

The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees

Stanislas Dehaene,^{1,2,*} Florent Meyniel,² Catherine Wacongne,^{1,2} Liping Wang,^{3,4} and Christophe Pallier²

¹Collège de France, 11 Place Marcelin Berthelot, 75005 Paris, France

²Cognitive Neuroimaging Unit, CEA DSV/I2BM, INSERM, Université Paris-Sud, Université Paris-Saclay, NeuroSpin center, 91191 Gif/Yvette, France

³Key Laboratory of Brain Functional Genomics (MOE & STCSM), Institute of Cognitive Neuroscience, School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China

⁴NYU-ECNU Institute of Brain and Cognitive Science at NYU Shanghai, Shanghai 200062, China

*Correspondence: stanislas.dehaene@cea.fr

<http://dx.doi.org/10.1016/j.neuron.2015.09.019>

A sequence of images, sounds, or words can be stored at several levels of detail, from specific items and their timing to abstract structure. We propose a taxonomy of five distinct cerebral mechanisms for sequence coding: transitions and timing knowledge, chunking, ordinal knowledge, algebraic patterns, and nested tree structures. In each case, we review the available experimental paradigms and list the behavioral and neural signatures of the systems involved. Tree structures require a specific recursive neural code, as yet unidentified by electrophysiology, possibly unique to humans, and which may explain the singularity of human language and cognition.

As early as the 1950s, the problem of serial order was identified by Karl Lashley as one of the pressing questions that behavioral and neural sciences should address (Lashley, 1951). The problem can be stated succinctly: how does the brain encode temporal sequences of items, such that this knowledge can be used to retrieve a sequence from memory, recognize it, anticipate on forthcoming items, and generalize this knowledge to novel sequences with a similar structure?

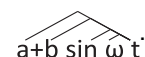
Lashley noted that language perception and production, but also bird song or rat spatial navigation behavior, presented special problems for the then-dominant view of associative chains. Humans and other animals do not simply associate each successive item with the next one at a particular delay, but they also grasp abstract multi-item sequential structures. This faculty is most evident in human language: even a single word such as “inexplicably” may consist in a nested structure of morphemes [in-[explic-able]]-ly].

Sixty years of linguistic analysis have confirmed that an accurate representation of language requires the postulation of nested tree structures (Chomsky, 1956). In parallel, behavioral and neurophysiological analyses of much simpler paradigms, involving for instance sequences of tones or gestures, have revealed a rich array of responses that go way beyond the simple associative chain (Restle, 1970; Restle and Brown, 1970). The purpose of the present article is to review those behavioral and neural findings and to provide a minimal taxonomy of brain mechanisms that any accurate model of sequence processing should emulate. We argue that there is evidence for a minimum of five distinct systems capable of representing sequence knowledge at increasing degrees of abstraction (Figure 1):

- Transition and timing knowledge: knowledge of the transitions from one item to the next (i.e., the identity and

approximate timing of the next item relative to the preceding ones).

- Chunking: the grouping of several contiguous items into a single unit that can be manipulated as a whole at the next hierarchical level.
- Ordinal knowledge: knowledge of which item comes first, which comes second, and so on, independently of their timing.
- Algebraic patterns: abstract schemas that capture the sequential regularities underlying a sequence of items; for instance, the word “cocolith” comprises twice the same syllable followed by a different one (AAB pattern).
- Nested tree structures generated by symbolic rules: at this level, characteristic of human languages, a sequence can be “parsed” according to abstract grammatical rules into a set of groupings, possibly embedded within each other, forming a nested structure of arbitrary depth, and possibly involving the recursive use of the same elements at multiple levels; an example is the parsing of the mathematical equation $a + b \sin \omega t$ as a nested set of parentheses $(a + (b (\sin (\omega t))))$ or, equivalently, a tree structure:



Transition and Timing Knowledge

Many animal species are able to represent the time intervals between sensory or motor events and use these temporal representations in simple computations. An excellent example is provided by a temporal choice task that has been used to probe temporal and probabilistic calculations in mice and humans (Balci et al., 2009; Kheifets and Gallistel, 2012). On each trial,

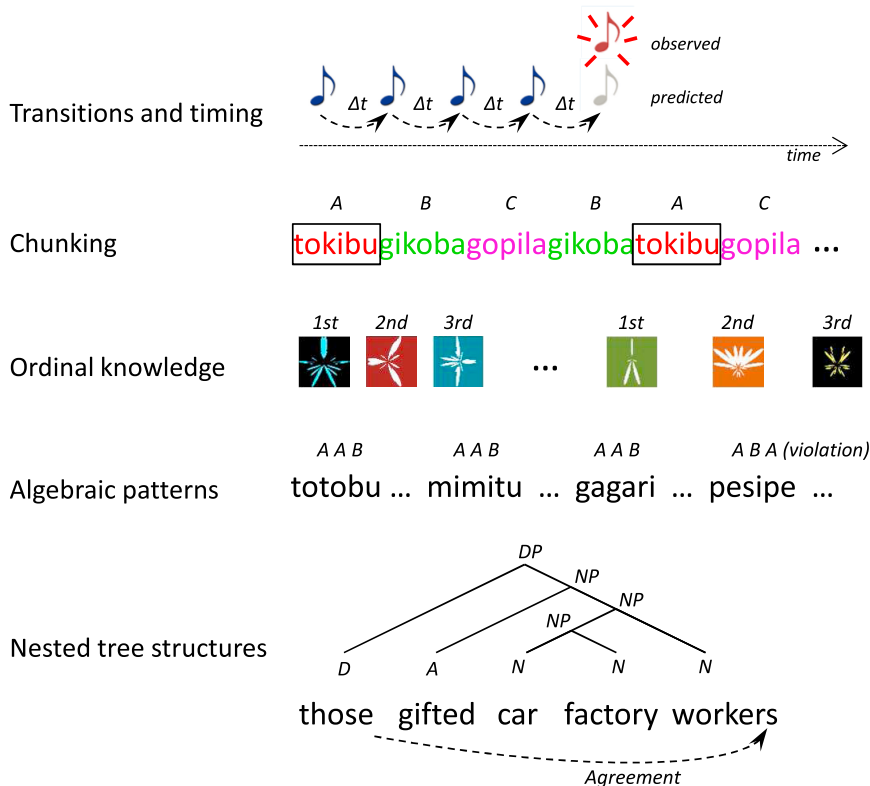


Figure 1. Illustration of the Proposed Taxonomy of Sequence Knowledge

According to our proposal, incoming sequences can be encoded internally at one of five possible levels of abstraction: (1) transitions between specific items at a specific time delay; (2) a sequence of “chunks,” for instance reproducible words within a stream of syllables; (3) an ordered list, with explicit knowledge of which item comes first, second, third...; (4) an algebraic pattern such as AAB, indicating that the first two items are identical while the third is different; and (5) a tree structure, with constituents nested inside other constituents, as observed in human languages.

when one of the stimuli is unexpectedly replaced by another one, the brain emits a mismatch response (MMR) (i.e., an activation increasing with the degree of mismatch between the expected and obtained stimuli) (Garrido et al., 2009). Most strikingly, a similar brain response is observed when the expected stimulus is omitted, peaking roughly at the time when the stimulus should have occurred (Raij et al., 1997; Wacongne et al., 2011). Such omission signals, which have been traced back to auditory cortices, incontrovertibly demonstrate that sensory circuits can internalize the timing of a regular

one of two events may occur: either, 3 s after trial initiation, a left lever press is rewarded or, 9 s after trial initiation, a right lever press is rewarded. Participants quickly learn to adjust their anticipations, first turning to the left lever, then if nothing occurs, switching to the right lever. The results show that mice and humans carefully and near-optimally adjust the duration after which they switch levers, taking into account both the imposed temporal delays and the internal and external uncertainties over which event is most likely. **Because the behavioral switches occur after a roughly fixed delay, which is not cued by any sensory event, it implies that the organism must maintain an internal representation of elapsed time** and base its decisions on a comparison of elapsed and memorized times of expected events. This and many other similar paradigms imply that time is one of the dimensions over which animals may compute (Gallistel, 1990).

A characteristic signature of this representation is its approximate nature, subject to scalar variability, also termed Weber’s law (Gibbon, 1977; Gibbon et al., 1997): **the imprecision (standard deviation) with which a delay is encoded is directly proportional to its duration, such that increasingly longer delays are represented with proportionally larger variability.**

It is likely that, whenever we hear a sequence of stimuli X_1, \dots, X_n , their transitions $X_i \rightarrow X_{i+1}$ and corresponding time delays are automatically and unconsciously registered, and **that such knowledge of temporal delays therefore constitutes a first level at which incoming sequences are internally represented.** The main paradigm that has been used to demonstrate this is the auditory oddball paradigm, in which a reproducible series of auditory stimuli is presented at regular intervals (Figure 2A). Crucially,

sequence and generate an endogenous response in the absence of any sensory input, purely in anticipation of an expected event. Variations in stimulus onset asynchrony indicate that this capacity to store temporal intervals operates up to delays of 2–5 s, with a decreasing amplitude and an increasing temporal dilution compatible with scalar variability (Mäntysalo and Näätänen, 1987; Pegado et al., 2010).

Another remarkable characteristic of temporal sequence encoding is its automaticity. MMR responses continue to be emitted even of the absence of attention, awareness of changes, or even of any consciousness, as during coma, vegetative state, or when falling sleep (Bekinschtein et al., 2009; Strauss et al., 2015). In spite of this automaticity, MMR studies have revealed that auditory sequences are internally stored with a great variety of details. Changing the presence, pitch, identity, intensity, or duration of the expected stimuli all lead to MMRs localized to distinct sites in primary and secondary auditory cortices as well as prefrontal cortex (PFC) (e.g., Giard et al., 1995). Precisely timed neuronal responses, including mismatch and omission responses, have also been observed in many other modalities, e.g., with visual stimuli in early visual cortex (Gavornik and Bear, 2014; Namboodiri et al., 2015) and inferotemporal cortex (Meyer and Olson, 2011; Meyer et al., 2014), with action sequences in premotor cortex and basal ganglia (Bartolo et al., 2014; Crowe et al., 2014; Mello et al., 2015; Merchant et al., 2011), and with anticipated reward in dopamine neurons (Fiorillo et al., 2008). It is therefore likely that the neural mechanisms for encoding temporal knowledge are replicated in several brain circuits that operate automatically and in parallel to each other.

