events. Would such semantically narrow systems engage language-processing mechanisms? We mentioned above that the nature and scope of expressible meanings differ between language and systems like math and computer code. Evaluation of communication systems with restricted sets of meanings will tell us whether it is the *nature* or the narrow *scope* of the meanings that makes math and computer languages draw on a distinct brain system.

Yet another question concerns communication systems that are not considered full-blown languages, including pidgins—communication systems that are created when two or more groups of speakers of different languages are brought together and need to interact (e.g., Bickerton, 1976; Muysken & Smith, 1995), homesign systems used by deaf children for communicating with hearing caretakers in the absence of access to sign language (e.g., Goldin-Meadow, 1993; Brentari & Coppola, 2013), and languages in the early stages of creation (e.g., Senghas et al., 2014; Ergin et al., 2020). Such systems can express quite a wide array of meanings but lack some of the structure and systematicity that characterizes fully developed languages. If structure and

systematicity are critical for engaging the language processing mechanisms, then such systems may be expected to not draw on the language areas, but if flexible expression of complex meanings is the driving force, they should draw on the language system.

To conclude, the human brain contains a network of areas that are highly selective for the processing of language: they are recruited when we understand and produce language (e.g., Menenti et al., 2010; Hu, Small et al., 2022) but show little or no response when we engage in diverse non-linguistic behaviors, from math and computer code, to music, to interpreting facial expressions and gestures, to reasoning about others' minds (e.g., see Fedorenko & Varley, 2016 and Fedorenko et al., subm, for reviews). Here, we evaluated brain responses to five constructed languages and showed that they engage the same brain areas as those that support the processing of natural languages, despite several notable differences between conlangs and natural languages. This work suggests that conlangs share some critical features of natural languages, including the ability to express diverse meanings about external and internal worlds. By testing neural responses to a wider array of naturally emerging and artificially constructed communication systems, future work should continue to distill the necessary and sufficient conditions for being a language according to this empirical criterion.

Methods

General approach. Each participant—all proficient speakers of one or more conlangs (see Participants for details)—performed several tasks in an fMRI scanner. The general approach we adopt relies on functionally identifying areas of interest in each individual participant (e.g., Saxe et al., 2006; Fedorenko et al., 2010). This approach has been shown to confer greater sensitivity, functional resolution, and interpretability than the traditional group-averaging fMRI approach (e.g., Saxe et al., 2006; Nieto-Castañón & Fedorenko, 2012; Fedorenko, 2021). As detailed below, each participant performed a) two 'localizer' tasks for the language network: a reading-based version (reading English sentences vs. nonword sequences; Fedorenko et al., 2010) and a listening-based version (listening to short passages in one's native language vs. to acoustically degraded versions of those passages; Malik-Moraleda, Ayyash et al., 2022); and b) the critical conlang task that was similar to the listening-based localizer but used the materials in the relevant conlang. In addition, participants performed c) two non-linguistic tasks, which were used as additional controls to which neural responses to conlangs could be compared (and could also be used for identifying additional candidate networks that could support conlang comprehension).

In addition to the fMRI component, participants self-assessed their conlang proficiency using the Common European Framework of Reference (CEFR; Council of Europe, 2001) grid and wrote short passages (timed) about three topics in the conlang(s) they were proficient in. Esperanto and

Klingon speakers additionally completed a lexical decision task.

Participants. Speakers of constructed languages (all native or highly proficient speakers of English) were recruited online and through word of mouth. Participants were selected based on a) their self-rated proficiency in the constructed language of interest on an informal scale from 1 ("know a couple words") to 5 ("can understand almost everything and express myself freely") (see **Supp. Table 1** for details) and b) geographical proximity to Boston, where the testing took place. Overall, 38 individuals (25 males; mean age = 32.31, SE = 1.83) participated in the study, five of whom spoke and were tested on two constructed languages. The pool of participants included speakers of Esperanto (n = 19, 12 males, mean age = 30.47, SE = 2.07), Klingon (n = 10, 6 males, mean age = 40.30, SE=3.64), Na'vi (n = 8, 5 males, mean age = 28.62, SE = 1.89), High Valyrian (n = 3, 1 male, mean age = 40.67, SE =5.49), and Dothraki (n = 3, 2 males, mean age = 37.00, SE = 3.06). A substantial number of participants (17 of the 38 participants) were polyglots, as defined by having some proficiency in 5 or more languages (e.g., Jouravlev et al., 2019). Most participants were right-handed (n = 33), as determined by the Edinburgh Handedness Inventory (n = 35) or by self-report (n=3), and had normal or corrected-to-normal vision. All participants gave informed written consent in accordance with the requirements of MIT's Committee on the Use of Human as Experimental Subjects, were paid for their participation and, if coming from outside of the Boston area, reimbursed for travel expenses.

