

*Research*

# An approach to separating the levels of hierarchical structure building in language and mathematics

Michiru Makuuchi<sup>1</sup>, Jörg Bahlmann<sup>2</sup> and Angela D. Friederici<sup>1,\*</sup>

<sup>1</sup>*Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*

<sup>2</sup>*Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA*

We aimed to dissociate two levels of hierarchical structure building in language and mathematics, namely ‘first-level’ (the build-up of hierarchical structure with externally given elements) and ‘second-level’ (the build-up of hierarchical structure with internally represented elements produced by first-level processes). Using functional magnetic resonance imaging, we investigated these processes in three domains: sentence comprehension, arithmetic calculation (using Reverse Polish notation, which gives two operands followed by an operator) and a working memory control task. All tasks required the build-up of hierarchical structures at the first- and second-level, resulting in a similar computational hierarchy across language and mathematics, as well as in a working memory control task. Using a novel method that estimates the difference in the integration cost for conditions of different trial durations, we found an anterior-to-posterior functional organization in the prefrontal cortex, according to the level of hierarchy. Common to all domains, the ventral pre-motor cortex (PMv) supports first-level hierarchy building, while the dorsal pars opercularis (POd) subserves second-level hierarchy building, with lower activation for language compared with the other two tasks. These results suggest that the POd and the PMv support domain-general mechanisms for hierarchical structure building, with the POd being uniquely efficient for language.

**Keywords:** arithmetic; complex cognition; functional magnetic resonance imaging; sentence processing

## 1. INTRODUCTION

The ability to compute symbol sequences in language and mathematics represents the powerful capability humans have for information processing. In language, sentences are produced as a linear sequence of words, but the words are organized to form syntactic chunks (i.e. phrases), which can be further grouped into larger chunks (i.e. clauses and sentences). For example, a simple sentence such as ‘The young man bought a big house’ can be grouped into [the young man], [bought] and [a big house], and further into [The young man] and [[bought][a big house]] [1–3]. It is assumed that repeated chunking during sentence processing leads to a build-up of hierarchically structured mental representations that allow the assignment of grammatical relations, and thus mapping from the word sequence to the meaning of the sentence. This linguistic hierarchical structure building process is known under different names, such as ‘Unification’ [4] or ‘Merge’ [5].

‘Merge’, for example, is a computation to build a minimal hierarchy. Multiple application of Merge

builds up the hierarchical structure of an entire sentence. The resulting hierarchical structures differ between sentence types. Sentences with the object noun phrase in sentence initial position (non-canonical structure) have more complex hierarchical structures than sentences with the subject noun phrase in sentence initial position (canonical structure). In an extremely complex case, the sentence structure can have multiple centre-embeddings such as ‘The rat the cat the dog chases caught ate the cheese’. Neuroimaging studies have revealed that the processing of minimal hierarchies in artificial grammar sequences [6] and minimal phrases [7] activates the frontal operculum in adults, but involves the lateral prefrontal cortex (PFC) when language processing is not yet fully automatic, as in second language learners [8] and children [9]. For the processing of complex hierarchical structures, the left pars opercularis (PO), which is part of the posterior part of the inferior frontal gyrus, is selectively recruited, for both object-first non-canonical structures [10–12] and embedded structures [6,10,11,13,14]. Studies have provided insights into the neural mechanism underlying the processing of hierarchical structures, by comparing sentences with different structural complexities.

Mathematics, another human complex cognition ability, also involves hierarchical structures such as  $(1 + 3) \times (4 - 2)$ . Although the neural bases of several

\* Author for correspondence (angelafr@cbs.mpg.de).

One contribution of 13 to a Theme Issue ‘Pattern perception and computational complexity’.

aspects of mathematical cognition have been uncovered [15–17] and compared with that of language [18], we still have little knowledge about the process of hierarchical structure building in mathematics. There is only one study that has investigated the processing of abstract formulae with hierarchical structures and compared these with formulae having a flat structure. In that study, the regions anterior to the PO in the left hemisphere were found to be activated [19]. The more anterior activation for the processing of hierarchical structures in mathematics compared with language could be either a reflection of a domain difference, or a reflection of the difference of automaticity of processes owing to a difference in the everyday use of those two domains (for a discussion, see [20]).

The goal of the present functional magnetic resonance imaging (fMRI) study was to investigate the neural mechanisms of hierarchical structure building in language and mathematics from linear input sequence, in a step-by-step manner. Specifically, we focused on whether the build-up of hierarchical structures at different levels in a hierarchy is supported by different brain regions in the posterior lateral PFC.

Our hypothesis concerning the different levels of hierarchy and their possible computational substrate was inspired by the view that processes at a higher level in the hierarchy recruit more anterior brain regions in the PFC than the processes at a lower level in the hierarchy [21–24]. This view was initially formulated to model executive functions as a domain-general hierarchical cognitive control process. According to this view, a simple button press task to external stimuli is executed by a simple stimulus–response mapping in the premotor cortex (‘sensory control’). When the stimulus–response mapping rule is dependent on immediate context, a more anterior region (the posterior lateral PFC) exerts the action selection in the premotor cortex (‘contextual control’). If there is a further factor that changes the mapping rule over time, in addition to the immediate context, then a yet more anterior region (the anterior lateral PFC) sends a control signal to the posterior lateral PFC to select the appropriate action (‘episodic control’). Finally, when multiple episodes switch serially, the highest level of cognitive control (‘branching’) is exerted to the anterior lateral PFC by the polar lateral PFC.

Here, we apply the hypothesis of a posterior to anterior gradient in the PFC that was formulated for executive functions [21–24] to the build-up of different hierarchical levels of abstract hierarchical structures. The present study was designed to test whether the build-up of a lower level hierarchy is computed in more posterior regions in the PFC than the build-up of a higher-level hierarchy, both in language and in mathematics.

In order to keep the hierarchical structure and the processing steps at lower and higher hierarchical levels as comparable as possible across language and mathematical sequences, we used Japanese language and presented mental arithmetic in the Reverse Polish notation. In this notation, binary arithmetic operators are given after the two natural numbers (figure 1*d*). This sequence is comparable to a simple

Japanese sentence in which two noun phrases (subject and object noun phrase) are given prior to the sentence-final verb (figure 1, note that the sentence structure displayed in figure 1*c* does not follow the convention of linguistics, but is meant to indicate the assumed operational similarity with the arithmetic domain in figure 1*d*). Japanese and other free word order languages (e.g. Korean) allow this structure, unlike languages with the verb in the second position, such as English (where it comes after the subject noun phrase). As a control condition, we added a verbal working memory task in which a serial string of letters was organized to have the identical hierarchical structure (figure 1*e*).

Crucially, we discriminated between the neural computations for two levels in the build-up of a hierarchical structure. The first-level hierarchy (HL1) refers to the build-up of a hierarchical structure with two elements and an operator given externally. The second-level hierarchy (HL2) means the build-up of a hierarchical structure using two internal representations resulting from the previous first-level processing (HL1) and an operator given externally. We differentiate the two levels in the hierarchical structure building, because we assume the ability to build the hierarchical structure with internal representation could be fundamental to the faculty of language.

## 2. METHODS

### (a) *Participants*

A total of 18 right-handed, healthy subjects took part in the study (11 females and 7 males). Handedness was assessed with the Edinburgh Handedness Inventory [25] ( $M = 92$ , range = 67–100). The mean age was 30.3 years, and the range was 21–40. All participants were native Japanese speakers living in Leipzig, Germany. All had no history of neurological disorders. The experimental procedures were approved by the Research Ethics Committees of the University of Leipzig. Written informed consent was given by all subjects.

### (b) *Stimuli and tasks*

#### (i) *Stimuli*

*LIST trials* (encoding of two items used as baseline control). The subjects encoded two items presented sequentially. In the language task (LANG), the items were plausible combinations of an adjective and a noun without a case-marking particle (e.g. *siroi ie* [white house]). In the arithmetic task (ARITH), two digits were given (from 1 to 9). In the string operation task (STRING), two capital letters were selected from A, B, C and D without repetition of letters (table 1).