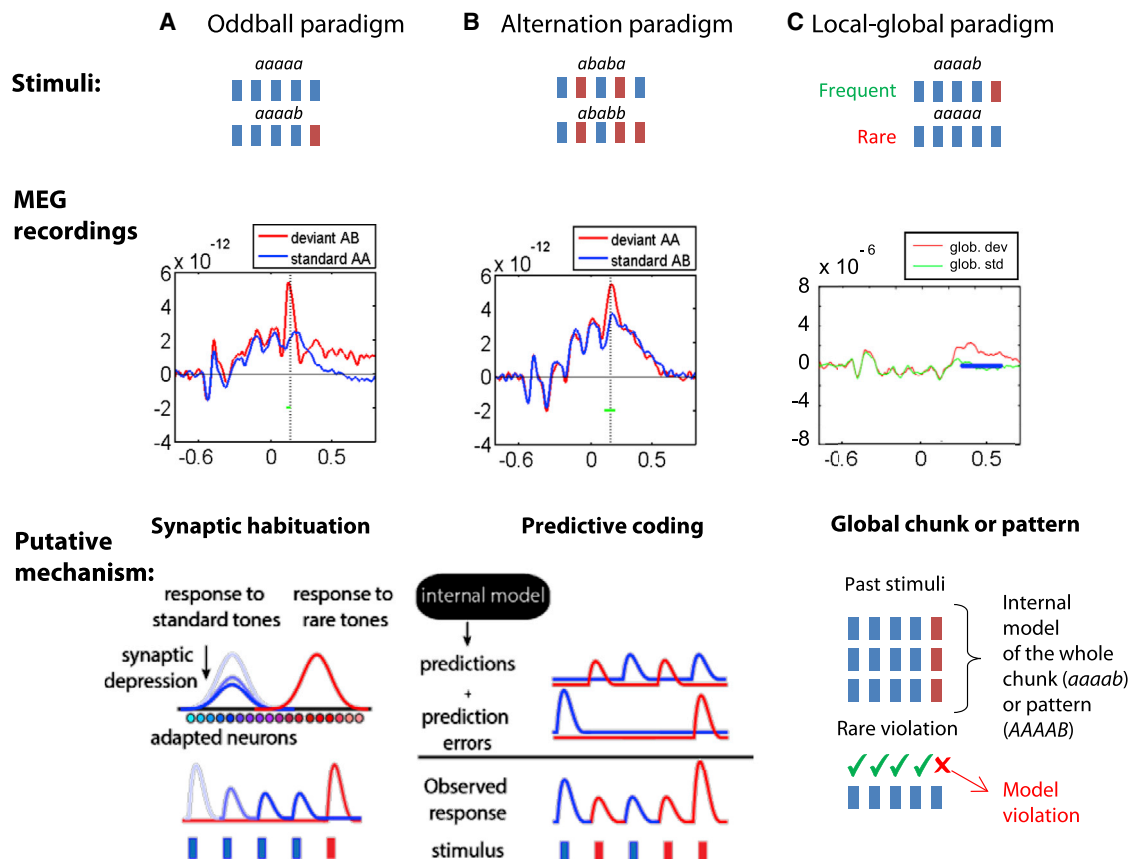


Figure 2. Simple Auditory-Violation Paradigms, Combined with Magneto-Encephalography Recordings, Can Dissociate Different Types of Sequence-Sensitive Mechanisms

Upon exposure to a sequence of repeated tones (*aaaaa* sequence, left column), auditory responses adapt, but recover when a novel tone is introduced (*aaaab* sequence, red curve). This “oddball” response is thought to result from two distinct mechanisms: passive sensory adaptation to individual tones (bottom left) and predictive coding based on transition probabilities between tones (bottom middle). The two mechanisms can be dissociated using a second auditory paradigm where two tones alternate (*ababa*, middle column): an unexpected repetition (*ababb*) leads to a sharp MMR, which is thought to result from a violation of the alternation prediction arising from past stimuli. Finally, a third paradigm, called local-global, presents a frequent sequence (in the illustrated example, the *aaaaa* sequence; right column). The occasional presentation of a rare deviant sequence (in this case, *aaaaa*) leads to a late novelty signal called “global response” because it arises from an integrated representation of the whole “chunk” (*aaaaa*) or abstract pattern (*AAAAA*) rather than of the individual tones or transitions between tones (adapted from data in [Strauss et al., 2015](#)).

At least two neural mechanisms for the generation of MMRs have been proposed: stimulus-specific adaptation (SSA) and predictive coding. SSA reflects the well-known fact that sensory responses tend to habituate over time, such as neurophysiological responses evoked by a stimulus decrease with repetition and recover when a fresh stimulus is presented. Predictive coding is the theory that cortical circuits form an internal model of input sequences, actively generate a prediction of upcoming items, and confront it with incoming stimuli. In this view, MMRs reflect the prediction error (i.e., the difference between prediction and reality) ([Friston, 2005](#)).

Although SSA undoubtedly contributes to the oddball paradigm ([May and Tiitinen, 2010](#)), adaptation cannot fully account for the whole range of mismatch effects. It fails to properly account for MMRs evoked by omissions, stimuli of lowered amplitude, or alternating stimuli. In the latter paradigm, two stimuli A and B are presented in regular alternation at a fixed interval: ABABA... (Figure 2B). When this regularity is occasionally

violated by stimulus repetition (e.g., AA), an MMR is observed ([Horváth and Winkler, 2004](#); [Strauss et al., 2015](#); [Todorovic and de Lange, 2012](#)), although SSA alone should predict a further decrease in evoked responses to a repeated stimulus. The results are better explained by assuming that this signal reflects the violation of an active expectation of the alternating stimulus. Similar findings in the visual modality indicate that the bulk of what is generally attributed to “repetition suppression” may in fact arise from predictions and their violations ([Egner et al., 2010](#); [Summerfield et al., 2008](#)). Such findings suggest that predictive coding is ubiquitous in various sensory cortices ([Friston, 2005](#)).

[Wacongne et al. \(2012\)](#) proposed a spiking-neuron model of predictive coding and mismatch detection. The model proposes that a cortical region, sensitive to a given aspect of stimulus identity (e.g., auditory pitch), learns a predictive-coding model that uses a temporal window on the recent past to generate anticipations of the future. When exposed to a sequence of events $X_{1:n}$,

the model cortex encodes it by strengthening a set of synapses that connect a representation of the past events $X_{i,t-\tau}$ to the present event $X_{i,t}$. Those synapses capture any regularity in the identity-by-delay matrix that connects past and present stimuli. Thus, the model assumes that what is being encoded at this level is the conditional probability of observing a new event $X_{i,t}$ at a specific time t , given the items $X_{i,t-\tau}$ observed in the recent past. Simulations show that this mechanism can store transition probabilities in sequences such as AAAA or ABAB and can account for major properties of the MMR, including its distribution across cortical layers and dependency on NMDA-receptor mechanisms (Javitt et al., 1996).

The Wacongne et al. (2012) model requires a neurophysiological mechanism of “time-stamp” neurons that emit spikes at a relatively fixed delay after a specific sensory or motor event. Such neurons tuned to a relatively specific temporal interval have been observed in dorsolateral PFC and caudate nucleus (CN) (Jin et al., 2009). Time may also be encoded by monotonically increasing or decreasing firing patterns, as observed in premotor cortex during a learned tapping task (Merchant et al., 2014). Neurons in the cerebellum (Johansson et al., 2014; Ohmae et al., 2013), parietal cortex (Leon and Shadlen, 2003), and hippocampus (Kraus et al., 2013) have also been observed to encode elapsed time. Thus, it is not implausible to suggest that time-stamp neurons are available, either locally or through basal ganglia or cerebellar loops, to virtually any cortical area. Theory and simulations show that, even if neurons exhibit noisy chaotic dynamics, the presence of partially reproducible trajectories, supplemented by a learning mechanism, suffices to encode temporal sequences in a stable manner, subject to Weber’s law (Jin et al., 2009; Laje and Buonomano, 2013).

In summary, at our first level of sequence representation, sequences are stored by keeping a record of the transitions between events and their approximate timing. At this level, the sequence code is unconscious, shallow, item specific, and temporally detailed. The signatures of this code are (1) precisely timed neuronal firing, (2) subject to Weber’s law, and (3) with the emission of mismatch or omission responses when the expected event sequence is violated. This mechanism seems to be duplicated in many cortical and subcortical circuits.

Chunking

When a sequence of events recurs, those events may be grouped together as a “chunk” and stored as a single unit. A “chunk” can be defined as a group of contiguous items that frequently recurs as a whole and that are therefore usefully encoded as a single group by the nervous system. An example is provided by a word such as “caramel,” which groups several successive syllables or phonemes into a single unit (in this example, it is important that the word “caramel” comprises a single morpheme: as discussed later, complex polymorphemic words are thought to involve nested trees structures, e.g. “repainted” = {{re,paint},ed}). Note that there is nothing in our definition of a chunk which is specific to sequences or to auditory stimuli. On the contrary, it is likely that many brain areas independently encode recurring chunks in multiple domains: frequent melodies and words are detected by auditory areas; recurrent series of motor actions are compiled in cortical and subcortical motor areas; groups of visual features

are detected as familiar faces, objects, and places in specialized areas of the visual system, etc.

There is evidence that temporal chunking occurs in a simple variant of the auditory oddball paradigm called the “local-global paradigm” (Figure 2C) (Bekinschtein et al., 2009). In this variant, the repeated sounds are no longer presented as a continuous stream, but as short sequences, thus offering the opportunity to store those sequences as “chunks.” In the critical condition, subjects are presented with a sequence of the *aaaab* type, where the first four tones are identical while the fifth sound differs. In human EEG and intracranial recordings, the deviancy of the last sound initially leads to a series of potentials: first a MMR, then a late surprising-elicited P3b wave. Repeatedly hearing the same *aaaab* sequence, however, leads to a dramatic reduction of those components. Interestingly, the MMR remains, confirming that this component reflects a shallow and automatic response to local transition probabilities that does not take global predictability into account. The P3b, however, disappears, suggesting that, at some level, the “global” sequence is memorized, including an expectation that the fifth tone will be different. This interpretation can be confirmed by presenting rare deviants consisting of five identical tones *aaaaa*: the P3b wave immediately reappears, indicating that even a monotonic sequence can be surprising if it violates prior expectations. The findings indicate the existence of a distinct level of global representation where the whole sequence is memorized. fMRI, ERP, MEG and intracranial recordings indicate that this mechanism is available to human adults (Bekinschtein et al., 2009; El Karoui et al., 2014; King et al., 2013, 2014; Strauss et al., 2015; Wacongne et al., 2011), 3-month-old infants (Basirat et al., 2014), and macaque monkeys (Uhrig et al., 2014).

The signatures of this mechanism distinguish it sharply from simple transition-probability learning. First, it is slower: early (~100–200 ms) event-related potentials such as the MMR are transient and primarily affected by local transition probabilities, and it typically takes an additional 100–200 ms before effects of global context arise in a sustained and stable manner (King et al., 2014). Second, the effect is no longer automatic: it essentially vanishes when subjects are distracted by a visual task, asleep, or unconscious (Bekinschtein et al., 2009; Strauss et al., 2015). Third, it arises from a distributed set of regions, including secondary auditory cortices; superior temporal sulcus; inferior frontal gyrus (IFG); and dorsolateral prefrontal, intraparietal, anterior, and posterior cingulate cortices (Uhrig et al., 2014).

The local-global paradigm does not provide a pure test of sequence chunking, because it may engage higher-order ordinal, algebraic, and tree learning mechanisms (as further discussed below). A simpler paradigm, however, has been used to specifically demonstrate chunking of a speech stream and its central role in word acquisition (Saffran et al., 1996). In this paradigm, subjects are exposed to a continuous auditory stream of syllables such as “tokibugikobagopilagikoba...” (Figure 1). Inconspicuously, the syllables form recurring groups of three syllables analogous to words (e.g., “gikoba” and “tokibu”). After being exposed to such a stream for 2 min, even 8-month-old babies show evidence of chunking: in a post-test, they attend less to three-syllable stimuli that correspond to the original words

than to otherwise similar stimuli consisting either in a random arrangement of the same syllables, or in the end of a word followed by the beginning of another word (Saffran et al., 1996). Event-related potentials in adults indicate that, after training, a left-lateralized rhythm emerges at one-third of the original syllable frequency, corresponding to the word presentation frequency, suggesting that the stream is now parsed into three-syllable chunks (Buiatti et al., 2009). fMRI studies suggest that such chunking arises from the left planum temporale (PT) and/or left IFG (Karuza et al., 2013; Tremblay et al., 2013).