Experimental tasks.

Language localizers

We included two localizer tasks: a visual (reading-based) version (e.g., Fedorenko et al., 2010), which is most commonly used in the Fedorenko lab (e.g., see Lipkin et al., 2022 for data from >600 participants on this version), and an auditory (listening-based) version. The latter was included because 2 of the participants were not native speakers of English, and although we have previously established that an English-based localizer works well for proficient speakers (Malik-Moraleda, Ayyash et al., 2022), we wanted to additionally include a localizer in their native language, and we only have *auditory* versions of localizers for those languages. In the main analyses, we used the reading-based version to define the language areas, and then examined the responses to the conditions of both localizers, to the conditions of the critical task, and the non-linguistic tasks (see **Supp. Figure 1** for evidence that the results are similar when the listening-based version of the localizer is used, as has been previously established; e.g., Fedorenko et al., 2010; Malik-Moraleda, Ayyash et al., 2022; Chen et al., 2023).

Reading-based language localizer task. Participants passively read English sentences and lists of pronounceable nonwords (the control condition) in a blocked design. The Sentences > Control contrast targets brain regions that support high-level linguistic processing. Each trial started with

100-ms pre-trial fixation, followed by a 12-word-long sentence or a list of 12 nonwords presented on the screen one word/nonword at a time at the rate of 450 ms per word/nonword. Then, a line drawing of a finger pressing a button appeared for 400 ms, and participants were instructed to press a button whenever they saw this icon. Finally, a blank screen was shown for 100 ms, for a total trial duration of 6 s. The simple button-pressing task was included to help participants stay awake and focused. Each block consisted of three trials and lasted 18 s. Each run consisted of 16 experimental blocks (eight per condition) and five fixation blocks (14 s each), for a total duration of 358 s (5 min 58 s). Each participant performed two runs. Condition order was counterbalanced across runs. This localizer is available from the Fedorenko laboratory website: https://evlab.mit.edu/funcloc/.

Listening-based language localizer task. Participants listened to passages from Alice in Wonderland (in their native language; English for n=36, German for n=1, and Dutch for n=1) and to acoustically degraded versions of the passages (the control condition). The acoustically degraded versions were created as described in Scott et al. (2017; see also Malik-Moraleda, Ayyash et al., 2022; **Supp. Text 1**) and sound like poor radio reception such that the linguistic content is not discernible. Similar to the reading-based version, the Sentences > Control contrast targets brain regions that support high-level linguistic processing. Each block (consisting of an intact or a degraded passage) lasted 18 s. Each run consisted of 12 experimental blocks (6 per condition) and 3 fixation blocks (12 s each), for a total duration of 252 s (4 min 12 sec). Each participant performed two runs. Condition order was counterbalanced across runs. The sound was delivered through Sensimetrics earphones (model S14). This localizer is available from the Fedorenko laboratory website: https://evlab.mit.edu/aliceloc.

Critical conlang task

Materials. For each conlang, a set of 100 phrases and sentences was recorded (all the materials are available at: https://osf.io/pe74v/, see **Supp. Table 2** for the English translations). The materials differed across languages, but all sentences were relatively simple, short (**Supp. Table 3**), semantically and syntactically diverse, and—for the fiction conlangs—did not include any show-/movie-specific references. For all conlangs except for High Valyrian, the materials were recorded by a highly proficient male speaker. For Esperanto and Klingon, initial sentence recordings were provided by researchers at Duolingo, but the sentences were then re-recorded by proficient male speakers; for Na'vi, P.F. recorded the sentences; for Dothraki, D.P. recorded the sentences. For High Valyrian, due to the low number of proficient speakers, we used the sentence recordings provided by Duolingo (recorded by several male speakers). Acoustically degraded versions of the materials (for the control condition) were created in the same way as the degraded versions in the auditory version of the language localizer (**Supp. Text 1**). The volume levels were normalized across all clips. The recordings are available from the authors upon request.

Procedure. For each conlang, the sentences were concatenated in order to create 16 blocks of similar length; a similar procedure was applied to the acoustically degraded versions of the sentences. Each block consisted of 5-8 sentences/degraded sentences (see **Supp. Table 3** for details by language), with a 0.5 second pause between sentences, and was between 20 and 24 seconds in duration depending on the conlang (**Supp. Table 3**). For blocks that were a little longer than the target length, a gradual volume fade-out was applied during the last 2 seconds of the block. The 32 blocks (16 Sentences, 16 Control) were further divided into two sets corresponding to two scanning runs, with each run containing 16 blocks (8 Sentences, 8 Control) as well as five 12 seconds-long fixation blocks and lasting between ~6 and 7.5 minutes depending on the conlang (**Supp. Table 3**). Participants were asked to listen attentively. (For Dothraki, a subset of the materials was discovered during the data analysis stage to not be proper Dothraki (**Supp. Table 2**). These materials affected 5 of the 16 blocks. We modeled those 5 blocks as a separate condition of no interest, so for Dothraki, the critical condition only consists of 11 blocks.)