#### *HL1 trials* (building of the first-level hierarchy).

In the HL1 trials, subjects had to build a hierarchy with two serially given items and the operator attached to the second item (table 1). In LANG, the operators were one of four case-marking particles (*ga*, *ni*, *wa*, *wo*), which were attached to the right of the noun. For example, a noun phrase ‘*wakai otoko-ga*’ (a young man [nominative case]) was presented. These particles signal a phrase closure and assign a thematic role, and subjects had to remember the exact combination of

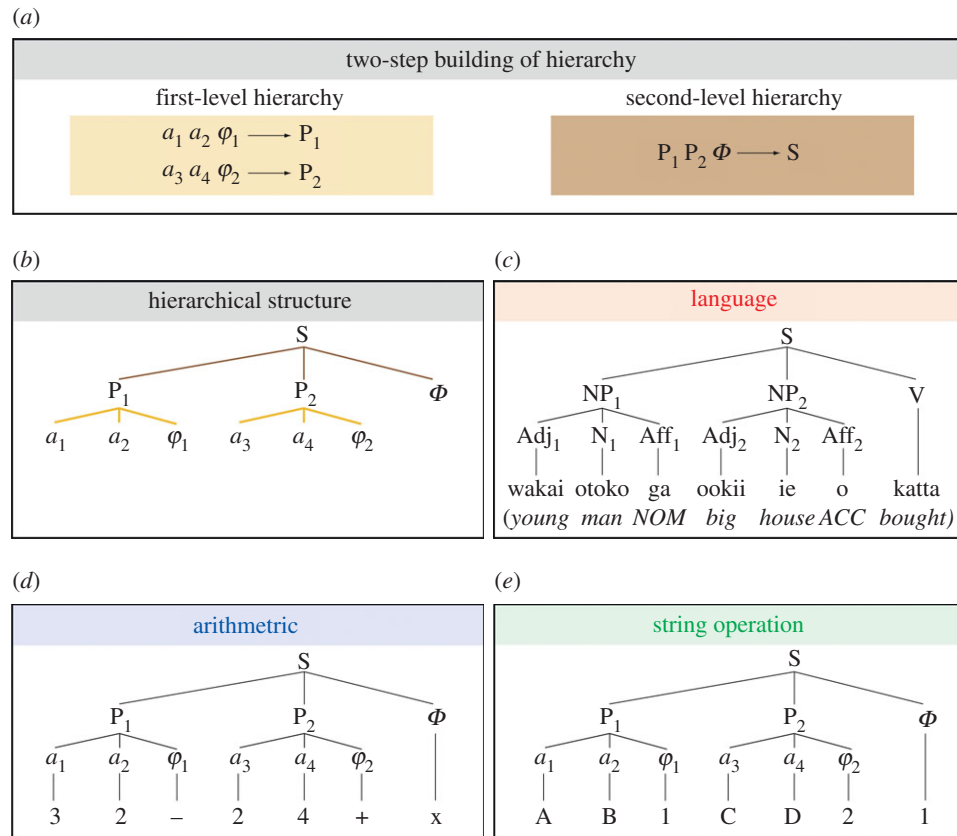


Figure 1. Two-level build-up of a hierarchical structure. (a) Two-step hierarchical structure building.  $a_1$ – $a_4$  are the elements given externally,  $\varphi_1$  and  $\varphi_2$  are the first-level operators, and  $\Phi$  is the second-level operator (table 1). The operations by  $\varphi$  and  $\Phi$  take two elements and build hierarchical structures with nodes P and S. Note that nodes such as NP, P and S are not externally given stimuli, but are the mental representations of the previously built hierarchical structures. (b) The resulting hierarchical structure from two-step hierarchical structure building. (c–e) Hierarchical structures of the HL2 sequences for language (LANG), arithmetic (ARITH) and string operation (STRING). Note that the structural description of the language example does not conform with standard linguistic descriptions in which NP<sub>2</sub> and V are subnodes of a verb phrase (VP). The present description was chosen to show the structural similarity across domains. NP, noun phrase; Adj, adjective; N, noun; Aff, affix (i.e. case marking particle); V, verb.

adjective, noun and particle. In ARITH, one of the four binary arithmetic operators (+, −, ×, /), which allow calculation, was given. We employed the reverse Polish notation (named after the Polish logician Jan Łukasiewicz), which is an expression for mathematical formulae without using parentheses. In this notation, the operator is given after the two elements. In this study, for example, when a sequence 1 2+ was given, subjects had to calculate 1 + 2 and remember the answer 3. In STRING, 1 or 2 was the operator that specified which item(s) had to be remembered. For example, for sequence A B1, A should be remembered, and for D C2, C had to be remembered (table 1).

#### HL2 trials (building the second-level hierarchy).

In the HL2 trials, subjects were given two HL1 stimuli in a row, and they had to build a hierarchical structure with the two internally represented outputs of the previous HL1 processing steps using the second-level operator presented in the final frame (figure 1 and table 1). In LANG, the sentence-final verb closed a sentence. Note that verbs always occupy the sentence-final position in Japanese sentences. Complete Japanese sentences were given, and the subjects were instructed to understand them. The verbs were in past tense and the voice of the verbs (i.e. active/passive/

causative/causative–passive) was varied to provide different sentence types. Although differences in voice can affect the activation pattern of the pars triangularis and posterior superior temporal sulcus [26], we did not treat them separately. The verbs were either in active (verb ending -ta, 10 verbs), passive (-rareta, 10 verbs), causative (-saseta, 10 verbs) or causative–passive (-saserareta, 10 verbs) voice.

This resulted in the following sentence types:

- animate NP-ha animate NP-ni V-passive.  
(five sentences)
- inanimate NP-ga animate NP-ni V-passive.  
(five sentences)
- animate NP-ha animate NP-wo V-active.  
(five sentences)
- animate NP-ga inanimate NP-wo V-active.  
(five sentences)
- animate NP-ni animate NP-wa or NP-ga V-causative-passive.  
(five sentences)
- animate NP-ni inanimate NP-wo V-causative e-passive.  
(five sentences)
- animate NP-ha inanimate NP-ni V-causative.  
(five sentences)
- animate NP-ga animate NP-ni V-causative.  
(five sentences)

Table 1. Stimulus types and examples of HL2 stimuli. Three types of stimulus (LIST, HL1, and HL2) with identical hierarchical structures are defined for the three domains.  $a_1$ – $a_4$  are elements.  $\varphi_1$  and  $\varphi_2$  are the first-level operators and  $\Phi$  is the second-level operator.  $a_1$ – $a_4$  are {adjective, noun}, {1,2,3,4,5,6,7,8,9,10}, and {A,B,C,D}, and  $\varphi_1$  and  $\varphi_2$  are {ga, wa, ni, wo (case marking particles)}, {+,−,×,/}, and {1,2} and  $\Phi$  is {verbs}, {+,−,×,/}, and {1,2} in LANG, ARITH and STRING, respectively. Elements were presented every one second and the presentation duration was 600 ms. In the LIST stimuli, the subject memorizes the two items presented serially. In the HL1 stimuli, a first-level operator is added to the second item. The subject integrates the two elements using an operator  $\varphi$  and builds a hierarchical structure with a node P. In the HL2 stimuli, the subject integrates the results of the two previously given HL1 stimuli using a second-level operator  $\Phi$  that appears in the final presentation frame and builds a hierarchical structure with a node S. In each trial, the subject keeps the final computation product (i.e. P or S) until the probe. NOM: nominative case, ACC: accusative case.  $1\ 2+4\ 2-\times$  indicates  $(1+2) \times (4-2)$  without parenthesis (the Reverse Polish notation). AB1 DC2 2 requires operations of  $A\ B1 \rightarrow A$ ,  $D\ C2 \rightarrow C$ , and  $A\ C2 \rightarrow C$ . See also figure 1.

| trial type                         |                                 |                                |                            |                               |                              |                        |  |
|------------------------------------|---------------------------------|--------------------------------|----------------------------|-------------------------------|------------------------------|------------------------|--|
| onset time                         | 0 s                             | 1                              | 2                          | 3                             | 4                            |                        | 6.8–7.2 s                                      |
| LIST                               | $a_1$                           | $a_2$                          | blank<br>2.6–3 s           | probe (yes/no)                |                              |                        |  |
| first-level<br>hierarchy<br>(HL1)  | $a_1$                           | $a_2 + \varphi_1$              | blank<br>2.6–3 s           | probe (yes/no)                |                              |                        |  |
| second-level<br>hierarchy<br>(HL2) | $a_1$                           | $a_2 + \varphi_1$              | $a_3$                      | $a_4 + \varphi_2$             | $\Phi$                       | blank<br>2.8–<br>3.2 s | probe (yes/no)                                 |
| examples of HL2 for each task      |                                 |                                |                            |                               |                              |                        |  |
| LANG                               | Wakai<br>(young<br>[adjective]) | Otoko-ga<br>(manNOM)<br>[noun] | ookii (big)<br>[adjective] | Ie-wo<br>(houseACC)<br>[noun] | Katta.<br>(bought)<br>[verb] |                        | Otoko-ga katta.<br>(a man bought.)<br>(yes/no) |
| ARITH                              | 1                               | 2+                             | 4                          | 2−                            | ×                            |                        | =8? (yes/no)                                   |
| STRING                             | A                               | B1                             | D                          | C2                            | 2                            |                        | C? (yes/no)                                    |

The interpretation of the case-marking particles can be dependent on the voice of the verb. For example, in the ‘animate NP1-ga animate NP2-ni’ construction, if the verb takes the active voice, NP1 is interpreted as the agent and NP2 as the recipient, but if the verb is in the passive voice, it is the reverse (for details see [26]). Only five of 40 sentences (inanimate NP-ga animate NP-ni V-passive) allowed the resolution of thematic roles before the verb and allowed prediction of the voice of the forthcoming verb at the second NP. The two noun phrases referred to either humans (65 nouns) or inanimate objects (15 nouns).

In ARITH and STRING, the two HL1 stimuli produced two values, and together with the second-level operator, these two produced the final result. For example, the mathematical expression ‘ $(1+2) \times (4-2)$ ’ is unambiguously expressed as ‘ $1\ 2+4\ 2-\times$ ’ by the Reverse Polish notation and results in 6. In STRING, AB1 DC2 2 requires two operations of the building of the HL1. In A B1, the first letter must be held in memory ( $A\ B1 \rightarrow A$ ), and in D C2, the second letter must be held in memory ( $D\ C2 \rightarrow C$ ). Then, finally the HL2 is built as  $A\ C2 \rightarrow C$ .