The Saffran et al. (1996) stimuli had a higher transition probability of syllables within a word than between words and could therefore be accounted for by transition-probability learning. Graf Estes et al. (2007), however, provided direct evidence that the three-syllable chunks were extracted as distinct units. In this experiment, 17-month-old babies who had been exposed to sequences of three-syllable words later showed evidence of a fast acquisition of semantic association of those words with visual objects, compared to non-exposed three-syllable strings. Evidently, the extracted chunks became available as lexical entries in the child's mental lexicon. Saffran and Wilson, (2003) likewise demonstrated that the chunked words could enter into a higher-order learning process: 12-month-old infants acquired the transition probabilities between those words, which implies that they were coded as independent entities. Indeed, chunking is thought to be one of the main mechanisms by which children identify words in the continuous speech stream (Hay et al., 2011).

A similar mechanism seems to operate in the visual and motor domains, allowing human adults (Orbán et al., 2008), infants (Bulf et al., 2011), and baboons (Minier et al., 2015) to chunk spatial or temporal groups of visual shapes, locations, or actions. Orbán et al. (2008) describe a Bayesian learning mechanism capable of discovering the minimal set of chunks that captures a regular visual scene, and they provide strong behavioral evidence that such a mechanism may be used by humans. Interestingly, Bor et al. (2003) observed intense bilateral IFG activation whenever a sequence of visual locations could be chunked into subsets forming familiar shapes such as a square, and similar observations were made in a related paradigm (Schapiro et al., 2013).

At the single-cell level, chunking has been primarily studied in the context of the acquisition of motor habits (Fujii and Graybiel, 2003; Graybiel, 1998; Jin et al., 2014; Smith and Graybiel, 2013). During the acquisition of a familiar motor sequence, single neurons in PFC and basal ganglia dramatically change their firing pattern: they initially emit a reproducible burst of spikes to every action in the sequence, but eventually end up firing only to the first and last items, thus signaling chunk boundaries (Figure 3). These data thus confirm the neurophysiological reality of chunking but do not yet indicate whether and how neurons are allocated to the encoding of entire chunks. Human fMRI suggests that, in action planning, chunking may be hierarchical, leading to the formation of "chunks of chunks" involving increasingly anterior prefrontal areas (Koechlin and Jubault, 2006).

In summary, our second sequence representation mechanism is one by which frequently co-occurring items are grouped together as a "chunk," which can then be manipulated as a sin-

gle higher-level unit. The precise neural mechanisms of chunk formation remain unknown, and are probably widely distributed in multiple cortical areas, but include inferior frontal gyrus and superior temporal regions for human auditory chunking and frontal/basal ganglia loops for motor chunking.

Ordinal Knowledge

The transition-timing and chunking mechanisms discussed so far can only recognize specific melodies or words and their violations. Indeed, the corresponding experimental paradigms typically do not evaluate the degree of abstraction of the sequence representations involved, and whether abstract features common to multiple sequences are extracted. Other paradigms, however, offer direct evidence that human and non-human primates extract at least one such feature: ordinal knowledge, i.e., a representation of which event comes first, second, or third, independently of when these events arise. The ordinal system abstracts away from specific timing information and encodes only relative temporal order. Possessing distinct mechanisms for timing and ordinal knowledge is clearly useful: there are many natural situations in which event timing is fixed and predictable (e.g., when an object falls from a certain height), but there are equally numerous situations in which it is only possible to predict that something will happen, or even how many events will occur, without knowing when (e.g., when one or more predators hide behind a tree). Prefrontal and parietal circuits for working memory and number may have evolved to bridge temporal gaps by predicting what and how many while leaving open when.

In humans, ordinal knowledge is typically studied by requesting to memorize a list of items and, after a short delay, to reproduce it in the same order. The classic "primacy" and "recency" effects indicate, respectively, that the first and last items are better remembered, suggesting that order information is used to organize the list. Importantly, the same conclusion was reached when this list learning paradigm was adapted to macaque monkeys (Chen et al., 1997; Orlov et al., 2000; Terrace et al., 2003). Strikingly, after learning two separate lists $a_1 a_2 a_3$ and $b_1 b_2 b_3$, monkeys were likely to confuse items with the same ordinal position, for instance erroneously reporting $a_1 b_2 a_3$ (much more frequently than, say, $a_1 b_3 a_3$). Thus, ordinal position is a salient parameter by which monkeys and humans organize incoming sequences. Both species also impose a global ordinal representation to items that are merely presented as pairs: for instance, presentation of pairs $a \rightarrow b$; $b \rightarrow c$; $c \rightarrow d$; $d \rightarrow e$, in random order, leads to the implicit knowledge that a is first, b second, and e last, and this ordinal knowledge transfers to a sequence task (Jensen et al., 2013; Merritt and Terrace, 2011).

Neurophysiological recordings in dorsolateral prefrontal and intraparietal cortex have revealed a neural code for ordinal number (Figure 4). In monkeys trained to remember a series of three shapes and reproduce it, Ninokura et al. (2004) observed that as many as 44% of prefrontal neurons were tuned to serial position in the list, thus responding identically to entirely different shapes, solely on the basis of order information (Figure 4). In monkeys trained to identify the total number in a series of identical sounds or light flashes, with variable duration and timing, Nieder and

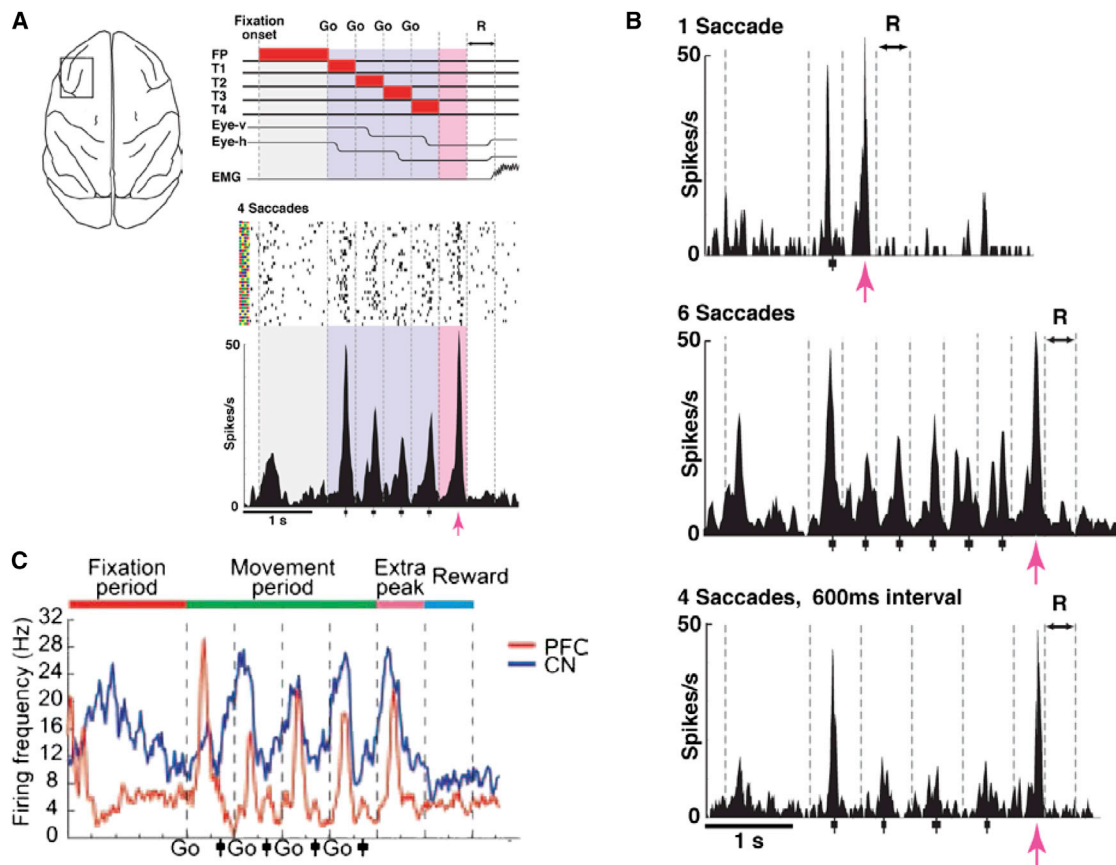


Figure 3. Evidence for Sequence Chunking in Monkey PFC and Basal Ganglia

(A) A PFC neuron exhibits enhanced firing at the start and end of a four-target sequential saccade task. Raster plots and peristimulus time histograms are shown for 35 different trials. Pink denotes the 400-ms window during which the end response is observed (“extra peak” denoted by a pink arrow). (B) Invariance of the start and end responses of the PFC neuron shown in (A) to different sequences: blocks with a single saccade (top), six saccades (middle), or four saccades with a longer inter-target interval (bottom). (C) Peri-event time histograms of the activity of two neurons from PFC and CN during the four-saccade task (after Fujii and Graybiel, 2003).

colleagues observed that many PFC and IPS neurons were tuned to a specific ordinal number and discharged maximally to one of four possible serial positions, with the characteristic signature of Weber’s law (Nieder, 2012; Nieder et al., 2006). A significant proportion of such ordinal neurons, particularly in PFC, were tuned to the same number of simultaneously presented visual items, confirming that they were encoding an abstract parameter of number. In supplementary and presupplementary areas, neurons may also signal the number of remaining movements needed before a reward is obtained (Sohn and Lee, 2007). Order-sensitive neurons are present in a broadly distributed set of areas including the supplementary motor area (SMA), presupplementary motor area (pre-SMA), supplementary eye field (SEF), and dorsolateral PFC (Berdyeva and Olson, 2010). Time- and order-sensitive neurons appear to be intermingled in the medial premotor cortex (Crowe et al., 2014).

While the above experiments involved highly trained animals, even untrained animals are sensitive to the numerical dimension of sequences. Experiments with untrained animals have primarily tested knowledge of the total number of items in a

sequence, not the ordinal position of each item, but it can be safely assumed that the two properties are intimately related, as knowledge of total number implies keeping track of how it evolves with each new item. Behaviorally, untrained or even newborn animals have been found to compare two numbers (Haun et al., 2010; McComb et al., 1994). A spontaneous capacity for matching the total numerosity of visual displays and auditory sequences has also been observed in newborn babies (Izard et al., 2009). When toddlers watch a puppet and are asked to imitate it, they spontaneously reproduce the approximate number of actions, again exhibiting Weber’s law (Figure 3D) (Sella et al., 2015). Neurophysiologically, number-tuned neurons have been recorded in PFC and IPS of untrained macaque monkeys (Viswanathan and Nieder, 2013), and spontaneous responses to number have been observed with fMRI (Wang et al., 2015). Thus, number is a very salient parameter that many species spontaneously take into account as they encode an incoming sequence.