Non-linguistic tasks

We included two non-linguistic tasks—a spatial working memory task and a naturalistic Theory of Mind task—which were used as additional controls to which neural responses to conlangs could be compared. The spatial working memory task (Fedorenko et al., 2013) also served as a localizer for the Multiple Demand network (Duncan, 2010, 2013)—the network that has been found to support comprehension of programming languages (e.g., Ivanova et al., 2020) and that served as the main alternative hypothesis for which brain system conlang comprehension would recruit.

Spatial Working Memory task. Participants had to keep track of four (easy condition) or eight (hard condition) locations in a 3 × 4 grid. In both conditions, participants performed a two-alternative forced-choice task at the end of each trial to indicate the set of locations that they just saw. Each trial lasted 8 s (see Fedorenko et al., 2011 for the trial timing details). Each block consisted of four trials and lasted 32 s. Each run consisted of 12 experimental blocks (six per condition) and four fixation blocks (16 s in duration each), for a total duration of 448 s (7 min, 28 s). Each participant performed two runs. Condition order was counterbalanced across runs. The hard > easy contrast targets brain regions engaged in diverse demanding cognitive tasks (e.g., Fedorenko et al., 2013; Shashidhara et al., 2019; Assem et al., 2020). This localizer is available from the Fedorenko laboratory website (https://evlab.mit.edu/funcloc/).

Naturalistic Theory of Mind task. Participants watched a short silent animated film ("Partly Cloudy", Pixar Animation Studios; 5 min 36 s in duration). The film has been coded for four types of events: i) mental (events that elicit thoughts about a character's mental state; e.g., the

cloud character believing that his friend has left him; 4 events, 44 s total); ii) social (events that involve social engagement but no specific mental/emotional content; e.g., the cloud character and a stork playing together; 5 events, 28 s total); iii) physical pain (e.g., the stork being bitten by a crocodile; 7 events, 26 s total); and iv) control (events where no characters are involved: e.g., shots of the landscape; 3 events, 24 s total). The mental > physical contrast targets brain regions engaged in thinking about another person's thoughts and beliefs (Jacoby et al., 2016; Richardson & Saxe, 2020; Shain, Paunov, Chen et al., 2022).

Behavioral tasks

CEFR self-assessment grid. Participants were asked to rate their proficiency in the relevant conlang(s) using the CEFR self-assessment grid developed by the Council of Europe (2001). In particular, they were asked to rate their proficiency in each of five different modalities: listening, reading, spoken interaction, spoken production, and writing. For each modality, participants were presented with six statements that describe basic to advanced skills and had to select the sentence that best describes their ability in the relevant conlang. For example, for listening, basic skills are described by the following statement: "I can understand familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly."; and advanced skills are described by: "I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided I have some time to get familiar with the accent." These statements correspond to the set of six common reference levels (A1, A2, B1, B2, C1, C2) developed by the Council of Europe to describe proficiency in a language (see **Supp. Table 4** for details on each reference level). For each participant, the scores were averaged across the five modalities to derive a single score (range: 1 (corresponding to A1) to 6 (corresponding to C2)).

Open-ended written production task. Participants were asked to write a few sentences about three topics in the conlang of interest. For each participant, the topics were randomly selected from the following set: family, hobbies, animals, home, friends, food, weather, sports, and clothes. On each trial, the topic was presented at the top of the screen along with the instructions below it ("please write about the topic above"). Participants had to type their responses into a text box window and then click the "submit" button at the bottom of the screen once they finished typing. The responses were timed from the first click on the text box until the "submit" button was pressed. For each participant, we calculated the average number of words per minute for each topic and averaged these values across the three topics to derive a single score for each participant.

Lexical decision task. Speakers of Esperanto and Klingon were presented with 40 strings of letters, one at a time, and were asked to decide as quickly as possible whether each string was a real word in the relevant conlang. Twenty of these strings were real words and 20 were

nonwords. To create the materials, we started with a set of 40 English words from the set of words used in the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1965) that exist in both Esperanto and Klingon. Twenty of these words were translated into Esperanto / Klingon and used as real words. The other twenty words were translated and then altered (by removing, adding, or transposing some of the letters) and cross-referenced with conlang dictionaries to ensure that the resulting strings did not constitute real words. The order of strings was randomized for each participant. On each trial, the letter string was presented in lower-case letters in the center of the screen, and participants were asked to press one of two buttons on the keyboard to indicate their response. For each participant, we calculated the accuracy score (number of correctly answered trials out of 40). Reaction times were recorded but not used in the analyses.