### (c) Formal description of elements, operators and tasks

We denote the building of hierarchical structure as  $a_i a_j \psi \rightarrow N$ , where  $a_i, a_j \in \Sigma$  (elements or terminal symbols)  $\cup N$  (non-terminal symbols),  $\varphi_1, \varphi_2, \Phi \in \psi$  (operator) and  $P, S \in N$  (mental representations for

the nodes of the HL1 and HL2, or non-terminal symbols).

The HL1 is built as

$$a_1 a_2 \varphi_1 \rightarrow P_1$$

and

$$a_3 a_4 \varphi_2 \rightarrow P_2.$$

Taking the products of the previous operation, the HL2 is built as

$$P_1 P_2 \Phi \rightarrow S.$$

The whole process can be equivalently expressed as  $((a_1 a_2 \varphi_1)(a_3 a_4 \varphi_2)) \Phi \rightarrow S$ . This two-level operation creates a hierarchical structure as shown in figure 1.

The elements in each domain are

*Language:*  $\Sigma = \{\text{adjective, noun}\}$ ,  $\psi = \{\text{ga, wo, wa, ni (case marking particle), verb}\}$ ,  $N = \{\text{NP, S}\}$ , where NP is noun phrase and S is sentence.

*Arithmetic:*  $\Sigma = \{1,2,3,4,5,6,7,8,9\}$ ,  $\psi = \{+, -, \times, /\}$ ,  $N = \{1,2,3,4,5,6,7,8,9\}$

*String operation:*  $\Sigma = \{A,B,C,D\}$ ,  $\psi = \{1,2\}$ ,  $N = \{A,B,C,D\}$ .

We modelled the online sequence processing of the hierarchically structured sequences as a repeated application of hierarchical structure building  $a_1 a_2 \varphi \rightarrow P$ , where  $a$  is an element,  $\varphi$  is an operator and  $P$  is the node of the hierarchical structure (figure 1a,b). We draw a distinction between the two levels in the hierarchical structure building. HL1 refers to the hierarchy with the elements given externally as visual stimuli. As a



control for HL1, we introduced encoding-only trials (LIST trials) that required participants to keep two items sequentially in memory without any structure building. HL2 is the hierarchy with internally represented elements resulting from the previous first-level building (HL1) that is stored in memory. Crucially, the build-up of hierarchical structure with the internal elements generates the hierarchical structures (figure 1b) inherent to language (figure 1c) and also to mathematics (figure 1d).

#### (d) Probes

Probes were presented after each trial to check the results of the computation. The time from the onset of the probe was modelled as offline processing. In LANG, for LIST and HL1 stimuli, the probes were noun phrases that were either correct (identical to those presented in the trials) or incorrect. Incorrect probes were created by replacing an adjective, noun or particle with another item that did not occur in the stimulus sentence, but was selected from the same lexical category. For example, for the stimulus of 'wakai otoko-ga (young man-ga)', an incorrect probe can be 'wakai otoko-wa' (young man-wa). For the HL2 stimuli, the correct probes were sentences that restated a part of the given sentence. Incorrect probes were constructed in a similar way to correct probes, but they were changed to express a different meaning. In some cases, one lexical item was replaced by one that did not appear in the stimulus trial. For example, for the sentence 'A good wine was found by a veteran sommelier', the probe (incorrect) was 'A white wine was found'. In other cases, the agent and the recipient of the action were exchanged. In ARITH and STRING, the correct probes were the identical pairs of digits/letters including the order (LIST stimuli) or one digit/letter, which was the same as the result of the operation (HL1 and HL2 stimuli). Incorrect probes were different pairs of digits/letters (LIST stimuli) or one digit/letter (HL1 and HL2 stimuli) different from the result of the operation.

#### (e) Qualitative differences in the stimuli and the operations among the domains

Although the stimuli were given in the sequence of an identical hierarchical structure, there are qualitative differences in the mental processing among the domains. Firstly, non-terminal symbols, or nodes, are qualitatively different. The non-terminal symbols in LANG are mental representations of a phrase or a sentence. In ARITH and STRING, any nodes are numbers and letters that are identical to the terminal symbols. Secondly, a few initial elements in a sentence in LANG may allow us to predict the forthcoming words based on the meaning of words or the context the words imply. Predictions are also possible based on syntactic information. According to the theta-criterion that holds that no theta role can be given twice in a sentence one can predict that, if the first NP has the case marker 'ni', then the second NP cannot have the case marker 'ni', but has to have other makers. Moreover, a sequence (inanimate NP-ga animate NP-ni) allows only passive voice marker for

the verb. ARITH and STRING do not allow the participants to predict the forthcoming elements. Thirdly, the commutability of elements differs. Elements are commutable with the operators  $+$  and  $\times$  but not with  $-$  and  $/$  in the ARITH task. For example,  $2 + 3$  is equal to  $3 + 2$ , but  $2 - 3$  is not equal to  $3 - 2$ . In LANG, the order of the elements is not changeable at all. Combination of noun-adjective-particle in this order results in ungrammatical products. In STRING, elements are commutable. Fourthly, there is a qualitative difference in the probes that test the participants' performance. While we presented other sentences as a probe in LANG to test subjects' comprehension, we gave one digit or one letter as the probe in the other two domains. Finally, proficiency in processing differs. Language is acquired in the early period of life and used every day. Arithmetic is learned in childhood and is automatized, but the frequency of use is less than that of language. The STRING task requires acquisition of a newly introduced rule, and proficiency should be the lowest. Even though there are many domain-specific features in the stimuli and their processing, the three domains share a two-step building of the hierarchical structures as common cognitive processes (figure 1).

#### (f) Experimental procedure

An event-related design was used and the stimuli of the nine conditions (i.e. three tasks (LANG, ARITH and STRING)  $\times$  three trial types (LIST, HL1 and HL2)) were presented in pseudo-random order. In each trial, two (LIST and HL1) or five (HL2) items were visually presented one by one for 600 ms every 1000 ms. After a blank screen of 3000 or 3400 ms, the probe was presented for 2000 ms (table 1). The time between the end of the probe and the beginning of the next stimulus varied, with a mean of 2600 ms and range 400–4800 ms. The beginning of stimuli presentation was jittered (0 or 800 ms) against the scanning. Eighteen LIST, 18 HL1 and 40 HL2 trials per domain were given, resulting in a total of 228 trials tested within 46 min. The experiment was performed in one session. The stimuli presentation was programmed with PRESENTATION v. 10.3 software (Neurobehavioral Systems, Inc., Albany, CA, USA) on a Windows PC. Stimuli were projected through an LCD projector (PLC-XP50L, SANYO, Tokyo, Japan) onto the back of a screen. Subjects viewed the images on the screen above their heads through a mirror attached to the head-coil.

#### (g) Image acquisition

Functional MRI data were acquired with a whole-body 3 Tesla Bruker MedSpec S300 scanner (Bruker BioSpin MRI GmbH, Germany) with a gradient-echo EPI sequence. The brain was covered with 4.8 mm thick 20 axial images with 1.2 mm gaps ( $TR = 1.6$  s,  $TE = 30$  ms, Flip angle =  $90^\circ$ ,  $FOV = 19.2 \times 19.2$  cm<sup>2</sup>,  $64 \times 64$  matrix). The resulting voxel size was  $3 \times 3 \times 6$  mm<sup>3</sup>. The slices were aligned to the AC–PC plane and placed to cover the whole brain. The subjects had one session of fMRI scanning with 1728 volumes per session in about 46 min. Structural high-resolution

images of the participants were also collected on a different day with a three-dimensional MDEFT sequence (TR = 1300 ms, TE = 3.93 ms, TI = 650 ms, Flip angle = 10°, FOV = 25.6 × 24 cm<sup>2</sup>, 256 × 240 matrix, sagittal 128 slices, 1 mm thick, 2 NEX) using a whole-body 3 Tesla Magnetom TRIO (Siemens Medical Solution, Erlangen, Germany). A stabilization cushion was laid under and to the sides of the head to reduce head motion.

## (h) Analysis

### (i) Preprocessing of structural and functional magnetic resonance imaging data

The first five volumes of the fMRI session were discarded to eliminate magnetic saturation effects. Data analysis were carried out using SPM8 (available at <http://www.fil.ion.ucl.ac.uk/spm/>) on Linux PC workstations. Structural images were corrected for signal intensity bias due to the magnetic field inhomogeneity using the bias correction tool in SPM8. The EPI images were realigned to the first image and then the difference in the slice acquisition time was corrected. Subsequently, the EPI images were coregistered to the subjects' three-dimensional high-resolution structural images. Normalizing an individual structural image to the SPM8 T1 brain template was processed in two steps: (i) estimation of the normalization parameters and (ii) writing the normalized images with the parameters. This parameter transformed the structural images and all the EPI volumes into a common stereotaxic space to allow multi-subject analyses. The EPI images were resampled into 3 × 3 × 3 mm<sup>3</sup> voxels and the structural images into 1 × 1 × 1 mm<sup>3</sup> voxels with the seventh degree B-spline interpolation.