While evidence for an abstract numerical code is overwhelming, an interesting question is how this abstract knowledge is applied to a specific sequence—how do we remember

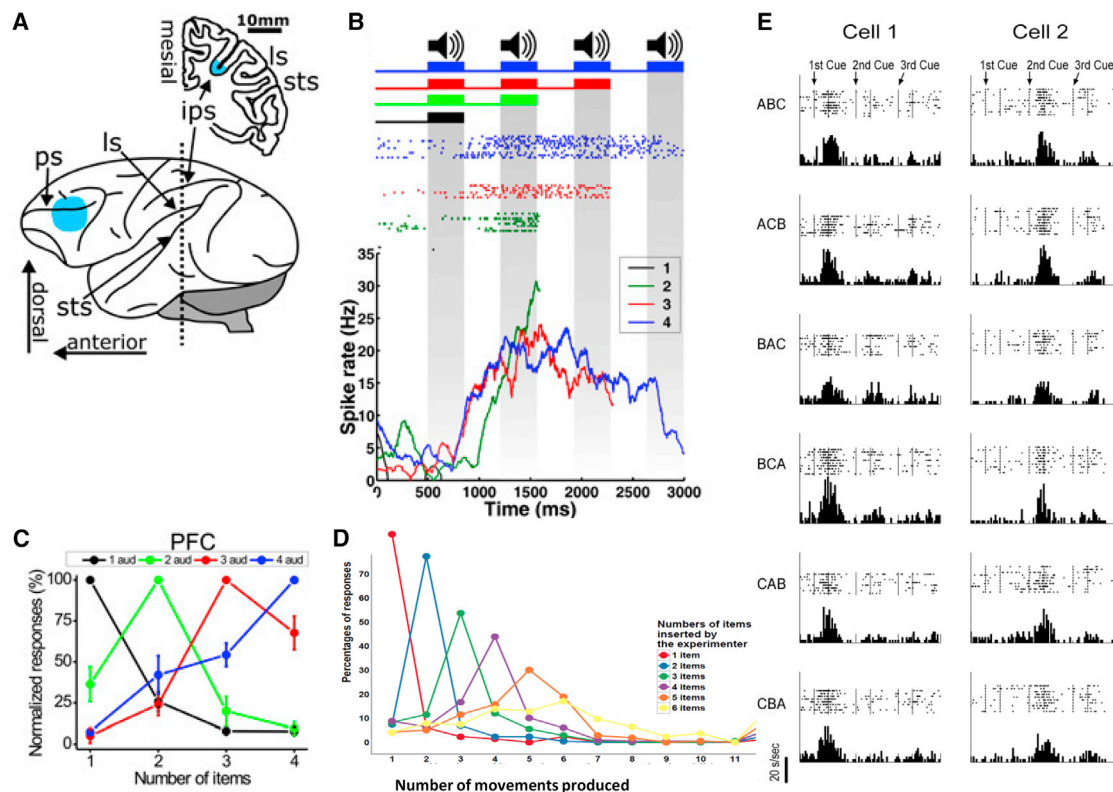


Figure 4. Representation of ordinal numerical information.

(A–C) Monkey neurons selective to ordinal number.

(A) Recording sites in PFC and IPS where such cells are found.

(B) (top) illustrates the temporal succession of individual tones represented by each square pulse, and the corresponding discharges of a cell whose firing peaks just after the second item.

(C) The average tuning curves for neurons tuned to first, second, third, or fourth items (after Nieder, 2012; Figure 3).

(D) Evidence that young children spontaneously use number knowledge when matching a specific number of actions (after Sella et al., 2015).

(E) Responses of two prefrontal cells in monkeys trained to remember the order of three visual objects (denoted by letters ABC). Cell 1 responds to the first item, regardless of its identity, while cell 2 responds to the second item (after Ninokura et al., 2004; Figure 2).

that, in the sequence of meaningless syllables “ba ki do,” “ba” was first, “ki” second, and “do” last? Botvinick and Watanabe (2007) have proposed a simple and effective neural code for serial order in working memory. Their model assumes that abstract ordinal information (presumably conveyed by IPS neurons) is conjoined with item-specific information (e.g., conveyed by auditory cortex) through a gain-field mechanism: prefrontal neurons would be tuned to the product of those two variables, thus exhibiting a joint preference for a specific item and a specific ordinal rank. Simulations demonstrated that such a coding scheme, combined with Weber’s law, could reproduce the main behavioral properties of working memory for ordered sequences, including primacy and recency effects as well as confusions between consecutive or similar items (Botvinick and Watanabe, 2007). Remarkably, conjunctive cells jointly sensitive to ordinal information and to stimulus identity have been reproducibly recorded in PFC, both for motor and for sensory sequences (Barone and Joseph, 1989; Inoue and Mikami, 2006). Gain fields are a very natural mechanism for neuronal coding, which has been observed in other contexts (Andersen et al., 1985). As further discussed below, the concept of neurons jointly sensitive to the

product of a specific identity or “filler” and to its abstract “role” (here ordinal position) may go a long way towards explaining the more abstract sense of syntax in humans (Smolensky and Legendre, 2006).

In summary, at our third level of sequence representation, a sequence is encoded by specifying which item comes first, second, third, etc. The signatures of this representation are primacy and recency effects, Weber’s law for number, and neurons tuned to ordinal number, either abstractly or in conjunction with specific item features.

Algebraic Patterns

The levels of representation that we considered so far all involve storing a series of specific items. We now consider the mental representation of more abstract schemas called “algebraic patterns” that capture the relationships between successive stimuli or stimulus categories. Consider, for instance, a 7-month-old baby listening to a sequence of syllables such as “totobu,” “mimika,” “paparo,” etc. Even a few minutes of such exposure appears sufficient for the baby to recognize that all such sequences obey a similar pattern that may be denoted as AAB (i.e., two

identical syllables followed by a different one). Thus, the baby is sensitive to the identity relation between successive sounds, irrespective of their specific identity. When this pattern is violated, for instance by an ABB item, the baby reacts with longer looking times, indicating that the change was perceived (Marcus, 2001; Marcus et al., 1999). Such evidence implies that the babies categorized the incoming syllables using concepts of same/different and grasped that the first two syllables were always identical.

Marcus introduced the term “algebraic pattern” to refer to an abstract schema (e.g., AAB) that captures the regularities behind several specific sequences (e.g., aab, ccd, etc). Following Marcus’ seminal work, research has concentrated on algebraic patterns defined by identity relationships (Endress et al., 2009; Marcus, 2001; Marcus et al., 1999, 2007). Here, we propose to apply this term whenever there is evidence that an input sequence is internally coded by a corresponding sequence of abstract relationships, concepts, or categories. For instance, there is evidence that infants perceive a sequence of increasing numbers as analogous to a sequence of lines of increasing length (de Hevia and Spelke, 2010)—this implies that infants possess an algebraic pattern for increasing size, which can be denoted as $A < B < C$.

Rudiments of the algebraic level of representation seem to be available to non-human primates (Saffran et al., 2008; Shima et al., 2007; Wang et al., 2015). For instance, Shima et al. (2007) trained monkeys to perform series of four actions on a lever (e.g., for a short period, they repeatedly did “push push pull pull”) (Figure 5). Remarkably, some PFC neurons were tuned to the abstract algebraic pattern underlying the motor sequence: some fired to repetition (AAAA), others to alternation (ABAB), and yet others to paired sequences (AABB), while totally disregarding the specific actions (push, pull, or turn).

Our laboratory used fMRI to explore macaque monkeys’ competence for algebraic patterns in the auditory domain, based on a variant of the local/global paradigm (Wang et al., 2015). In short blocks, naive macaque monkeys were exposed to sequences of tones varying in pitch, tempo, and duration, yet with a constant algebraic pattern (either AAAB or AAAA). Using fMRI, we then tested for brain responses to violations affecting the total number of items (e.g., going from four sounds to two sounds or to six sounds), the sound-repetition pattern (going from AAAA to AAAB or vice-versa), or both (e.g., going from AAAA to AAAAAB). Monkeys showed responses to both number (in intraparietal cortex, anterior cingulate and SMA) and sound-repetition pattern (in basal ganglia, ventral inferior prefrontal, and temporal cortex). Those findings confirm that naive non-human primates are capable of representing the abstract numerical and algebraic patterns of sequences (Nieder, 2012; Shima et al., 2007). Many other species such as corvids (Veit and Nieder, 2013) and bees (Avarguès-Weber et al., 2012; Giurfa et al., 2001) may represent abstract relationships such as same-different, above-below, and right-left.

This is not to say, of course, that such abstract knowledge is identically encoded in human and non-human brains. When comparing fMRI activations in humans and macaques exposed to identical stimuli exemplifying a simple algebraic pattern (e.g., AAAB, as explained above), we observed a strik-

ing difference (Wang et al., 2015): in monkeys, disjoint cortical sites responded to violations in number and in sound-repetition pattern, while in humans, those changes elicited overlapping and precisely correlated responses at the same sites in inferior frontal and posterior superior temporal cortex—sites that also happened to be activated during linguistic sentence processing. Monkeys did show ventrolateral PFC responses to pattern changes, but there was no sensitivity to number at this site. On this basis, we speculated that human language areas exhibit a superior ability to assemble abstract sequence information into a single unified “expression” (e.g., “3 identical sounds, then a different one”); monkeys, on the other hand, may only possess disjoint knowledge of number (“four sounds in total”) and identity relationships (“one item is different”) (Wang et al., 2015).

The reason why algebraic patterns are of great interest, indeed, is that they may provide a model domain for the types of patterns present in language, music, mathematic, and other domains of singular competence in the human species. For instance, each human language is characterized by specific “phonotactic” rules (i.e., a specification of which phoneme sequences are acceptable). For instance, the English language accepts syllables with consonant clusters (CC or CCC structures; e.g., /pr/ in “price” or “spl” in /split/), while Japanese does not. Human infants quickly learn such phonotactic rules (Jusczyk et al., 1994). At a higher grammatical levels, young children learn abstract patterns such as “determiners precede nouns” (in English) and “determiners agree with nouns in gender and number” (in French), and they quickly generalize them to novel instances (e.g., Bernal et al., 2010; Melançon and Shi, 2015). It is not clear that such rules are available to any non-human species. For instance, even a highly trained chimpanzee, after learning to use sign language and to produce some combinations of signs, still lacked the combinatorial productivity evident in human children’s very first linguistic productions (Yang, 2013).

In summary, our fourth level of sequence representation is characterized by the capacity to abstract away from the specific identity and timing of the sequence items and to grasp their generic underlying pattern. The main signatures of this level are a behavioral capacity to generalize to novel items and a neurophysiological tuning of PFC neurons to abstract patterns. There is, to the best of our knowledge, no accepted neural network mechanism by which such an abstract encoding is achieved, as all current neural network proposals seem too limited to account for human abstract rule-extraction abilities (Hadley, 2009; Marcus, 2001).