fMRI data acquisition. Structural and functional data were collected on a whole-body 3 Tesla Siemens Prisma scanner with a 32-channel head coil at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT. T1-weighted, Magnetization Prepared Rapid Gradient Echo (MP-RAGE) structural images were collected in 208 sagittal slices with 1 mm isotropic voxels (TR = 1,800 ms, TE = 2.37 ms, TI = 900 ms, flip = 8 degrees). Functional, blood oxygenation level-dependent (BOLD) data were acquired using an SMS EPI sequence with a 90° flip angle and using a slice acceleration factor of 3, with the following acquisition parameters: seventy-two 2 mm thick near-axial slices acquired in the interleaved order (with 10% distance factor), 2 mm × 2 mm in-plane resolution, FoV in the phase encoding (A >> P) direction 208 mm and matrix size 104 × 104, TR = 2,000 ms, TE = 30 ms, and partial Fourier of 7/8. The first 10 s of each run were excluded to allow for steady state magnetization.

fMRI data pre-processing and first-level modeling. fMRI data were analyzed using SPM12 (release 7487), CONN EvLab module (release 19b) and other custom MATLAB scripts. Each participant's functional and structural data were converted from DICOM to NIFTI format. All functional scans were co-registered and resampled using B-spline interpolation to the first scan of the first session. Potential outlier scans were identified from the resulting subject-motion estimates as well as from BOLD signal indicators using default thresholds in CONN pre-processing pipeline (5 s.d. above the mean in global BOLD signal change or framewise displacement values above 0.9 mm). Functional and structural data were independently normalized into a common space (the Montreal Neurological Institute (MNI) template, IXI549Space) using SPM12 unified segmentation and normalization procedure with a reference functional image computed as the mean functional data after realignment across all timepoints omitting outlier scans. The output data were resampled to a common bounding box between MNI-space coordinates (−90, −126 and −72) and (90, 90 and 108), using 2 mm isotropic voxels and fourth-order spline interpolation for the functional data and 1-mm isotropic voxels and tri-linear interpolation for the structural data. Lastly, the functional data were smoothed spatially using spatial convolution with a 4 mm full-width half-maximum (FWHM) Gaussian kernel. For the language localizer task and the non-linguistic tasks, effects were estimated using a general

linear model (GLM) in which each experimental condition was modeled with a boxcar function convolved with the canonical hemodynamic response function (HRF) (fixation was modeled implicitly). Temporal autocorrelations in the BOLD signal time series were accounted for by a combination of high-pass filtering with a 128 s cutoff and whitening using an AR(0.2) model (first-order autoregressive model linearized around the coefficient a = 0.2) to approximate the observed covariance of the functional data in the context of restricted maximum likelihood (ReML) estimation. In addition to main condition effects, other model parameters in the GLM design included first-order temporal derivatives for each condition (for modeling spatial variability in the HRF delays) as well as nuisance regressors to control for the effect of slow linear drifts, subject-motion parameters and potential outlier scans on the BOLD signal.

fROI definition and response estimation. For each participant, language fROIs were defined using the Group-constrained Subject-Specific (GSS) approach (Fedorenko et al., 2010) whereby a set of parcels or 'search spaces' (that is, brain areas within which most individuals in prior studies showed activity for the localizer contrast) is combined with each individual participant's activation map for the same contrast. (As noted above, in the main analysis, we used the reading-based language localizer, but see **Supp. Figure 1** for evidence that the results are similar when using the listening-based language localizer.)

We used five parcels derived from a group-level representation of data for the Sentences > Nonwords contrast in 220 independent participants. These parcels (used in much prior work; e.g., Malik-Moraleda, Ayyash et al., 2022) included three regions in the left frontal cortex—two in the inferior frontal gyrus (LIFG and LIFGorb) and one in the middle frontal gyrus (LMFG)—and two regions in the left lateral temporal cortex (LAntTemp and LPostTemp; **Figure 2**). Individual fROIs were defined by selecting within each parcel the 10% of most localizer-responsive voxels based on the *t*-values for the relevant contrast (Sentences > Control). In our set of participants, 2 participants showed right-hemisphere dominance for language based on the English reading localizer. For these participants, we used the RH parcels to define the language fROIs, but we also report a version of the analysis where the LH parcels are used for all participants (**Supp. Figure 3**). We then extracted the responses from these fROIs (averaging the responses across the voxels in each fROI) to each condition of the language localizer (using an across-runs cross-validation procedure to ensure independence between the data used to define the fROIs vs. to examine the response magnitudes), to the Sentences and Control conditions of the auditory language localizer and of the critical conlang experiment, and to the non-linguistic tasks. Statistical tests were then performed on the percent BOLD signal change values extracted from the fROIs.