### (ii) Whole brain analysis

As a first step, individual analyses were performed. All the functional volumes were smoothed with a 6-mm full width at half-maximum isotropic Gaussian kernel prior to statistical calculation in order to improve the signal-to-noise ratio and to compensate for the anatomical variability among individual brains. Each subject's haemodynamic responses induced by the stimuli and probes were modelled with a box-car function with the duration of the stimulus presentation (1600 ms for LIST and HL1, 4600 ms for HL2), and for probes (2000 ms) and convolved with a haemodynamic function. The stimuli and probes were modelled as distinct events to separate online hierarchical processing from offline probe matching. The global mean intensity of each session was normalized to 100. Confounds by global signal changes were removed by applying a high pass filter with a cut-off cycle of 128 s. Signal increase relative to the baseline in each condition of each participant was estimated according to the general linear model. The resulting individual contrast images were submitted to the second-level (group) analysis, including a conjunction analysis (against the 'conjunction null' hypothesis [27]). Statistical inferences were drawn at  $p < 0.05$  (corrected) at cluster level: the statistical maps (SPM{T}) were thresholded at  $p < 0.001$  (not

corrected), and then thresholded by the cluster size ( $k = 50$ ).

### (iii) Volumes of interest analysis

To visualize the activation time course and to perform further detailed statistical analysis, we reconstructed the trial time courses (TTCs) of the activated regions in the PFC for each condition. These were the ventral premotor cortex (PMv) and the dorsal portion of the pars opercularis (POd). For the volumes of interest (VOI) of the PMv, the activated cluster revealed by the conjunction analysis of 'HL1 > LIST' was used (figure 3 and table 2), since this contrast reveals neural correlates for the HL1 building. Assuming that the posterior-to-anterior gradient hypothesis [22] holds for abstract hierarchy building we take the POd as a neural correlate for the HL2. We defined the POd VOI as the activation cluster of the conjunction analysis of 'HL2 > baseline' (figure 3 and table 2) that locates within a cytoarchitectonically defined PO (i.e. Brodmann's area 44 [28,29]). Since we cannot compare the brain activation for HL2 with that for HL1 directly (i.e. 'HL2 > HL1') due to the difference in the trial duration, we used the contrast 'HL2 > baseline'. The cytoarchitectonic map is a digitized three-dimensional population map (in the MNI space) of cytoarchitectonic areas created with 10 postmortem human brains in which borders of areas were determined by statistically significant changes in laminar density patterns of neuronal cell bodies. We truncated a population map of Brodmann's area 44 of the left hemisphere at 30 per cent and created a mask image. Mean time-series data were extracted from these VOIs using MARSBAR v. 0.41 (available at <http://marsbar.sourceforge.net>). The TTCs for each condition were modelled with 21 variables representing the BOLD signal at every 0.8 s from 0 to 16 s after the stimulus onset, and the simultaneous equation was solved against the time-series data of VOIs under the assumption of linear time invariant system. In this analysis, we did not assume any shapes for haemodynamic responses, and online and offline processing were not separately modelled, resulting in two peaks in the reconstructed TTC.

To evaluate the difference in the building cost according to the levels of hierarchy, we calculated the BOLD signal difference between HL2 and HL1, and HL1 and LIST (we call this difference in the building cost  $\Delta\text{COST}$ ). We obtained the time to reach peaks from the mean TTCs of HL2 and HL1 trials across the three domains, and then collected the peak signal for the sequence processing from the individual TTCs. The  $\Delta\text{COST}$ s are computed with a novel method that was applied to correct for the peak BOLD signal height difference due to the difference in the duration of the stimuli in LIST, HL1 and HL2. The correction was performed in a procedure as follows. First, we calculated the convolution of a canonical haemodynamic function and a rectangular function with different durations (figure 2a). Then, we plotted the peak BOLD signal height against stimulus duration (figure 2b). Using this plot, we derived the weights to transform the BOLD signals (%) measured

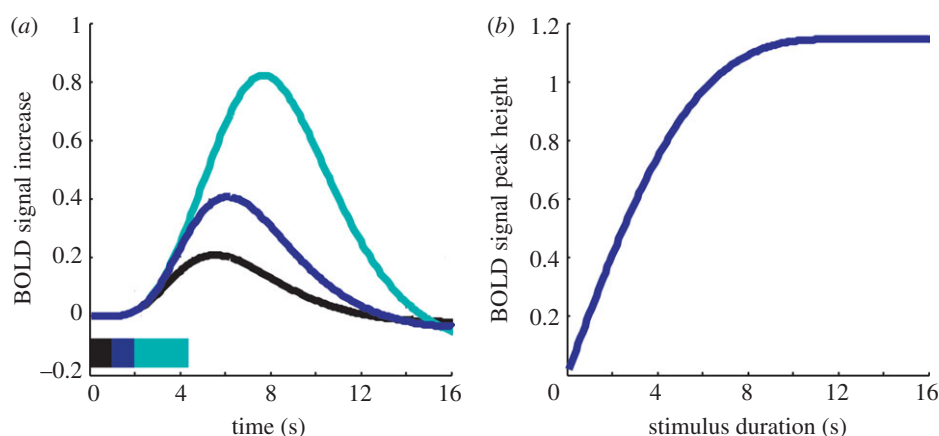


Figure 2. Modelled time course and stimulus duration. (a) BOLD response curves calculated by the convolution of the canonical haemodynamic function with rectangular functions with different duration (1 s, black; 2 s, blue; 4.6 s, cyan). The horizontal bars in black, blue and cyan at the left bottom represent the duration (1, 2 and 4.6 s) of the stimuli. (b) Plot of the peak height of BOLD response of different durations. The unit of the BOLD signal amplitude is arbitrary, but the same for both panels.

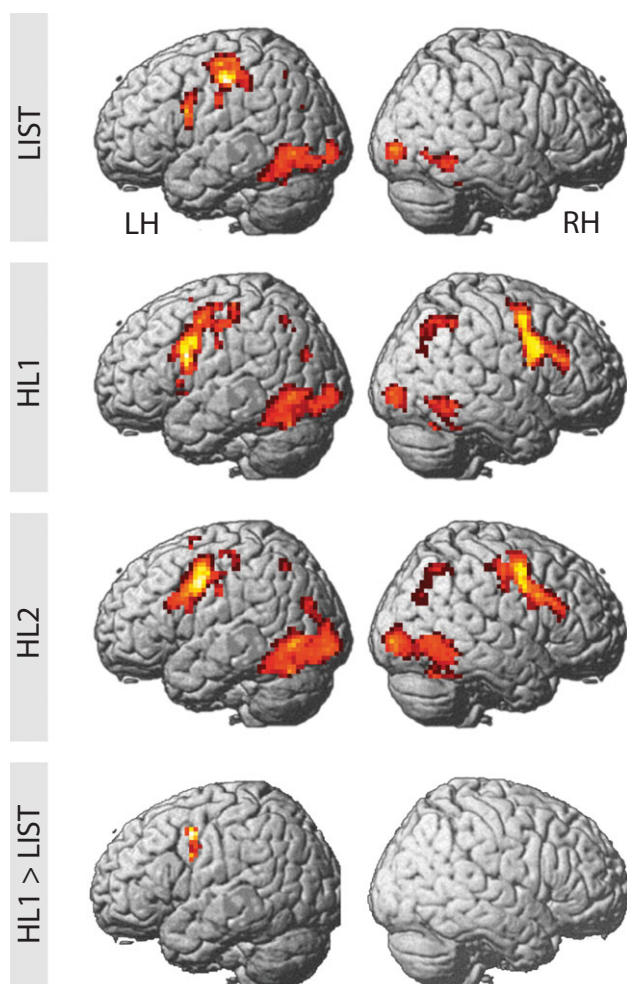


Figure 3. Conjunction analysis across the three conditions. Conjunction analysis across the three conditions for the LIST, HL1, HL2 and HL1 > LIST during online (i.e. stimuli) processing. Statistical threshold is  $p < 0.05$  (corrected). HL1 > LIST is marginally significant. See table 2 for the list of the activated foci.

in trials of different durations into the signal unit when the would have a duration of one second (corrected %). The weights were the ratios of the peak height

for the duration of the stimuli over that for the duration of 1 s. The peak BOLD signals sampled from the reconstructed TTC were divided by these weights to enable the comparison of BOLD signal increase between conditions of different durations. Although this correction is based on an assumption of linearity in the convolution of a canonical haemodynamic function and rectangular functions (figure 2), we believe that this method serves as a reasonable correction in this instance. For sequence processing,  $HL2/c_2 - HL1/c_1$  and  $HL1/c_1 - LIST/c_0$  were computed, where  $c_0, c_1, c_2 = 1.58, 1.58, 3.93$ . Note that these two  $\Delta COST$ s are qualitatively different.  $HL1/c_1 - LIST/c_0$  shows the BOLD signal difference between  $a_1 a_2 \varphi$  (HL1) and  $a_1 a_2$  (LIST), namely for  $\varphi$ , the building of HL1. On the other hand,  $HL2/c_2 - HL1/c_1$  indicates the corrected difference between  $P_1 P_2 \Phi$  and  $a_1 a_2 \varphi_1$ , which is not equal to  $\Phi$ , the HL2 building.