Nested Tree Structures

Do transition-timing, chunking, numerical, and algebraic patterns suffice to account for the most advanced form of sequence processing, namely language? Most linguists believe that human language abilities imply an additional competence for nested tree structures (Chomsky, 1956). Consider for instance the phrase “black taxi driver.” It can be parsed as “black [taxi driver]” or as “[black taxi] driver.” Such an ambiguity would be impossible to express in a system comprising only “flat” sequence structures such as transition probabilities between

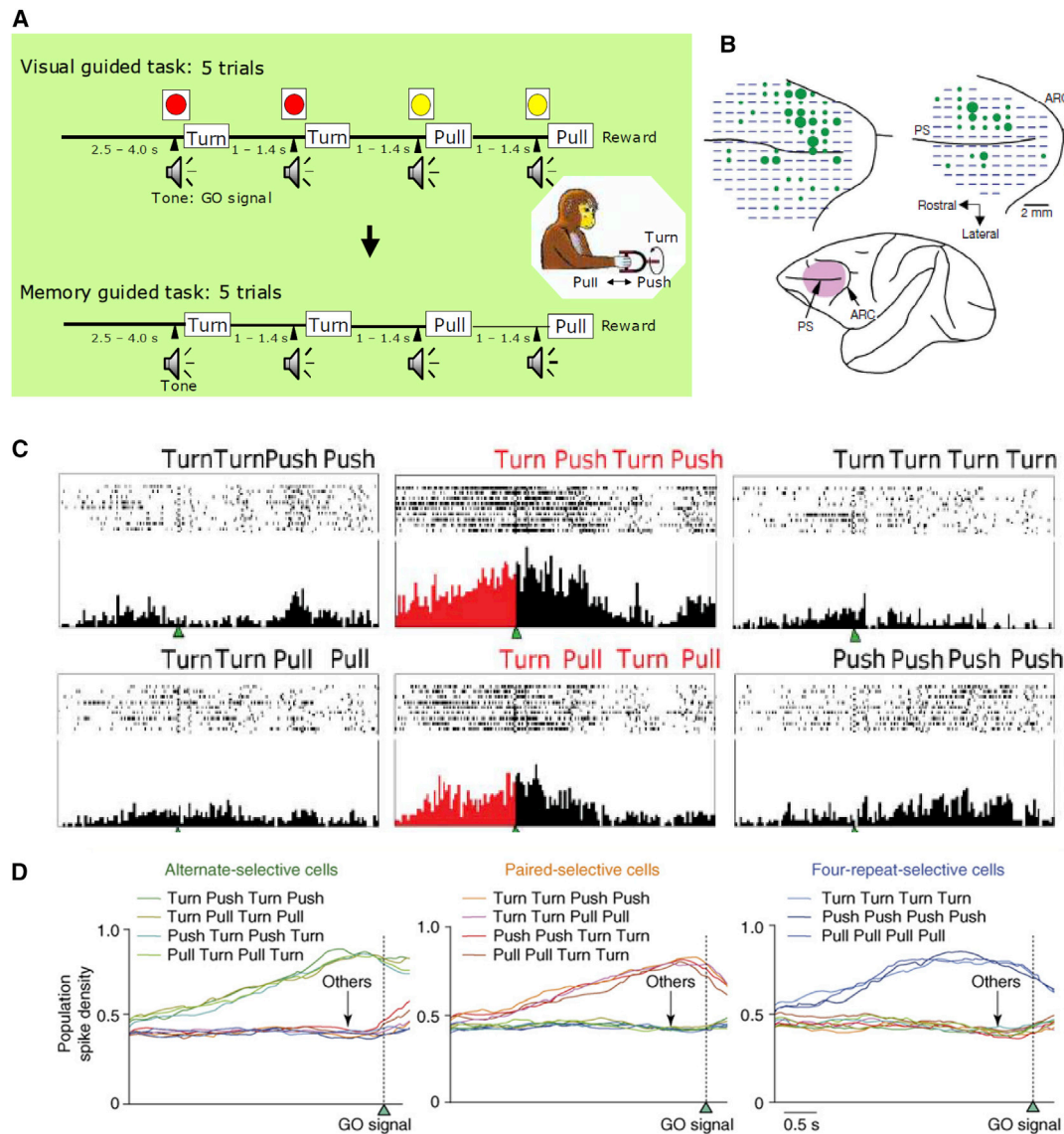


Figure 5. Neurophysiological Evidence that Monkey Prefrontal Neurons Encode the Algebraic Patterns Underlying Motor Sequences

(A) Diagram illustrating the temporal sequences of events in a behavioral task requiring the categorization of multiple sequences. Initially, a correct sequence of four movements is cued with visual stimuli. Monkeys are trained under visual guidance for five trials and then required to perform the same sequence of movements from memory.

(B) Single-unit activity was recorded in monkey PFC. Green dots denote the density of pattern-selective cells.

(C) Activity of a prefrontal cell selective to the "alternate" ABAB pattern during motor planning, just prior to sequence reproduction.

(D) Time courses of population responses for PFC cells tuned to three different patterns (left: alternate ABAB pattern; middle: paired AAB pattern; right: repeat AAAA pattern).

words or even between categories of words. Rather, such relationships require substitution rules such as NP \rightarrow A NP (a noun phrase can comprise an adjective followed by a noun phrase) and NP \rightarrow NP N (a noun phrase can be formed by compounding a noun phrase with a noun). Once such rules are available, their nested application can generate the appropriate structures: [A [N N]] or [[A N] N].

Linguists call "constituents" any of the nested blocks of words that belong together in a sentence—thus, "black taxi driver" is a constituent, itself comprising "black taxi" and

"taxi driver" as two potential sub-constituents. A great variety of linguistic observations support the notion that constituent structures are a fundamental building block of language (Sporich et al., 2013). Thus, ellipsis and substitution indicate that certain groups of words form constituents because they can be deleted and/or replaced by a shorter word—consider for instance "he [drove [to [this [big house]]]]" \rightarrow "he drove to this one," "he drove to it," "he drove there," "he did." Coordination indicates that constituents of the same type (e.g., noun phrases) can be combined by "and," as in "[my [big dog]] and [I]."

Constituents may also move as a unit, for instance when forming a cleft sentence: “it is [to this house] that he drove.” The linguistic rules that govern such movements (called “wh-movement” because it is used to form questions in English) operate over tree structures rather than over mere word sequences. Tree structures are also indispensable to account for agreement (e.g., the link between subject and verb) and binding (the link between a pronoun and its antecedent noun). For instance, in the sentence “The cars [that pass this truck] are red,” the plural subject (cars) agrees with the verb (are), irrespective of the intervention of an arbitrarily large relative sentence including another noun in the singular (truck). Such “long-distance dependencies,” which characterize languages, can only be formulated by assuming that the underlying mental representation is a tree structure that directly links the words “cars” and “are” in spite of their temporal separation. Conversely, note how the above sentence, although obviously grammatical, contains the superficially ungrammatical sequence “this truck are red” (which is not a constituent, is barely detectable, and plays no role in syntactic operations). This and many other examples indicate that transition probability, ordinal, chunking, or pattern-detection mechanisms that are only sensitive to “flat” local order cannot capture the grammaticality of constituent structures—nested trees are required.

In his recent “minimalist” research program, Noam Chomsky has championed the idea that a single mental operation, called “merge,” underlies the formation of tree structures in all of the world’s languages. Merge is a hypothetical operation that forms a minimal binary tree by taking two mental objects X , Y and creating a new object $Z = \{X, Y\}$, the pair formed by X and Y (Chomsky, 2013). The resulting object Z can itself enter into further merge operations, thus creating extended tree structures—for instance the noun phrase “very happy man” can result from two nested merges $\{\{\text{very, happy}\}, \text{man}\}$. Note that the outcome of merge is postulated to be an *unordered* pair, and thus this view of language predicts that the internal representation of syntax abstracts away from the temporality of the word sequence: at a deep syntactic level, there is no temporal or ordinal information, only structure. Various properties of word order, including movement, would be “traced back to the necessity for natural language to organize words in linear order at the interface with the perceptual-articulatory module” (Moro, 1997). In other words, linearization into an ordered sequence would be imposed during sentence production, but the inner representations of language would be unordered binary trees, thus explaining the primacy of constituent structures and long-distance dependencies over temporal proximity in various language phenomena.

Chomsky and many other linguists postulate that even simple sentences are represented by long and complex tree structures. Marcus (2013) argues that this aspect is implausible and difficult to implement within the constraints of human memory and proposes instead that “mental representations of sentences [is] realized via a system of overlapping but incompletely bound sub-trees or treelets that are unified only in transitory and imperfect ways.” Each dependency relationship, for instance, between a verb and its object, an auxiliary and a verb, or a determiner and a noun, would be represented by a small tree or “treelet,”

but there would not any overarching structure, just a list of small trees that would need to be re-explored whenever a specific question or mental transformation is required. A similar proposal, “tree-adjoining grammar,” also assumes that complex phrase structures are built out of a small set of elementary tree structures (Kroch and Joshi, 1985).

While the minimalist program and the optimal linguistic formalism remain hotly debated issues (e.g., Pinker and Jackendoff, 2005), there is very little doubt among linguists that trees comprising nested constituents will have to be an essential component of any future theory of language. Furthermore, although the neural codes underlying linguistic trees remain unknown, the brain areas involved are beginning to be localized. Over a century of lesion studies and 20 years of brain imaging have delineated a left-lateralized network of areas for sentence comprehension, involving primarily the left superior temporal sulcus (STS), middle temporal gyrus (MTG), temporal poles (TP), temporo-parietal junction (TPJ), and especially the left inferior frontal gyrus or “Broca’s area” (including Brodmann areas [BA] 44 and 45) (Friederici, 2011; Mazoyer et al., 1993). The left putamen (Pallier et al., 2011) and left CN are also frequently involved (Monti et al., 2009; Moro et al., 2001). Specific studies of syntax have homed in on a narrower circuit involving left putamen, STS, and IFG, interconnected by the arcuate fascicle and the extreme capsule (Rolheiser et al., 2011; Saur et al., 2010; Tyler et al., 2011). Activation of this network correlates tightly with predictors of syntactic complexity derived from formal linguistics, including nested constituent structures, wh-movement, and syntactic ambiguity (Brennan et al., 2012; Musso et al., 2003; Ben-Shachar et al., 2004; Shetreet and Friedmann, 2014; Shetreet et al., 2009; Snijders et al., 2009). For instance, Musso et al. (2003) trained subjects to transform sentences in what was described to them as a new language (e.g., Japanese). Unbeknownst to them, the learned transformations operated either according to “real” linguistic rules based on nested trees (e.g., forming a question by moving the questioned constituent to the front of the sentence; i.e., wh-movement) or according to “unreal” rules based on superficial sequence features such as the ordinal position of words (e.g., forming a question by inverting the linear order of all words). Although subjects learned both types of rules, only real rules led to a progressive increase of fMRI activation in IFG (Brodmann’s area 45). Likewise, this region increases its activation whenever constituents are moved out of their normal location by wh-movement (Ben-Shachar et al., 2004; Shetreet and Friedmann, 2014).