Statistical analysis. For each of the tasks (and each conlang separately in some analyses), we fitted a model that predicted the BOLD responses (to the relevant contrasts, as elaborated below) of the LH language fROIs, with participants and fROIs included as random intercepts:

$$EffectSize \sim Condition + (1|Participant) + (1|ROI)$$

Note that for the main analyses, we consider the five language regions as an integrated system given that these regions a) have similar functional profiles with respect to their selectivity for language (e.g., Fedorenko et al., 2011; Fedorenko & Blank, 2020) and their role in lexico-semantic and combinatorial processing during language comprehension and production (e.g., Fedorenko et al., 2010, 2016, 2020; Bautista & Wilson, 2016; Hu, Small et al., 2022) and b) exhibit strong inter-region correlations in their activity during naturalistic cognition paradigms (e.g., Blank et al., 2014; Paunov et al., 2019; Braga et al., 2020; Malik-Moraleda, Ayyash et al., 2022).

To assess the selectivity of the language regions' responses to language relative to control conditions, we examined the Sentences vs. Control contrast for the reading-based localizer (estimated using the across-runs cross-validation to ensure independence between the data used for fROI definition vs. response estimation), the listening-based localizer, and the critical conlang task both across all participants and within each conlang group. To assess the specificity of the language regions' responses to conlangs relative to non-linguistic tasks, we examined the contrasts between the Sentences condition of the conlang task and each of the following conditions: the hard and easy conditions of the spatial working memory task and the mental and physical conditions from the naturalistic Theory of Mind task.

References

Amalric, M., & Dehaene, S. (2019). A distinct cortical network for mathematical knowledge in the human brain. *NeuroImage*, *189*, 19–31. https://doi.org/10.1016/J.NEUROIMAGE.2019.01.001

Apperly, I. A., Samson, D., Carroll, N., Hussain, S., & Humphreys, G. (2007). Intact first- and second-order false belief reasoning in a patient with severely impaired grammar. *Social Neuroscience*, *1*(3–4), 334–348. https://doi.org/10.1080/17470910601038693

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

Barham, L., & Everett, D. (2021). Semiotics and the origin of language in the Lower Palaeolithic. *Journal of Archaeological Method and Theory*, *28*(2), 535–579.

Bautista, A., & Wilson, S. M. (2016). Neural responses to grammatically and lexically degraded speech. *Language, Cognition and Neuroscience*, *31*(4), 567–574. https://doi.org/10.1080/23273798.2015.1123281/SUPPL_FILE/PLCP_A_1123281_SM7002.WAV

Bers, M. U. (2019). Coding as another language: A pedagogical approach for teaching computer science in early childhood. *Journal of Computers in Education*, *6*(4), 499–528.

Bickerton, D. (1976). Pidgin and creole studies. *Annual Review of Anthropology*, *5*(1), 169–193.

Blank, I. A., Kanwisher, N., & Fedorenko, E. (2014). A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *Journal of Neurophysiology*, *112*(5), 1105–1118. https://doi.org/10.1152/jn.00884.2013

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

Brentari, D., & Coppola, M. (2013). What sign language creation teaches us about language. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(2), 201–211.

Call, J., & Tomasello, M. (2020). *The gestural communication of apes and monkeys*. Psychology press.

Chee, M. W. L., Tan, E. W. L., & Thiel, T. (1999). Mandarin and English single word processing studied with functional magnetic resonance imaging. *Journal of Neuroscience*, *19*(8), 3050–3056. https://doi.org/10.1523/jneurosci.19-08-03050.1999

Chee, M. W. L., Weekes, B., Lee, K. M., Soon, C. S., Schreiber, A., Hoon, J. J., & Chee, M. (2000). Overlap and Dissociation of Semantic Processing of Chinese Characters, English Words, and Pictures: Evidence from fMRI. *NeuroImage*, *12*(4), 392–403. https://doi.org/10.1006/NIMG.2000.0631

Chen, X., Affourtit, J., Ryskin, R., Regev, T. I., Norman-Haignere, S., Jouravlev, O., Malik-Moraleda, S., Kean, H., Varley, R., & Fedorenko, E. (2023). The human language system, including its inferior frontal component in "Broca's area," does not support music perception. *Cerebral Cortex*, bhad087.

Cohn, N. (2013). Visual narrative structure. *Cognitive Science*, *37*(3), 413–452.

Council of Europe. (2001). *Common European Framework of Reference*.