### 3. RESULTS

#### (a) Behavioural results

Overall, participants performed well on all tasks as summarized in the following. Accuracy and reaction time (RT) for LANG: LIST (mean 96.9% (s.d. 3.3%), 1118.6 ms (261.7 ms)), HL1 (98.5% (5.2%), 1142 ms (273.8 ms)) and HL2 (93.9% (6.4%), 1473.3 ms (288.2 ms)); for ARITH: LIST (97.8% (6.2%), 913.5 ms (224.6 ms)), HL1 (98.5% (3.1%), 876.9 ms (200.9 ms)) and HL2 (94.3% (6.7%), 840.6 ms (215.9 ms)); for STRING: LIST (98.8% (2.3%), 932.2 ms (224.2 ms)), HL1 (97.8% (4.6%), 867.5 ms (195.2 ms)) and HL2 (96.4% (3.6%), 828.9 ms (187.1 ms)). We performed a  $3 \times 3$  within-subject ANOVA with factors HIERARCHY LEVEL and DOMAIN for accuracy and RT, respectively. For accuracy, a significant main effect of HIERARCHY LEVEL was revealed ( $F_{2,34} = 11.7^{**}$ , hereafter \* and \*\* denote  $p < 0.05$  and  $0.01$ , respectively), but not for DOMAIN ( $F_{2,34} = 1.56$  ns) nor for interaction ( $F_{4,68} = 0.99$  ns). A simple main effect analysis revealed  $LIST \approx HL1 > HL2$ . For RT, a significant interaction of the factors DOMAIN and



Table 2. Activation revealed by conjunction analysis for online sequence processing for LIST, HL1 and HL2 stimuli and the contrast HL1 > LIST across three domains. sup, superior; mid, middle; med, medial; inf, inferior; ant, anterior; post, posterior; g, gyrus; s, sulcus; L, left; R, right; PMv, ventral premotor area; POd, dorsal portion of the pars opercularis; SMA, supplementary motor area. Cluster size is given in number of voxels (1 voxel is  $3 \times 3 \times 3$  mm). When cluster p and cluster size are blank, it means the peak is found in the same cluster listed immediately above.

| anatomical region        | cluster p | cluster size | peak Z | coordinates |     |     |
|--------------------------|-----------|--------------|--------|-------------|-----|-----|
| LIST                     |           |              |        |             |     |     |
| PMv L                    | 0.009     | 69           | 4.06   | -57         | 5   | 19  |
| POd L                    |           |              | 3.26   | -42         | 5   | 22  |
| postcentral g. L         | 0         | 458          | 4.88   | -39         | -25 | 49  |
| SMA L                    | 0         | 181          | 4.91   | -6          | -7  | 52  |
| post. inf. temporal g. L | 0         | 305          | 5.53   | -45         | -64 | -5  |
| inf. occipital g. R      | 0.020     | 57           | 5.57   | 33          | -91 | -5  |
| cerebellum R             | 0         | 132          | 4.34   | 21          | -52 | -23 |
| HL1                      |           |              |        |             |     |     |
| PMv L                    | 0         | 513          | 6.17   | -51         | 5   | 40  |
| POd L                    |           |              | 5.58   | -57         | 8   | 25  |
| PMv R                    | 0         | 445          | 5.45   | 51          | 8   | 37  |
| post. inf. temporal g. L | 0         | 413          | 5.38   | -45         | -61 | -8  |
| SMA L                    | 0         | 444          | 4.96   | -6          | 8   | 52  |
| intraparietal s. L       | 0         | 185          | 5.26   | -27         | -61 | 43  |
| intraparietal s. R       | 0         | 215          | 5.36   | 30          | -55 | 43  |
| retrosplenial cortex     | 0         | 58           | 4.86   | 0           | -34 | 28  |
| inf. occipital g. R      | 0.009     | 68           | 5.30   | 33          | -91 | -5  |
| cerebellum R             | 0         | 317          | 5.28   | 21          | -49 | -23 |
| globus pallidus L        | 0         | 241          | 5.09   | -24         | -4  | 4   |
| HL2                      |           |              |        |             |     |     |
| PMv L                    | 0         | 443          | 5.95   | -54         | 2   | 40  |
| POd                      |           |              | 4.96   | -42         | 8   | 31  |
| precentral g. R          | 0         | 452          | 5.18   | 48          | -1  | 52  |
| preSMA L                 | 0         | 399          | 5.66   | -3          | 14  | 49  |
| precentral g. L          | 0.035     | 59           | 4.08   | -27         | -16 | 55  |
| intraparietal s. R       | 0         | 191          | 5.19   | 33          | -58 | 46  |
| inf. temporal g. L       | 0         | 840          | 6.16   | -42         | -64 | -8  |
| inf. occipital g. R      | 0         | 550          | 6.38   | 33          | -91 | -5  |
| globus pallidus L        | 0         | 155          | 4.04   | -21         | 5   | -2  |
| HL1 > LIST               |           |              |        |             |     |     |
| PMv L                    | 0.054     | 44           | 3.11   | -51         | 2   | 46  |

HIERARCHY LEVEL was detected ( $F_{4,68} = 76.2^{**}$ ). Analyses for simple main effect revealed a significant effect of DOMAIN in all hierarchy levels (LANG > ARITH  $\approx$  STRING), and effect of HIERARCHY LEVEL at LANG (HL1  $\approx$  LIST < HL2), ARITH (LIST > HL2, LIST  $\approx$  HL1, HL1  $\approx$  HL2) and STRING (LIST > HL1 > HL2), respectively.

### (b) Imaging results

Several different analyses were conducted to evaluate the brain activation of sequence processing in the different domains. First, a conjunction analysis across the three domains was performed in LIST, HL1 and HL2 trials, respectively. The results are presented in figure 3. Sequence processing during HL1 and HL2 activates the ventral premotor area (PMv) and the POd, including the depth of the inferior frontal sulcus, and inferior temporal/occipital regions (figure 3 and table 2). The cluster in the PMv was the only region that was seen to be activated for the HL1 > LIST contrast in the conjunction analysis across the three domains (figure 3 and table 2), indicating that the PMv is involved in the building of the HL1 common to the three domains.

The domain-specific activation of sequence processing was revealed by LANG > STRING and ARITH > STRING contrasts (table 3). Compared with STRING, LANG shows more activation in the pars triangularis, pars orbitalis, superior temporal sulcus, hippocampal gyrus and retrosplenial cortex. Compared with STRING, ARITH showed increased activation in the PO bilaterally.

We reconstructed a TTC for each stimulus type in the POd and PMv VOIs (for VOI definition, see §2) (figure 4) and obtained the peak BOLD signals used in the computation of difference in the building cost ( $\Delta$ COST). In the reconstruction of the TTCs, we did not explicitly model stimuli and probe as distinct neural events, but checked whether different BOLD signal peaks corresponding to stimuli and probes are visible in the plots (figure 4). We averaged the TTCs of all the domains and confirmed two separate peaks for the processing of stimuli and probes. We calculated the times to peaks for the sequence stimuli processing as 3.2 s for LIST and HL1, and 8.0 s for HL2 trials after the stimuli onset. The TTC plots revealed that the POd activation for stimulus processing (between 4 and 8 s) is much lower for LANG



Table 3. Activation revealed by the direct comparison of online sequence processing of HL2 stimuli of LANG versus STRING and that of ARITH versus STRING. sup, superior; mid, middle; med, medial; inf, inferior; ant, anterior; post, posterior; g, gyrus; s, sulcus; L, left; R, right; PTr, pars triangularis; PO, pars orbitalis. Cluster size is given in number of voxels (1 voxel is  $3 \times 3 \times 3$  mm). When cluster p and cluster size are blank, the peak is found in the same cluster listed immediately above.

| anatomical region        | cluster p | cluster size | peak Z | coordinates |     |     |
|--------------------------|-----------|--------------|--------|-------------|-----|-----|
| LANG > STRING            |           |              |        |             |     |     |
| PO L                     | 0         | 501          | 5.79   | −48         | 26  | −2  |
| PTr L                    |           |              | 5.76   | −51         | 26  | 10  |
| ant. sup. temporal s. L  |           |              | 5.42   | −48         | 5   | −20 |
| mid. sup. temporal s. L  |           |              | 5.17   | −54         | −19 | −2  |
| PO R                     | 0         | 150          | 4.31   | 45          | 32  | −11 |
| ant. sup. temporal s. R  |           |              | 4.85   | 63          | 2   | −11 |
| med. sup. frontal g. L   | 0         | 202          | 5.48   | −9          | 56  | 34  |
| post. sup. temporal s. L | 0.004     | 75           | 4.67   | −51         | −67 | 25  |
| retrosplenial cortex L   | 0         | 135          | 5.01   | −6          | −52 | 16  |
| post. inf. temporal g. R | 0         | 660          | 5.89   | 30          | −73 | −14 |
| hippocampal g. L         | 0         | 915          | 5.82   | −21         | −7  | −17 |
| mid. occipital g. L      | 0.003     | 78           | 4.38   | −12         | −94 | 25  |
| ARITH > STRING           |           |              |        |             |     |     |
| PO L                     | 0.003     | 77           | 5.07   | −48         | 8   | 22  |
| PO R                     | 0.008     | 64           | 5.05   | 48          | 8   | 19  |

than ARITH and STRING, but higher for probe processing (between 9 and 16 s). Although the conjunction analyses (figure 3) suggest that the POd is domain-generally involved in online hierarchical structure building of the stimulus sequence (in both HL1 and HL2), the TTC analysis provides a finer grained picture, suggesting the effectiveness of the POd for native language processing at higher hierarchy levels.