Additional cues as to how constituent structures are encoded were recently obtained from a parametric study of constituent size (Pallier et al., 2011). Subjects were exposed to sequences of 12 words while parametrically varying how many of those words could be bound into nested constituents (Figure 6). Subjects read or heard sequences of words or pseudowords forming either no constituents at all (C01, word-list condition), word pairs (C02), triplets (C03), quadruplets (C04), sextuplets (C06), or an entire sentence of 12 words (C12). The reasoning was that an increasingly larger number of nested constituents would lead to an increase in the activation of relevant brain areas. Indeed, a chain of areas in the left STS and IFG showed increasing

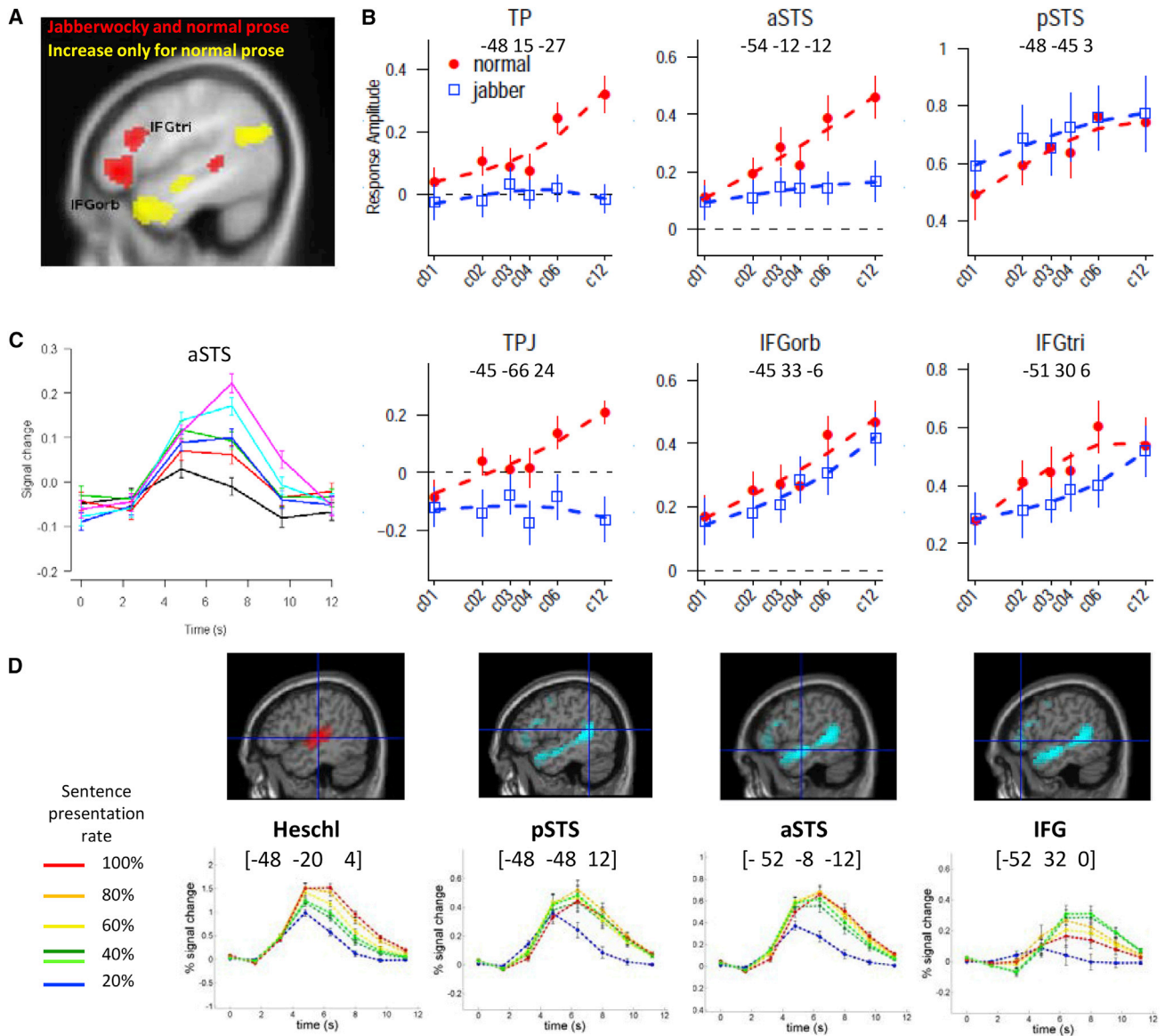


Figure 6. Parametric Properties of Human Brain Regions Responding to the Constituent Structure of Language

(A–C) Variation of constituent size. Although the stimuli always comprise a fixed total number of words (12), the regions in (A) show increasingly higher activation as a function of the number of words that can be integrated into a linguistic constituent. Their response profile as a function of constituent size appears in (B). Regions in red respond to phrases with both words and pseudowords, suggesting a sensitivity to syntax, while regions in yellow respond only to meaningful phrases with real words, suggesting a sensitivity to semantics (from [Pallier et al., 2011](#)). In most regions, the fMRI response is also increasingly delayed for larger constituents (C).

(D) Response of this language processing network (light blue) as well as auditory cortex (red) to sentences with various degrees of acceleration (from 100% down to 20% of their original duration). Auditory activation is monotonically accelerated, but language activation is not: it is constant (STS) or even increases with compression (IFG) up to a critical value, somewhere between 20% and 40% of original sentence duration, where activation suddenly collapses (from [Vagharchakian et al., 2012](#)). These findings suggest that language areas host a slow process of constituent structure formation.

fMRI activation with the logarithm of the number of words that could be entered in a constituent. Remarkably, when the stimuli were “delexicalized” by substituting all content words with meaningless pseudowords while maintaining all grammatical words and inflections, a core set of areas in left IFG and pSTS continued to respond identically, suggesting their central role in the construction of abstract syntactic trees. Furthermore, analysis of the phase of the fMRI response indicated that the

build-up of constituents took increasingly longer time for larger structures ([Pallier et al., 2011](#)). During sentence processing, anterior temporal and inferior frontal areas respond much more slowly than primary or secondary auditory areas ([Dehaene-Lambertz et al., 2006](#); [Hasson et al., 2008](#); [Pallier et al., 2011](#)), and experiments with speech compression indicate that this slow processing speed cannot be accelerated beyond a fixed limit ([Figure 6D](#)) ([Vagharchakian et al., 2012](#)).

Overall, fMRI experiments that have manipulated nesting depth (Goucha and Friederici, 2015; Pallier et al., 2011; Shetreet et al., 2009) lead to the tentative suggestion that constituents are constructed through a slow process of progressive word accretion, reflected in slowly increasing brain activity: the innermost words would be bound first (e.g., “black taxi”), then this constituent would itself be bound at the next level (“{black taxi} driver”), and so on. An MEG study found a correlate of this sentence-construction process in a progressive increase of power in the beta and theta bands (Bastiaansen et al., 2010). Unfortunately, this study did not pinpoint the source of these signals, but other studies suggest that left IFG and anterior temporal regions are primarily involved (Brennan et al., 2012; Pallier et al., 2011). While the syntax-related fMRI activation in Pallier et al. (2011), with French stimuli and subjects, was primarily located to Brodmann’s areas 45 and 47, a recent replication and extension in German demonstrated a joint involvement of BA 44 and 45, with BA44 playing a specific role for “pure syntax” once lexical information and derivational morphology were carefully suppressed (Goucha and Friederici, 2015). In Hebrew, when manipulating syntactic embedding depth, Shetreet et al. (2009) only observed BA 45. At present, therefore, the contribution of distinct IFG areas is not fully understood and might even differ according to language. An interesting possibility is that IFG participates in the “merge” (Chomsky, 2013) or “unification” (Snijders et al., 2009) operations needed to assembled tree structures, while different regions of the STS provide the “raw materials” (words, grammatical categories, thematic roles, etc.) to be assembled. Compatible with this idea, recent fMRI suggests that the STS may be subdivided into subregions or “slots” that encode the agent, the verb, and the patient of an action, regardless of temporal order (i.e., coding identically “the dog bites the man” and “the man was bitten by the dog”) (Frankland and Greene, 2015). It is still unclear, however, how this mechanism could create recursive structures such as relative sentences with multiple agents, verbs, or patients.

It is important to acknowledge that the formation of nested trees is by no means unique to language. Rather, a form of combinatorial syntax is also present in motor action, music, or mathematics. Already in the 1970s, behavioral studies by Frank Restle revealed that simple sequences of actions are stored in an abstract compressed form that goes beyond the simple associative chain and requires the postulation of an internal tree structure (Restle, 1970; Restle and Brown, 1970). More recently, fMRI studies have revealed that musical syntax, like language, engages the left IFG and pSTS, but also their right-hemispheric homologs (Koelsch, 2005). Concerning the syntax of written mathematical expressions such as “ $(2 \times 4 + 1) - 3$,” a distinct dorsal network involving fusiform gyrus, parietal cortex and dorsal BA44 is involved (Maruyama et al., 2012; Monti et al., 2012; Nakai and Sakai, 2014). Even within language, distinct parallel IFG/STS circuits contribute to combinatorial structures for morphology, syntax, and semantics (Xiang et al., 2009). As noted earlier, a single word such as “repainted” may comprise a tree of morphemes. Even a single syllable is thought to be internally represented in humans by a hierarchical tree structure that distinguishes onset and rhyme, and within the rhyme, vocalic nucleus and coda, themselves comprising multiple phonemes (\ll plant

$\gg = \{\{p,l\},\{a,\{n,t\}\}\}$). What these examples show is that, in humans, tree structures are ubiquitous: the human brain may exhibit a specific \ll dendrophilia \gg (Fitch, 2014) (i.e., a propensity to impose tree structures to virtually any domain of perception, action, or thought). One may formulate the tentative hypothesis that multiple parallel IFG areas, with distinct cytoarchitectony (Amunts et al., 2010) and connectivity to posterior cortices (Clos et al., 2013; Neubert et al., 2014) and basal ganglia, may be involved in the construction of tree structures in different domains.

Artificial stimuli obeying simpler regularities than those of human languages have been used to determine the minimal properties that engage Broca’s area. This approach has demonstrated that BA44 can be activated by meaningless sequences of syllables that mimic the nested structures of language (Bahlmann et al., 2008; Fitch and Friederici, 2012; Friederici et al., 2006). For instance, Bahlmann et al. (2008) had subjects learn sequences of meaningless syllables such as “ge bi di tu po ko” in which the items were paired according either to temporal adjacency ([ge bi] [di tu] [po ko]) or to nested rules (ge [bi [di tu] po] ko). Although all stimuli were virtually identical, nested structures led to increased activation in left IFG (BA 44) as well as additional sites in right premotor cortex, bilateral anterior insula and basal ganglia. As described above, the human IFG even responds to simple sequences of tones with a minimal algebraic structure such as AAAB (three identical tones, then a different one) (Wang et al., 2015). Even when submitted to a random sequence of stimuli, the IFG responds to violations in a chance pattern of repetition or alternation (Huettel et al., 2002). The evidence therefore suggests that the human IFG exhibits a remarkable sensitivity to detect all sorts of regularities, from simple repetitions and alternations to complex nested structures.