Dediu, D., & Levinson, S. C. (2013). On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. *Frontiers in Psychology*, *4*, 397.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. In *Trends in Cognitive Sciences* (Vol. 14, Issue 4, pp. 172–179). Elsevier Current Trends. https://doi.org/10.1016/j.tics.2010.01.004

Duncan, J. (2013). The Structure of Cognition: Attentional Episodes in Mind and Brain. *Neuron*, *80*(1), 35–50. https://doi.org/10.1016/J.NEURON.2013.09.015

Duncan, J., Assem, M., & Shashidhara, S. (2020). Integrated intelligence from distributed brain activity. *Trends in Cognitive Sciences*, *24*(10), 838–852.

Dunn, L. M., & Dunn, L. M. (1965). Peabody Picture Vocabulary Test. In *APA PsycTests*.

Ergin, R., Raviv, L., Senghas, A., Padden, C., & Sandler, W. (2020). Community structure affects convergence on uniform word orders: evidence from emerging sign languages. *Proceedings of 13th International Conference on the Evolution of Language*.

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

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

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

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

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

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

Fedorenko, E., Ivanova, A., Dhamala, R., & Bers, M. U. (2019). The language of programming: a cognitive perspective. *Trends in Cognitive Sciences*, *23*(7), 525–528.

Fedorenko, E., Piantadosi, S. T., & Gibson, E. (n.d.). Language is a tool for communication, not for complex thought. *Nature Reviews Neuroscience*.

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

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

Fitch, W. T. (2017). Empirical approaches to the study of language evolution. *Psychonomic Bulletin and Review*, *24*(1), 3–33. https://doi.org/10.3758/S13423-017-1236-5

Gibson, E., Futrell, R., Piandadosi, S. T., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How Efficiency Shapes Human Language. In *Trends in Cognitive Sciences* (Vol. 23, Issue 5, pp. 389–407). Elsevier Ltd. https://doi.org/10.1016/j.tics.2019.02.003

Goldin-Meadow, S. (1993). When does gesture become language? A study of gesture used as a primary communication system by deaf children of hearing parents. In *Tools, language and cognition in human evolution* (pp. 63–85).

Goodall, G. (2022). Constructed Languages. *Annual Review of Linguistics*, *9*(1). https://doi.org/10.1146/ANNUREV-LINGUISTICS-030421-064707

Hagendoorn, I. (2010). Dance, language and the brain. *International Journal of Arts and Technology*, *3*(2–3), 221–234. https://doi.org/10.1504/IJART.2010.032565

Hockett Charles F. (1959). Animal" languages" and human language. *Human Biology*, *31*(1), 32–39.

Hu, J., Small, H., Kean, H., Takahashi, A., Zekelman, L., Kleinman, D., Ryan, E., Nieto-Castañón, A., Ferreira, V., & Fedorenko, E. (2022). Precision fMRI reveals that the language-selective network supports both phrase-structure building and lexical access during language production. *Cerebral Cortex (New York, N.Y.□: 1991)*. https://doi.org/10.1093/CERCOR/BHAC350

Illes, J., Francis, W. S., Desmond, J. E., Gabrieli, J. D. E., Glover, G. H., Poldrack, R., Lee, C. J., & Wagner, A. D. (1999). Convergent cortical representation of semantic processing in bilinguals. *Brain and Language*, *70*(3), 347–363. https://doi.org/10.1006/BRLN.1999.2186

Ivanova, A. A., Mineroff, Z., Zimmerer, V., Kanwisher, N., Varley, R., & Fedorenko, E. (2021). The Language Network Is Recruited but Not Required for Nonverbal Event Semantics. *Neurobiology of Language*, *2*(2), 176–201. https://doi.org/10.1162/NOL_A_00030

Ivanova, A. A., Srikant, S., Sueoka, Y., Kean, H. H., Dhamala, R., O'Reilly, U. M., Bers, M. U., & Fedorenko, E. (2020). Comprehension of computer code relies primarily on domain-general executive brain regions. *ELife*, *9*, 1–24. https://doi.org/10.7554/ELIFE.58906

Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences*, *3*(7), 272–279. https://doi.org/10.1016/S1364-6613(99)01333-9

Jackendoff, R. (2010). Your theory of language evolution depends on your theory of language. *The Evolution of Human Language: Biolinguistic Perspectives*, 63–72. https://doi.org/10.1017/CBO9780511817755.004

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

Joshua, F. (2012). Utopian for beginners. *The New Yorker*, *88*, 41.

Kirby, S. (2017). Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin & Review*, *24*, 118–137.

Kirby, Simon, Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(31), 10681–10686. https://doi.org/10.1073/PNAS.0707835105/SUPPL_FILE/0707835105SI.PDF

Kirby, Simon, Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. , 28, 108-114. *Current Opinion in Neurobiology*, *28*, 108–113.

Kirby, Simon, Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87-102.