In order to statistically compare  $\Delta$ COST at the lower and higher hierarchy levels, we computed the difference between the BOLD signal peak height of HL2 and HL1, and HL1 and LIST. Since we had to compare the BOLD signals of the trials of different durations, we converted the BOLD signals into the unit of signal increase per cent per second. Without this conversion of the unit, the difference between HL2 and HL1 trials is difficult to compare due to the difference in the duration of the stimuli. The  $\Delta$ COSTs were analysed in a within-subject three-way ANOVA with factors  $\Delta$ COST (HL2–HL1/HL1–LIST), DOMAIN (LANG/ARITH/STRING) and REGION (PMv/POd). A significant interaction between  $\Delta$ COST and REGION ( $F_{1,17} = 17.06^{**}$ ) was found (figure 5), but the three-way interaction ( $F_{2,34} = 0.17$ ) and interaction between  $\Delta$ COST and DOMAIN ( $F_{1,17} = 2.75$  ( $p < 0.1$ )) did not reach significance. These results suggest no differences between domains with respect to  $\Delta$ COST (i.e. the difference in brain activation between hierarchy levels.) The latter interaction suggests that the opposite patterns of the  $\Delta$ COSTs in the PMv and POv are statistically significant. The simple main effect analysis for the latter interaction revealed a significant effect of  $\Delta$ COST at the PMv ( $F_{1,17} = 5.97^{*}$ ) but not at the POd ( $F_{1,17} = 0.48$  ns), and significant effects of REGION at both  $\Delta$ COSTs ( $F_{1,17} = 8.47^{**}$ ,  $17.56^{**}$  for HL2–HL1 and HL1–LIST, respectively).

#### 4. DISCUSSION

The present study investigated the neural basis of hierarchical structure building at different hierarchy levels in two cognitive domains, language and mathematics.

##### (a) Level of hierarchy in structure building in the lateral prefrontal cortex

The results suggest that posterior sub-regions of the lateral PFC are functionally organized in a posterior-to-anterior gradient to implement hierarchical structure building at different levels in language and mathematics. The analyses of the building costs revealed that hierarchical structure building with two successive externally given stimuli (HL1), relative to mere encoding (LIST), activates the PMv, while the building of a higher-level hierarchy with two mental representations (HL2) activates a more anterior region, weighted on the dorsal part of the pars opercularis (POd), i.e. the posterior dorsal part of Broca's area (Brodmann's area 44; figure 5). Although the estimation of the difference in the integration cost between the hierarchy conditions of different trial durations is based on a novel method, and may therefore need empirical validation in the same as well as in different modalities, the current results provide first evidence for a posterior-to-anterior gradient as structural hierarchy increases.

The shift of activation pattern from a posterior to a more anterior region in the lateral PFC, as a function of the increase in the hierarchy level, appears to be compatible with the view that the lateral PFC is hierarchically organized along the posterior-to-anterior axis for different levels of cognitive control [21–24]. However, we face a gap when applying the hierarchical cognitive control hypothesis with its posterior-to-anterior gradient for contextual, episodic and branching control to language. Firstly, accumulating evidence suggests that centre-embedding structures are processed

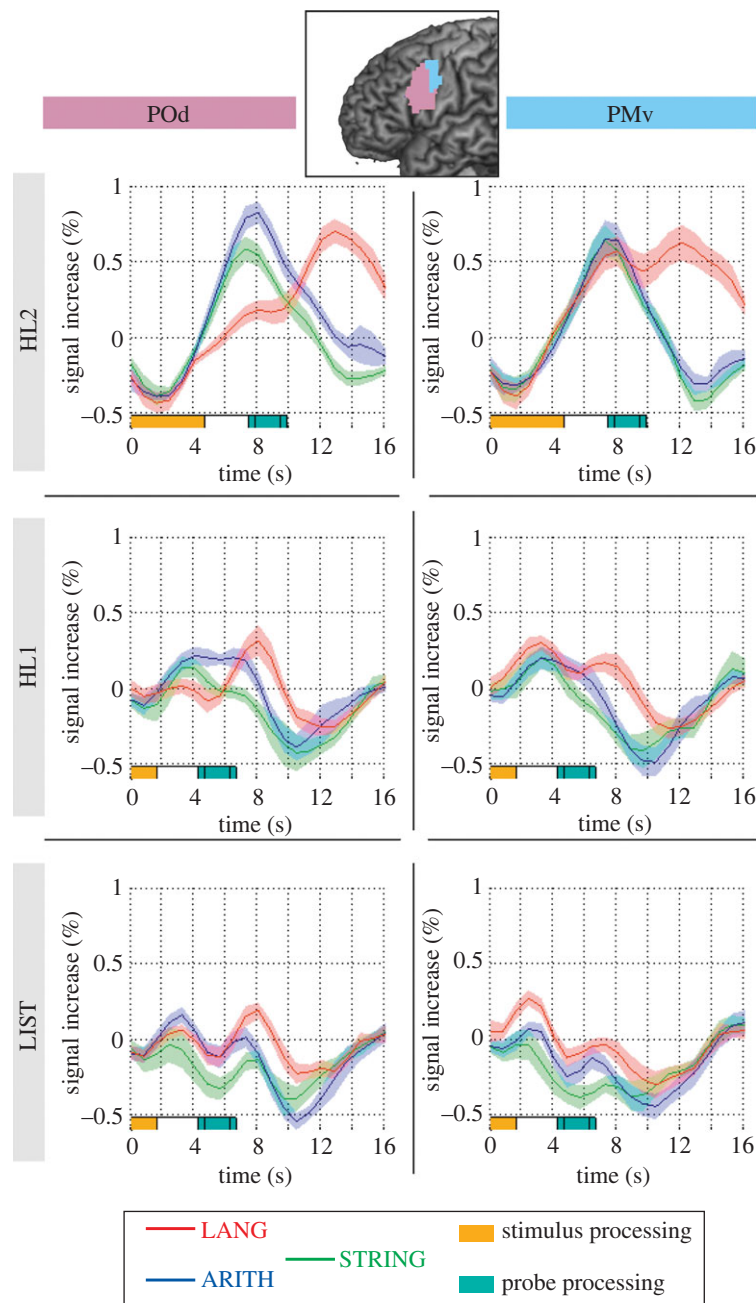


Figure 4. TTC plots for PMv and POd VOIs. The heading schematically shows the locations of the POd (pink, left column) and PMv (light blue, right column) activation clusters. TTC plots for HL2 trials (first row), HL1 trials (second row) and LIST trials (third row) are shown. In HL2 trials, two peaks appear at around 8 s and 13 s, corresponding to the responses to the stimulus and to the probe processing, respectively. In HL1 and LIST trials, two peaks appear at around 3 and 8 s, which correspond to the responses to the stimulus and to the probe processing, respectively. Red, language (LANG); blue, arithmetic (ARITH); green, string operation (STRING). Coloured bands indicate s.e. The bars at the left bottom of each plot represent the period when stimuli (yellow) and probes (cyan) are presented. The probes were presented: either 7.4–9.4 s or 7.8–9.8 s for the HL2 condition, and either 4.2–6.2 s or 4.6–6.6 s for the LIST and HL1 conditions.

in the PO [6,10,11,13,14], and that such structures require cognitive branching. For example, in a centre-embedded structure such as in  $[A_1[A_2B_2]B_1]$ , the  $A_1$  is processed first and kept on hold, then the inner structure is processed before the  $B_1$  is processed. According to the hierarchical cognitive control hypothesis [21,23], such branching processes should activate the anterior polar cortex. Secondly, it is conceptually unclear whether hierarchical structure building in language/arithmetic processing corresponds to ‘action selection’ in the hierarchical cognitive control hypothesis. Thus, further

theoretical and empirical work is needed to integrate language processing into the hypothesis.

#### (b) Processing of event sequences in the lateral prefrontal cortex

In the present study, the dorsal part of PO was seen to be activated for HL2 building in language and mathematics. In general, there is convergent evidence suggesting that the PO is involved in the processing of event sequences in various forms, such as visual