A fascinating question for future research is whether the capacity to form nested representations of sequences is unique to human brains. Undoubtedly, non-human primates can learn to detect structures within auditory sequences (Wilson et al., 2013) and may even represent certain abstract features of such sequences such as their number or algebraic pattern (Nieder, 2012; Nieder et al., 2006; Wang et al., 2015). However, those competences need not imply that nested rules are involved: specific experiments are needed to probe this representational level. An early behavioral experiment suggested that tamarin monkeys were unable to grasp center-embedded nested tree structures of the type $A[A[AB]B]B$, while they succeeded in learning non-nested pairwise structures $[AB] [AB] [AB]$ (Fitch and Hauser, 2004). More recently, some researchers have claimed that macaque monkeys and even songbirds could encode such nested structures (Abe and Watanabe, 2011; Gentner et al., 2006). However, those claims have been heavily disputed on the ground that the paradigms failed to include sufficiently abstract generalization items and that the behavior was in fact driven by simpler non-recursive representation of transition probabilities and transition patterns (Beckers et al., 2012; ten Cate and Okanoya, 2012; Fitch and Friederici, 2012; van Heijningen et al., 2009). Instead, a variety of behaviors such as sequence learning, tool use, music, or mathematics suggest that the capacity to acquire and manipulate embedded tree

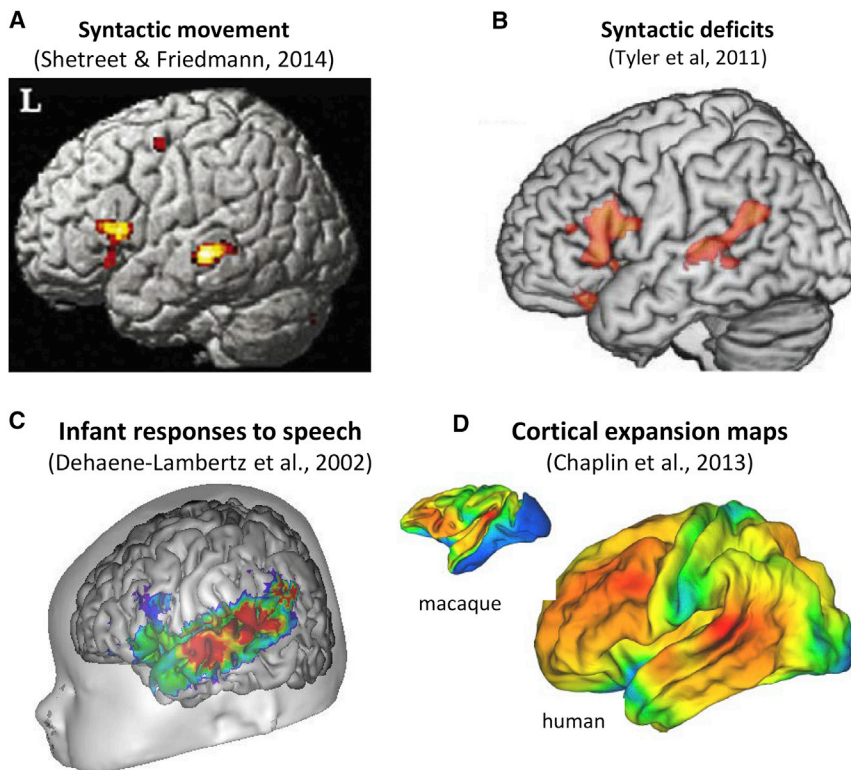


Figure 7. Converging Evidence for a Core Language Circuit and Its Recent Evolution in Humans

(A) Areas activated by syntactic movement, that is, showing a greater response to sentences with unusual word order (object subject verb, after transformation by a linguistic process called wh-movement) than to sentences with usual word order (subject verb object) (from Shetreet and Friedmann, 2014).

(B) Areas whose lesioning correlates with deficits in a syntactic judgment task in patients with left-hemisphere lesions (from Tyler et al., 2011).

(C) Areas activated in two-month infants listening to sentences in their mother tongue (from data in Dehaene-Lambertz et al., 2006).

(D) Areas of recent cortical expansion across the primate lineage (from Chaplin et al., 2013).

structures may be a uniquely human feat (Fitch, 2014; Hauser et al., 2002; Penn et al., 2008).

This conclusion, if validated, would fit well with neuroscientific evidence which indicates that (1) the areas that subserve language, including human IFG, superior temporal cortex, inferior parietal cortex, and underlying white matter have expanded and show greater hemispheric asymmetry in humans (Chaplin et al., 2013; Glasel et al., 2011; Hill et al., 2010; Rilling et al., 2008; Schoenemann et al., 2005) (Figure 7); (2) the IFG shows human-specific functional connectivity to superior temporal (Neubert et al., 2014) and inferior parietal areas (Mantini et al., 2013) that is not observed in macaque monkeys; (3) in 2-month-old human babies, the IFG and STS are already active, at their classical adult location, when they listen to sentences in their mother tongue (Dehaene-Lambertz et al., 2002, 2006); and (4) early pathologies, such as thiamine deprivation in infancy, can dramatically impair the processing of syntactic movement and embedding (Fattal et al., 2011). While such evidence clearly suggests an early and genetically determined channeling of language inputs to reproducible areas of the human brain, it does not however conclusively demonstrate that nested structures are the key difference between human and non-human brains—this is an important area for further research.

A second, essential limit to our current knowledge concerns the neurophysiological representation of linguistic structures. Although some neural-network simulations have claimed to capture basic aspects of language (e.g., Dominey et al., 2009; Elman, 1990; van der Velde and de Kamps, 2006), they fall short of mimicking the detailed knowledge accrued by linguists (Spor-

tiche et al., 2013). Smolensky and Legendre (2006) and Eliasmith et al. (2012) argue that progress will only be made through the identification of mathematical principles capable of bridging between low-level neurophysiological and high-level linguistic descriptions. Building upon earlier work by Smolensky (1990) and Plate (1995), they propose that tree structures are encoded neurophysiologically by “vector symbolic architectures.”

This concept is an extension of the classical notion that each mental representation can be implemented by a vector of activity over a population of neurons in a given area (i.e., a set of approximate numerical values for the firing rate of each neuron, with some neurons firing at an elevated rate while other remain silent). Evidence for vector codes has been observed in various domains: the direction of a planned movement can be represented by a vector of firing rate over a population of motor neurons (Georgopoulos et al., 1989), color and motion can be jointly represented by superimposed vectors in PFC (Mante et al., 2013), etc. What is unique to “vector symbolic architectures” is the proposal that nested, syntactically organized representations can be encoded by vectors combinations. Specifically, the proposal entails:

1. Assigning vectors of neural activity to each possible sequence item (called a “filler”).
2. Assigning other vectors to the abstract syntactic “roles” that these items play in the sequence (e.g., subject, verb, object).
3. Binding each filler to its role by computing a mathematical operation called “tensor product” or “circular convolution” of those two vectors—simplifying somewhat, these operations essentially involve allocating a new vector to each combination of filler and role through a product-like operation.
4. Summing those products to represent the total set of role-filler relationships in a phrase or constituent.
5. Making the resulting vector sum available for further binding by the same mechanism, thus allowing for the recursive construction of nested structures.

Box 1. Current Status of the Field

- Transition probabilities and timing of sensory and motor sequences can be represented by a great diversity of neural circuits including PFC, basal ganglia, and cerebellum.
- Similar computational principles of prediction and prediction error underlie sequence knowledge in many different domains and sensory modalities.
- Monkeys and humans share a capacity for representing the abstract numerical and sequence structure of temporal sequences.
- Transitions and timing, chunking, ordinal knowledge, algebraic patterns, and nested rules are all quickly available to humans during infancy.
- A reproducible network of cortical and subcortical areas underlies the representation and manipulation of language syntax in humans.

While still unsupported by direct physiological observation, this theory is appealing for several reasons. The proposed mechanism (vector codes that combine according to a multiplicative relationship) is similar to a previous proposal for serial order in working memory, as described earlier (Botvinick and Watanabe, 2007). It predicts the observation of “gain fields” (i.e., neurons tuned to the product of two experimental variables that, as noted above, have indeed been reported for spatial dimensions in posterior parietal cortex) (Andersen et al., 1985) and for binding of ordinal and identity information in PFC (Barone and Joseph, 1989; Inoue and Mikami, 2006). Most importantly, assuming that one of the two vector codes is sparse (i.e., only a small subset of neurons are active), the proposed summation mechanism may explain why brain activation increases monotonically with the number of nested constituents being encoded, as observed experimentally with fMRI (Pallier et al., 2011) and with MEG (Bastiaansen et al., 2010). Thus, the proposal has some face validity and should be further tested using intracranial recordings or other high-resolution methods in humans.

In summary, the formation of nested tree structures, as exemplified by the constituent structure of language, relates to an identified circuit in IFG and pSTS. One signature of this code is an increasingly stronger and delayed activation as the constituents comprise more words. Nevertheless, the exact neural code that underlies those effects remains elusive.

CONCLUSION

We have reviewed the evidence for a minimum of five levels of sequence knowledge: transition-timing, chunking, ordinal, algebraic, and nested structures (see Box 1). Fragmentary evidence suggests that aspects of the last two of these levels may be available only to humans and may relate to the uniquely human competences for language, music, and mathematics. These representations are currently understudied at the neural level. A focused research program, using dedicated methods to visualize human and non-human primate brain activity at a high level of spatial and temporal resolution, is urgently needed to resolve the mechanisms by which those representations are encoded

Box 2. Future Directions

- How is predictive coding implemented? What, if any, is the specific contribution of cortical layers and subcortical regions to the extraction of sequence structure and the prediction of forthcoming items?
- What are the respective roles of cortical, basal ganglia, and cerebellar circuits to sequence coding?
- How are abstract algebraic patterns encoded in neural circuits?
- How are nested trees and recursive structures encoded in the human brain?
- Is the human brain endowed with early, genetically determined circuits for language processing and extraction of nested rules?
- Do the uniquely human abilities for abstraction stem from a domain-specific system for language processing, or from a broader competence for representing abstract rules, possibly duplicated in several parallel circuits engaging in language, mathematics, music...?

and to determine if vector codes and vector product operations are involved, as predicted by some theories (see Box 2).

The coexistence, in different brain circuits, of multiple systems for sequence learning raises an interesting issue for further research: how does the brain determine what is the best model for a given sequence? Do all systems compete to reduce the prediction error in the sensory input, until one of them effectively manages it to predict it and “blocks” the learning in other systems? Or, on the contrary, do all systems operate independently of each other, each attempting to grasp an aspect of the incoming sequence? The latter possibility is supported by the experimental finding that local transition probabilities (as reflected in the MMN) are extracted independently of the concomitant knowledge of the global sequence (as reflected by the P3 wave) (Bekinschtein et al., 2009; Wacongne et al., 2011). However, specific experiments, putting multiple interpretations of the same sequence in competition with each other, will be needed to clarify this point.

More generally, we hope that our proposed taxonomy may facilitate the identification of minimal stimuli capable of isolating a specific level of representation. In the current literature, confusingly, diverse sequence-learning paradigms are lumped together under the term “artificial grammar learning,” even when the stimuli do not require anything more than the learning of transition probabilities. Careful stimulus design, particularly in the selection of generalization items, will be essential in order to bolster claims of abstract rule learning in animals and humans (Beckers et al., 2012). Whether a given test requires abstract sequence knowledge should be explicitly tested by demonstrating that simpler properties such as sensory adaptation, transition probabilities, or chunking cannot account for it. To this aim, two detailed neurophysiological models are available for simulation, one for transition probabilities and their timing (Wacongne et al., 2012), and the other for ordinal knowledge in working memory (Botvinick and Watanabe, 2007). The field would greatly benefit from the development of similar models for chunking, algebraic

patterns, and nested tree structures (e.g., see Frank and Tenenbaum, 2011).

ACKNOWLEDGMENTS

The research for this review was funded by the European Commission Seventh Framework Programme (FP7/2007-2013) under grant agreement 604102 (Human Brain Project), the program 'Investissements d'Avenir' ANR-10-IAIHU-06, the French National Agency for Research (ANR 2010 BLANC 1403 01), as well as INSERM, CEA, Collège de France, and the Bettencourt-Schuller foundation.