Klein, D., Zatorre, R. J., Chen, J. K., Milner, B., Crane, J., Belin, P., & Bouffard, M. (2006). Bilingual brain organization: A functional magnetic resonance adaptation study. *NeuroImage*, *31*(1), 366–375. https://doi.org/10.1016/J.NEUROIMAGE.2005.12.012

Krifka, M. (2001). Quantifying into question acts. *Natural Language Semantics*, *9*(1), 1–40.

Levshina, N. (2022). *Communicative Efficiency*. Cambridge University Press.

Lipkin, B., Tuckute, G., Affourtit, J., Small, H., Mineroff, Z., Kean, H., Jouravlev, O., Rakocevic, L., Pritchett, B., Siegelman, M., Hoeflin, C., Pongos, A., Blank, I. A., Kline Struhl, M., Ivanova, A., Shannon, S., Sathe, A., Hoffmann, M., Nieto-Castañón, A., … Zhan, M. E. (2022). LanA (Language Atlas): A probabilistic atlas for the language network based on fMRI data from >800 individuals. *Nature Scientific Data*, *9*(1). https://doi.org/10.1101/2022.03.06.483177

Liu, Y. F., Kim, J., Wilson, C., & Bedny, M. (2020). Computer code comprehension shares neural resources with formal logical inference in the fronto-parietal network. *ELife*, *9*, 1–22. https://doi.org/10.7554/ELIFE.59340

Lupyan, G., & Dale, R. (2010). Language Structure Is Partly Determined by Social Structure. *PLOS ONE*, *5*(1), e8559. https://doi.org/10.1371/JOURNAL.PONE.0008559

Malik-Moraleda, S., Ayyash, D., Gallée, J., Affourtit, J., Hoffmann, M., Mineroff, Z., Jouravlev, O., & Fedorenko, E. (2022). An investigation across 45 languages and 12 language families reveals a universal language network. *Nature Neuroscience*, *25*(8), 1014–1019. https://doi.org/10.1038/s41593-022-01114-5

Malik-Moraleda, S., Jouravlev, O., Mineroff, Z., Cucu, T., Taliaferro, M., Mahowald, K., Blank, I. A., & Fedorenko, E. (2023). Functional characterization of the language network of polyglots and hyperpolyglots with precision fMRI. *BioRxiv*.

McGee, V. (1997). How we learn mathematical language. *The Philosophical Review*, *106*(1),

35–68.
Menenti, L., Petersson, K. M., & Hagoort, P. (2010). From reference to sense: An fMRI adaptation study on semantic encoding in language production. *HBM 2010-16th Annual Meeting of the Organization for Human Brain Mapping*.
Mineroff, Z., Blank, I. A., Mahowald, K., & Fedorenko, E. (2018). A robust dissociation among the language, multiple demand, and default mode networks: Evidence from inter-region correlations in effect size. *Neuropsychologia*, *119*, 501–511. https://doi.org/10.1016/j.neuropsychologia.2018.09.011
Monti, M. M., Parsons, L. M., & Osherson, D. N. (2009). The boundaries of language and thought in deductive inference. *Proceedings of the National Academy of Sciences*, *106*(30), 12554–12559.
Monti, M. M., Parsons, L. M., & Osherson, D. N. (2012). Thought Beyond Language: Neural Dissociation of Algebra and Natural Language. *Psychological Science*, *23*(8), 914–922. https://doi.org/10.1177/0956797612437427/ASSET/IMAGES/LARGE/10.1177_0956797612437427-FIG2.JPEG
Muysken, P., & Smith, N. (1995). The study of pidgin and creole languages. In *Pidgins and creoles: An introduction* (pp. 3–14).
Nieto-Castañón, A., & Fedorenko, E. (2012). Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *NeuroImage*, *63*(3), 1646–1669. https://doi.org/10.1016/J.NEUROIMAGE.2012.06.065
Okrent, A. (2010). *In the land of invented languages: Adventures in linguistic creativity, madness, and genius*. Random House.
Patel, A. (2010). *Music, language, and the brain*. Oxford University Press.
Paunov, A. M., Blank, I. A., & Fedorenko, E. (2019). Functionally distinct language and Theory of Mind networks are synchronized at rest and during language comprehension. *Journal of Neurophysiology*, *121*(4), 1244–1265.
Perreault, C., & Mathew, S. (2012). Dating the Origin of Language Using Phonemic Diversity. *PLoS ONE*, *7*(4), 35289. https://doi.org/10.1371/JOURNAL.PONE.0035289
Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291.
Premack, D. (1985). “Gavagai!” or the future history of the animal language controversy. *Cognition*, *19*(3), 207–296. https://doi.org/10.1016/0010-0277(85)90036-8
Quijada, J. (2011). *A Grammar of the Ithkuil Language*.
Richardson, H., & Saxe, R. (2020). Development of predictive responses in theory of mind brain regions. *Developmental Science*, *23*(1). https://doi.org/10.1111/DESC.12863
Rombouts, S. A. R. B., Barkhof, F., Hoogenraad, F. G. C., Sprenger, M., Valk, J., & Scheltens, P. (1997). Test-retest analysis with functional MR of the activated area in the human visual cortex. *Am Soc Neuroradiology*, *18*, 195–6108. http://www.ajnr.org/content/18/7/1317.short
Roux, F. E., & Trétmoulet, M. (2002). Organization of language areas in bilingual patients: a cortical stimulation study. *Journal of Neurosurgery*, *97*(4), 857–864. https://doi.org/10.3171/JNS.2002.97.4.0857
Saxe, R., Moran, J. M., Scholz, J., & Gabrieli, J. (2006). Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Social Cognitive and Affective Neuroscience*, *1*(3), 229–234. https://doi.org/10.1093/scan/nsl034
Schor, E. (2016). *(2016). Bridge of words: Esperanto and the dream of a universal language. Macmillan.* Macmillan.