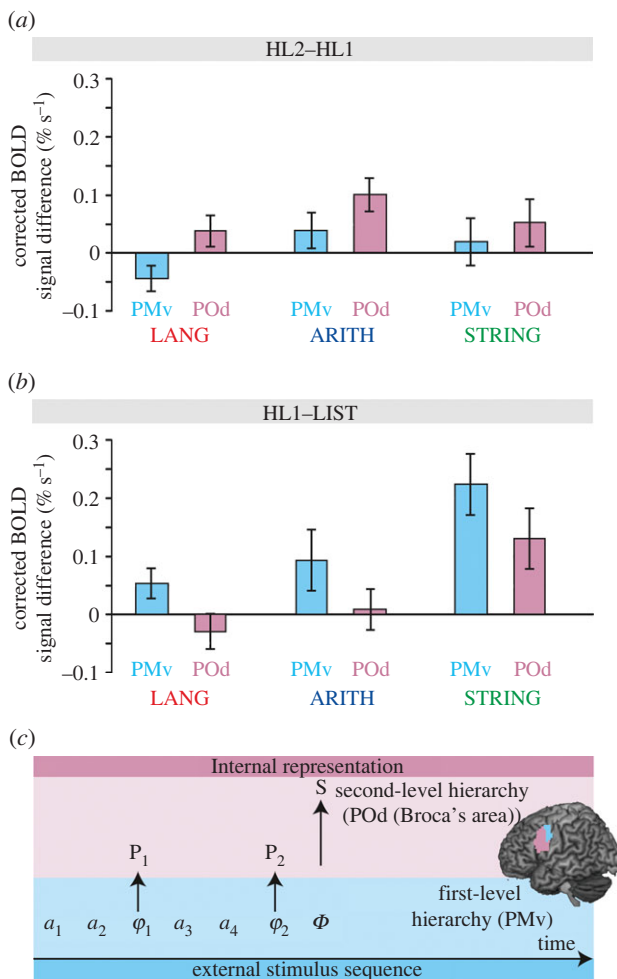


Figure 5. Building of the hierarchical structure at the first- and second-levels. (a) The difference in hierarchical structure building costs ( $\Delta\text{COST}$ ) between HL2 and HL1 in online sequence processing. This is the difference between  $\text{PP}\Phi$  and  $\text{aa}\phi$  with correction for the difference in duration of the stimuli. (b)  $\Delta\text{COST}$  between HL1 and LIST in online sequence processing. This is the difference between  $\text{aa}\phi$  and  $\text{aa}$ . Note that the difference computed in HL1–LIST corresponds to the HL1 building cost  $\phi$ , the difference computed in HL2–HL1 does not correspond to the HL1 building cost  $\Phi$ , but to the difference between  $\text{PP}\Phi$  and  $\text{aa}\phi$ . Error bars show standard errors. See figure 3 for the reconstructed TTCs in the two VOIs. (c) Schema for hierarchical structure building and the supporting neural basis. For details, see figure 1 and table 1. PMv (light blue) and POd (pink) are projected on the surface of a template brain.

object sequences [30–32] and human action sequences [33,34]. Crucially, hierarchically structured sequences of syllables [6,13] and sentences with similar hierarchical structures selectively activate the PO [10,11,14]. From a neuroanatomical perspective, the present activation clusters covering the POd and PMv (figure 3) are close to or include the left inferior frontal junction (IFJ), which is activated by a task switch: a cognitive process required when several distinct tasks have to be applied in a sequence [35,36]. These findings suggest that a larger PFC region involving the PMv, POd and IFJ may play a role in updating internal representations, be it in processing structured sequences or in task representation during online sequence processing.

### (c) Language-specific activation pattern in Broca's area

When comparing language processing to the mathematics and working memory task, we observed that POd activation is significantly lower in LANG compared with ARITH and STRING (figure 4). This lower activation suggests that LANG stimuli are computed more efficiently in the POd when building hierarchical structures. Because the structural processing for native language sentences is over-learned and highly automatic, the local neural circuit may involve fewer neurons compared with the less practiced arithmetic and the newly learned string operation working memory task [37]. The reduced activation for language may also be due to facilitation by additional information that exists in language but not in other conditions. For example, in simple sentences like the ones used here, the thematic relations ('who is doing what to whom') can be recovered using case marking information and also semantic information (animacy) of noun phrases. Recent neuroimaging studies revealed that both information types activate PO when it comes to assigning thematic roles [38–40]. Psycholinguistic studies using English as the test language have demonstrated that people can comprehend sentences even by using a shallow syntactic analysis on some occasions, suggesting that heuristics can drive sentence processing without building complete syntactic hierarchical representation [41,42]. Such heuristic processing of a sentence may be reflected in respective brain regions in the LANG > STRING contrast (table 3). One region is the pars triangularis (PT<sub>r</sub>), which is known to reflect the lexico-semantic processing of words [7,43–47]. Other regions are the hippocampus/parahippocampal gyri and retrosplenial cortex. These two regions were found to be activated more in language compared with the working memory task in the present study. Some authors have reported that these two regions are involved in the processing of contextually associated memories [48] for generating predictions of forthcoming events [49]. Thus, in the argument paradigm in which elements are presented at slow pace semantic/heuristic processes may come into play (for non-syntactic strategies in sentence processing, see also [50]), thereby reducing the computational burden in the POd during sentence processing.

### (d) Hierarchical structure processing in mathematics

The present results are partly compatible with work on the processing of hierarchies in mathematics. Although the maximum of the activation for hierarchical structure processing in mathematics was found anterior to Broca's area, a post hoc ROI analysis revealed that Broca's area (PO) was also involved [19,20]. The recruitment of the more anterior region in the Friedrich & Friederici study can be put down to the fact that their mathematical formulae were highly abstract, whereas the present mathematical sequences contained natural numbers. The present findings are indeed compatible with behavioural results that showed structural priming from



mathematical sequences to language, i.e. from the parenthetical grouping in mathematical calculations to the syntactic structure of sentences [51]. These behavioural findings argue for a domain generality of structural representations and possibly processes.

However, our results are at variance with a recent imaging study comparing the brain activation across different domains, including language, mathematics and working memory [18]). Using a localizer task reflecting both lexical and syntactic aspects of language processing, the authors found little or no response by the language regions to non-linguistic domains. They argued for domain specificity for language of the pars orbitalis and more posterior IFG region, possibly including both the PO and PTr. Activations in the PO for the language condition in the former study would, however, be compatible with the present study. In addition, their working memory results are largely compatible with the present results. This is in contrast to the results for the mathematical conditions in the two studies. This may be due to differences between the actual calculations required in the study by Fedorenko *et al.* [18] and the present one. A critical difference between the arithmetic calculation following the Reverse Polish notation (figure 1) and those used by Fedorenko *et al.* ( $21 + 7 + 8 + 6$ ) is that in the latter case, the participant has to successively hold only one internal representation in order to perform the next processing steps. Our design, in contrast, required the operation on two internal representations in order to build the HL2. This difference may explain the activation of the PO in the mathematics condition in our design compared with that of Fedorenko *et al.* Thus, the PO seems to be recruited only when two (or more) internally represented elements are integrated to build a hierarchical structure.

### (e) Specific effects in the string operation task

In the HL1 > LIST contrast, POd activation is much higher in STRING than in LANG and ARITH, leading to non-significant  $\Delta$ COST effect in the POd (figure 5b). Interestingly, the STRING operation condition also differs from the other conditions in RT. An ANOVA for RT revealed significantly longer RT in HL1 compared with LIST only in the STRING condition. These results suggest that the processes required in the STRING condition are fundamentally different from those in the two other domains: adjective and noun in language as well as the two digits in arithmetic are actually integrated, whereas in the string operation working memory task they are only selected.

## 5. CONCLUSION

Using a novel method to compute the difference in the integration cost between the conditions of different trial durations, we showed that the left posterior PFC is functionally organized to subserve the build-up of hierarchical structures at different levels in online sequence processing. The build-up of the HL1 with the externally given elements is supported by the PMv, whereas the second-level hierarchical build-up with internally represented elements recruits the POd. This mechanism holds across language and

mathematics, but with more efficient processing in language compared with other domains.

We are grateful to Kayo Yamauchi and Aya Kugele for the recruitment of participants, Annet Wiedemann, Simone Wipper and Anke Kummer for MRI data acquisition, to Kerstin Flake and Andrea Gast-Sandmann for the graphics, and to Helga Smallwood for English correction.

## REFERENCES

- 1 Johnson, N. F. 1965 The psychological reality of phrase-structure rules. *J. Verbal Learn. Verbal Behav.* **4**, 469–475. (doi:10.1016/S0022-5371(65)80044-5)
- 2 Fodor, J. A. & Bever, T. G. 1965 The psychological reality of linguistic segments. *J. Verbal Learn. Verbal Behav.* **4**, 414–420. (doi:10.1016/S0022-5371(65)80081-0)
- 3 Levelt, W. 1970 Hierarchical chunking in sentence processing. *Percept. Psychophys.* **8**, 99–103. (doi:10.3758/BF03210182)
- 4 Shieber, S. M. 1986 *An introduction to unification-based approaches to grammar*. CSLI Lecture Notes Series, 4 Stanford, CA: Center for the Study of Language and Information.
- 5 Chomsky, N. 1995 *minimalist program*. Cambridge, MA: MIT Press.
- 6 Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I. & Anwander, A. 2006 The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl Acad. Sci. USA* **103**, 2458–2463. (doi:10.1073/pnas.0509389103)
- 7 Friederici, A. D., Ruschmeyer, S. A., Hahne, A. & Fiebach, C. J. 2003 The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb Cortex* **13**, 170–177. (doi:10.1093/cercor/13.2.170)
- 8 Ruschmeyer, S. A., Zysset, S. & Friederici, A. D. 2006 Native and non-native reading of sentences: an fMRI experiment. *NeuroImage* **31**, 354–365. (doi:10.1016/j.neuroimage.2005.11.047)
- 9 Brauer, J. & Friederici, A. D. 2007 Functional neural networks of semantic and syntactic processes in the developing brain. *J. Cogn. Neurosci.* **19**, 1609–1623. (doi:10.1162/jocn.2007.19.10.1609)
- 10 Stromswold, K., Caplan, D., Alpert, N. & Rauch, S. 1996 Localization of syntactic comprehension by positron emission tomography. *Brain Lang.* **52**, 452–473. (doi:10.1006/brln.1996.0024)
- 11 Caplan, D., Chen, E. & Waters, G. 2008 Task-dependent and task-independent neurovascular responses to syntactic processing. *Cortex* **44**, 257–275. (doi:10.1016/j.cortex.2006.06.005)
- 12 Opitz, B. & Friederici, A. D. 2007 Neural basis of processing sequential and hierarchical syntactic structures. *Hum. Brain Mapp.* **28**, 585–592. (doi:10.1002/hbm.20287)
- 13 Bahlmann, J., Schubotz, R. I. & Friederici, A. D. 2008 Hierarchical artificial grammar processing engages Broca's area. *NeuroImage* **42**, 525–534. (doi:10.1016/j.neuroimage.2008.04.249)
- 14 Makuuchi, M., Bahlmann, J., Anwander, A. & Friederici, A. D. 2009 Segregating the core computational faculty of human language from working memory. *Proc. Natl Acad. Sci. USA* **106**, 8362–8367. (doi:10.1073/pnas.0810928106)
- 15 Dehaene, S., Molko, N., Cohen, L. & Wilson, A. J. 2004 Arithmetic and the brain. *Curr. Opin. Neurobiol.* **14**, 218–224. (doi:10.1016/j.conb.2004.03.008)
- 16 Anderson, J. R., Fincham, J. M., Qin, Y. & Stocco, A. 2008 A central circuit of the mind. *Trends Cogn. Sci.* **12**, 136–143. (doi:10.1016/j.tics.2008.01.006)



- 17 Nieder, A. & Dehaene, S. 2009 Representation of number in the brain. *Annu. Rev. Neurosci.* **32**, 185–208. (doi:10.1146/annurev.neuro.051508.135550)
- 18 Fedorenko, E., Behr, M. K. & Kanwisher, N. 2011 Functional specificity for high-level linguistic processing in the human brain. *Proc. Natl Acad. Sci. USA* **108**, 16 428–16 433. (doi:10.1073/pnas.1112937108)
- 19 Friedrich, R. & Friederici, A. D. 2009 Mathematical logic in the human brain: syntax. *PLoS ONE* **4**, e5599. (doi:10.1371/journal.pone.0005599)
- 20 Friederici, A. D., Bahlmann, J., Friedrich, R. & Makuuchi, M. 2011 The neural basis of recursion and complex syntactic hierarchy. *Biolinguistics* **5**, 87–104.
- 21 Koechlin, E., Ody, C. & Kouneiher, F. 2003 The architecture of cognitive control in the human prefrontal cortex. *Science* **302**, 1181–1185. (doi:10.1126/science.1088545)
- 22 Koechlin, E. & Jubault, T. 2006 Broca's area and the hierarchical organization of human behavior. *Neuron* **50**, 963–974. (doi:10.1016/j.neuron.2006.05.017)
- 23 Koechlin, E. & Summerfield, C. 2007 An information theoretical approach to prefrontal executive function. *Trends Cogn. Sci.* **11**, 229–235. (doi:10.1016/j.tics.2007.04.005)
- 24 Badre, D., Hoffman, J., Cooney, J. W. & D'Esposito, M. 2009 Hierarchical cognitive control deficits following damage to the human frontal lobe. *Nat. Neurosci.* **12**, 515–522. (doi:10.1038/nn.2277)
- 25 Oldfield, R. C. 1971 The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97–113. (doi:10.1016/0028-3932(71)90067-4)
- 26 Hirotani, M., Makuuchi, M., Ruschemeyer, S. A. & Friederici, A. D. 2011 Who was the agent? The neural correlates of reanalysis processes during sentence comprehension. *Hum. Brain Mapp.* **32**, 1775–1787. (doi:10.1002/hbm.21146)
- 27 Nichols, T., Brett, M., Andersson, J., Wager, T. & Poline, J. B. 2005 Valid conjunction inference with the minimum statistic. *NeuroImage* **25**, 653–660. (doi:10.1016/j.neuroimage.2004.12.005)
- 28 Roland, P. E. & Zilles, K. 1998 Structural divisions and functional fields in the human cerebral cortex. *Brain Res. Rev.* **26**, 87–105. (doi:10.1016/S0165-0173(97)00058-1)
- 29 Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K. & Zilles, K. 2005 A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage* **25**, 1325–1335. (doi:10.1016/j.neuroimage.2004.12.034)
- 30 Schubotz, R. I. & von Cramon, D. Y. 2001 Functional organization of the lateral premotor cortex: fMRI reveals different regions activated by anticipation of object properties, location and speed. *Brain Res. Cogn. Brain Res.* **11**, 97–112. (doi:10.1016/S0926-6410(00)00069-0)
- 31 Schubotz, R. I. & von Cramon, D. Y. 2004 Sequences of abstract nonbiological stimuli share ventral premotor cortex with action observation and imagery. *J. Neurosci.* **24**, 5467–5474. (doi:10.1523/JNEUROSCI.1169-04.2004)
- 32 Bahlmann, J., Schubotz, R. I., Mueller, J. L., Koester, D. & Friederici, A. D. 2009 Neural circuits of hierarchical visuospatial sequence processing. *Brain. Res.* **1298**, 161–170. (doi:10.1016/j.brainres.2009.08.017)
- 33 Binkofski, F. & Buccino, G. 2006 The role of ventral premotor cortex in action execution and action understanding. *J. Physiol.* **99**, 396–405.
- 34 Fazio, P., Cantagallo, A., Craighero, L., D'Ausilio, A., Roy, A. C., Pozzo, T., Calzolari, F., Granieri, E. & Fadiga, L. 2009 Encoding of human action in Broca's area. *Brain* **132**, 1980–1988. (doi:10.1093/brain/awp118)
- 35 Brass, M., Derrfuss, J., Forstmann, B. & von Cramon, D. Y. 2005 The role of the inferior frontal junction area in cognitive control. *Trends Cogn. Sci.* **9**, 314–316. (doi:10.1016/j.tics.2005.05.001)
- 36 Derrfuss, J., Brass, M., Neumann, J. & von Cramon, D. Y. 2005 Involvement of the inferior frontal junction in cognitive control: meta-analyses of switching and Stroop studies. *Hum. Brain Mapp.* **25**, 22–34. (doi:10.1002/hbm.20127)
- 37 Grill-Spector, K., Henson, R. & Martin, A. 2006 Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* **10**, 14–23. (doi:10.1016/j.tics.2005.11.006)
- 38 Bornkessel, I., Zysset, S., Friederici, A. D., von Cramon, D. Y. & Schlesewsky, M. 2005 Who did what to whom? The neural basis of argument hierarchies during language comprehension. *NeuroImage* **26**, 221–233. (doi:10.1016/j.neuroimage.2005.01.032)
- 39 Grewe, T., Bornkessel, I., Zysset, S., Wiese, R., von Cramon, D. Y. & Schlesewsky, M. 2005 The emergence of the unmarked: a new perspective on the language-specific function of Broca's area. *Hum. Brain Mapp.* **26**, 178–190. (doi:10.1002/hbm.20154)
- 40 Bornkessel, I. & Schlesewsky, M. 2006 The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychol. Rev.* **113**, 787–821. (doi:10.1037/0033-295X.113.4.787)
- 41 Ferreira, F., Ferraro, V. & Bailey, K. G. D. 2002 Good-enough representations in language comprehension. *Curr. Direct. Psychol. Sci.* **11**, 11–15. (doi:10.1111/1467-8721.00158)
- 42 Ferreira, F. & Patson, N. D. 2007 The 'good enough' approach to language comprehension. *Lang. Linguist. Compass* **1**, 71–83. (doi:10.1111/j.1749-818X.2007.00007.x)
- 43 Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M. & Raichle, M. E. 1988 Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* **331**, 585–589. (doi:10.1038/331585a0)
- 44 Devlin, J. T., Matthews, P. M. & Rushworth, M. F. 2003 Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *J. Cogn. Neurosci.* **15**, 71–84. (doi:10.1162/089892903321107837)
- 45 Newman, S. D., Just, M. A., Keller, T. A., Roth, J. & Carpenter, P. A. 2003 Differential effects of syntactic and semantic processing on the subregions of Broca's area. *Brain Res. Cogn. Brain Res.* **16**, 297–307. (doi:10.1016/S0926-6410(02)00285-9)
- 46 Hagoort, P., Hald, L., Bastiaansen, M. & Petersson, K. M. 2004 Integration of word meaning and world knowledge in language comprehension. *Science* **304**, 438–441. (doi:10.1126/science.1095455)
- 47 Newman, S. D., Ikuta, T. & Burns, T. 2010 The effect of semantic relatedness on syntactic analysis: an fMRI study. *Brain Lang.* **113**, 51–58. (doi:10.1016/j.bandl.2010.02.001)
- 48 Bar, M. & Aminoff, E. 2003 Cortical analysis of visual context. *Neuron* **38**, 347–358. (doi:10.1016/S0896-6273(03)00167-3)
- 49 Bar, M. 2007 The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci.* **11**, 280–289. (doi:10.1016/j.tics.2007.05.005)
- 50 Bever, T. G. 1970 The cognitive basis for linguistic structures. In *Cognition and language development* (ed. R. Hayes), pp. 277–360. New York, NY: Wiley & Sons, Inc.
- 51 Scheepers, C., Sturt, P., Martin, C. J., Myachikov, A., Teevan, K. & Viskupova, I. 2011 Structural priming across cognitive domains: from simple arithmetic to relative-clause attachment. *Psychol. Sci.* **22**, 1319–1326. (doi:10.1177/0956797611416997)