Senghas, R., Senghas, A., & JE Pyers. (2014). The emergence of Nicaraguan Sign Language: Questions of development, acquisition, and evolution. In *Biology and knowledge revisited* (pp. 305–324). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781410611970-16/emergence-nicaraguan-sign-language-questions-development-acquisition-evolution-richard-senghas-ann-senghas-jennie-pyers

Shain, C., Paunov, A., Chen, X., Lipkin, B., & Fedorenko, E. (2022). No evidence of theory of mind reasoning in the human language network. *Cerebral Cortex*. https://doi.org/10.1093/CERCOR/BHAC505

Shashidhara, S., Spronkers, F. S., & Erez, Y. (2019). Individual-subject functional localization increases univariate activation but not multivariate pattern discriminability in the "multiple-demand" frontoparietal network. *Journal of Cognitive Neuroscience*, *32*(7), 1348–1368. https://doi.org/10.1162/jocn_a_01554

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial LIfe*, *9*(4), 371–386.

Srikant, S., Lipkin, B., Ivanova, A., Fedorenko, E., & O'Reilly, U. M. (2022). Convergent Representations of Computer Programs in Human and Artificial Neural Networks. *Advances in Neural Information Processing Systems*, *35*, 18834–18849.

Sulpizio, S., Del Maschio, N., Fedeli, D., & Abutalebi, J. (2020). Bilingual language processing: A meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *108*, 834–853. https://doi.org/10.1016/J.NEUBIOREV.2019.12.014

Swadesh, M. (2017). The origin and diversification of language. *The Origin and Diversification of Language*, 1–350. https://doi.org/10.4324/9781315133621/ORIGIN-DIVERSIFICATION-LANGUAGE-MORRIS-SWADESH

Varley, R. A., Klessinger, N. J. C., Romanowski, C. A. J., & Siegal, M. (2005). Agrammatic but numerate. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(9), 3519–3524. https://doi.org/10.1073/PNAS.0407470102/SUPPL_FILE/07470FIG4.JPG

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

Wikipedia. (2023). *Ithkuil*. https://en.wikipedia.org/wiki/Ithkuil

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, *21*(4), 463–489.

Acknowledgements

We would like to thank Bozena Pajak, Jessica Becker, and Joseph Rollinson from Duolingo for providing the initial recordings for Esperanto, Klingon, and High Valyrian; Matthew Michael Early and Alan Anderson for re-recording the Esperanto and Klingon materials, respectively; So Hee Ahn and Rucha Kelkar for help with stimulus creation; and all members of the Fedorenko and Gibson labs for helping during data collection (especially, Colton Casto, Lia Washington, Aalok Sathe, Anya Ivanova, Hee So Kim, Eric Martinez, Sammy Floyd, Chengxu Zhuang, Moshe Poliak, Greta Tuckute, Hope Kean, Carina Kauf and Elizabeth Lee). We would also like to thank Damián Blasi and Arika Okrent for their presentations (along with the four conlang creators, who are authors on the manuscript) at the 'Brains on Conlangs' event at MIT (https://mcgovern.mit.edu/2022/12/12/brains-on-conlangs/), and Thomas Clark for helping host the event. Finally, we are grateful to Elise Malvicini, Julie Pryor, and Gayle Lutchen for their help in planning and organizing the event and to the McGovern Institute for Brain Research for sponsorship. For comments on an earlier draft, we thank Rebecca Saxe and Steve Piantadosi. This work was supported by research funds to EF from the McGovern Institute for Brain Research, the Brain and Cognitive Sciences Department, the Simons Center for the Social Brain, and the Middleton Professorship. EF was additionally supported by NIH awards R01-DC016607, R01-DC016950, and U01-NS121471.