REFERENCES

- Abe, K., and Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat. Neurosci.* 14, 1067–1074.
- Amunts, K., Lenzen, M., Friederici, A.D., Schleicher, A., Morosan, P., Palomero-Gallagher, N., and Zilles, K. (2010). Broca's region: novel organizational principles and multiple receptor mapping. *PLoS Biol.* 8, <http://dx.doi.org/10.1371/journal.pbio.1000489>.
- Andersen, R.A., Essick, G.K., and Siegel, R.M. (1985). Encoding of spatial location by posterior parietal neurons. *Science* 230, 456–458.
- Avarguès-Weber, A., Dyer, A.G., Combe, M., and Giurfa, M. (2012). Simultaneous mastering of two abstract concepts by the miniature brain of bees. *Proc. Natl. Acad. Sci. USA* 109, 7481–7486.
- Bahlmann, J., Schubotz, R.I., and Friederici, A.D. (2008). Hierarchical artificial grammar processing engages Broca's area. *Neuroimage* 42, 525–534.
- Balci, F., Freestone, D., and Gallistel, C.R. (2009). Risk assessment in man and mouse. *Proc. Natl. Acad. Sci. USA* 106, 2459–2463.
- Barone, P., and Joseph, J.P. (1989). Prefrontal cortex and spatial sequencing in macaque monkey. *Exp. Brain Res.* 78, 447–464.
- Bartolo, R., Prado, L., and Merchant, H. (2014). Information processing in the primate basal ganglia during sensory-guided and internally driven rhythmic tapping. *J. Neurosci.* 34, 3910–3923.
- Basirat, A., Dehaene, S., and Dehaene-Lambertz, G. (2014). A hierarchy of cortical responses to sequence violations in three-month-old infants. *Cognition* 132, 137–150.
- Bastiaansen, M., Magyar, L., and Hagoort, P. (2010). Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *J. Cogn. Neurosci.* 22, 1333–1347.
- Beckers, G.J.L., Bolhuis, J.J., Okanoya, K., and Berwick, R.C. (2012). Bird-song neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport* 23, 139–145.
- Bekinschtein, T.A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., and Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proc. Natl. Acad. Sci. USA* 106, 1672–1677.
- Ben-Shachar, M., Palti, D., and Grodzinsky, Y. (2004). Neural correlates of syntactic movement: converging evidence from two fMRI experiments. *Neuroimage* 21, 1320–1336.
- Berdyeva, T.K., and Olson, C.R. (2010). Rank signals in four areas of macaque frontal cortex during selection of actions and objects in serial order. *J. Neurophysiol.* 104, 141–159.
- Bernal, S., Dehaene-Lambertz, G., Millotte, S., and Christophe, A. (2010). Two-year-olds compute syntactic structure on-line. *Dev. Sci.* 13, 69–76.
- Bor, D., Duncan, J., Wiseman, R.J., and Owen, A.M. (2003). Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron* 37, 361–367.
- Botvinick, M., and Watanabe, T. (2007). From numerosity to ordinal rank: a gain-field model of serial order representation in cortical working memory. *J. Neurosci.* 27, 8636–8642.
- Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D.J., and Pykkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain Lang.* 120, 163–173.
- Buiatti, M., Peña, M., and Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage* 44, 509–519.
- Bulf, H., Johnson, S.P., and Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition* 121, 127–132.
- Chaplin, T.A., Yu, H.-H., Soares, J.G.M., Gattass, R., and Rosa, M.G.P. (2013). A conserved pattern of differential expansion of cortical areas in simian primates. *J. Neurosci.* 33, 15120–15125.
- Chen, S., Swartz, K.B., and Terrace, H.S. (1997). Knowledge of the ordinal position of list items in rhesus monkeys. *Psychol. Sci.* 8, 80–86.
- Chomsky, N. (1956). Three models for the description of language. *IEEE Trans. Inf. Theory* 2, 113–124.
- Chomsky, N. (2013). Problems of projection. *Lingua* 130, 33–49.
- Clos, M., Amunts, K., Laird, A.R., Fox, P.T., and Eickhoff, S.B. (2013). Tackling the multifunctional nature of Broca's region meta-analytically: co-activation-based parcellation of area 44. *Neuroimage* 83, 174–188.
- Crowe, D.A., Zarco, W., Bartolo, R., and Merchant, H. (2014). Dynamic representation of the temporal and sequential structure of rhythmic movements in the primate medial premotor cortex. *J. Neurosci.* 34, 11972–11983.
- de Hevia, M.D., and Spelke, E.S. (2010). Number-space mapping in human infants. *Psychol. Sci.* 21, 653–660.
- Dehaene-Lambertz, G., Dehaene, S., and Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science* 298, 2013–2015.
- Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mériaux, S., Roche, A., Sigman, M., and Dehaene, S. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proc. Natl. Acad. Sci. USA* 103, 14240–14245.
- Dominey, P.F., Inui, T., and Hoen, M. (2009). Neural network processing of natural language: II. Towards a unified model of corticostriatal function in learning sentence comprehension and non-linguistic sequencing. *Brain Lang.* 109, 80–92.
- Egner, T., Monti, J.M., and Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *J. Neurosci.* 30, 16601–16608.
- El Karoui, I., King, J.-R., Sitt, J., Meyniel, F., Van Gaal, S., Hasboun, D., Adam, C., Navarro, V., Baulac, M., Dehaene, S., et al. (2014). Event-Related Potential, Time-frequency, and Functional Connectivity Facets of Local and Global Auditory Novelty Processing: An Intracranial Study in Humans. *Cereb. Cortex*. Published online June 26, 2014.
- Eliasmith, C., Stewart, T.C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., and Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science* 338, 1202–1205.
- Elman, J.L. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211.
- Endress, A.D., Nespor, M., and Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends Cogn. Sci.* 13, 348–353.
- Fattal, I., Friedmann, N., and Fattal-Valevski, A. (2011). The crucial role of thiamine in the development of syntax and lexical retrieval: a study of infantile thiamine deficiency. *Brain* 134, 1720–1739.
- Fiorillo, C.D., Newsome, W.T., and Schultz, W. (2008). The temporal precision of reward prediction in dopamine neurons. *Nat. Neurosci.* 11, 966–973.
- Fitch, W.T. (2014). Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition. *Phys. Life Rev.* 11, 329–364.
- Fitch, W.T., and Friederici, A.D. (2012). Artificial grammar learning meets formal language theory: an overview. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1933–1955.

- Fitch, W.T., and Hauser, M.D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380.
- Frank, M.C., and Tenenbaum, J.B. (2011). Three ideal observer models for rule learning in simple languages. *Cognition* 120, 360–371.
- Frankland, S.M., and Greene, J.D. (2015). An architecture for encoding sentence meaning in left mid-superior temporal cortex. *Proc. Natl. Acad. Sci. USA* 112, 11732–11737.
- Friederici, A.D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392.
- Friederici, A.D., Bahlmann, J., Heim, S., Schubotz, R.I., and Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl. Acad. Sci. USA* 103, 2458–2463.
- Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836.
- Fujii, N., and Graybiel, A.M. (2003). Representation of action sequence boundaries by macaque prefrontal cortical neurons. *Science* 301, 1246–1249.
- Gallistel, C.R. (1990). *The organization of learning* (MIT Press).
- Garrido, M.I., Kilner, J.M., Stephan, K.E., and Friston, K.J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clin. Neurophysiol.* 120, 453–463.
- Gavornik, J.P., and Bear, M.F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nat. Neurosci.* 17, 732–737.
- Gentner, T.Q., Fenn, K.M., Margoliash, D., and Nusbaum, H.C. (2006). Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207.
- Georgopoulos, A.P., Lurito, J.T., Petrides, M., Schwartz, A.B., and Massey, J.T. (1989). Mental rotation of the neuronal population vector. *Science* 243, 234–236.
- Giard, M.H., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., and Näätänen, R. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: an event-related potential and dipole-model analysis. *J. Cogn. Neurosci.* 7, 133–143.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychol. Rev.* 84, 279–325.
- Gibbon, J., Malapani, C., Dale, C.L., and Gallistel, C. (1997). Toward a neurobiology of temporal cognition: advances and challenges. *Curr. Opin. Neurobiol.* 7, 170–184.
- Giurfa, M., Zhang, S., Jenett, A., Menzel, R., and Srinivasan, M.V. (2001). The concepts of 'sameness' and 'difference' in an insect. *Nature* 410, 930–933.
- Glasel, H., Leroy, F., Dubois, J., Hertz-Pannier, L., Mangin, J.F., and Dehaene-Lambertz, G. (2011). A robust cerebral asymmetry in the infant brain: the rightward superior temporal sulcus. *Neuroimage* 58, 716–723.
- Goucha, T., and Friederici, A.D. (2015). The language skeleton after dissecting meaning: A functional segregation within Broca's Area. *Neuroimage* 114, 294–302.
- Graf Estes, K., Evans, J.L., Alibali, M.W., and Saffran, J.R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol. Sci.* 18, 254–260.
- Graybiel, A.M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Mem.* 70, 119–136.
- Hadley, R.F. (2009). The problem of rapid variable creation. *Neural Comput.* 21, 510–532.
- Hasson, U., Yang, E., Vallines, I., Heeger, D.J., and Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.* 28, 2539–2550.
- Haun, D.B.M., Jordan, F.M., Vallortigara, G., and Clayton, N.S. (2010). Origins of spatial, temporal and numerical cognition: Insights from comparative psychology. *Trends Cogn. Sci.* 14, 552–560.
- Hauser, M.D., Chomsky, N., and Fitch, W.T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579.
- Hay, J.F., Pelucchi, B., Graf Estes, K., and Saffran, J.R. (2011). Linking sounds to meanings: infant statistical learning in a natural language. *Cognit. Psychol.* 63, 93–106.
- Hill, J., Inder, T., Neil, J., Dierker, D., Harwell, J., and Van Essen, D. (2010). Similar patterns of cortical expansion during human development and evolution. *Proc. Natl. Acad. Sci. USA* 107, 13135–13140.
- Horváth, J., and Winkler, I. (2004). How the human auditory system treats repetition amongst change. *Neurosci. Lett.* 368, 157–161.
- Huettel, S.A., Mack, P.B., and McCarthy, G. (2002). Perceiving patterns in random series: dynamic processing of sequence in prefrontal cortex. *Nat. Neurosci.* 5, 485–490.
- Inoue, M., and Mikami, A. (2006). Prefrontal activity during serial probe reproduction task: encoding, mnemonic, and retrieval processes. *J. Neurophysiol.* 95, 1008–1041.
- Izard, V., Sann, C., Spelke, E.S., and Streri, A. (2009). Newborn infants perceive abstract numbers. *Proc. Natl. Acad. Sci. USA* 106, 10382–10385.
- Javitt, D.C., Steinschneider, M., Schroeder, C.E., and Arezzo, J.C. (1996). Role of cortical N-methyl-D-aspartate receptors in auditory sensory memory and mismatch negativity generation: implications for schizophrenia. *Proc. Natl. Acad. Sci. USA* 93, 11962–11967.
- Jensen, G., Altschul, D., Danly, E., and Terrace, H. (2013). Transfer of a serial representation between two distinct tasks by rhesus macaques. *PLoS ONE* 8, e70285.
- Jin, D.Z., Fujii, N., and Graybiel, A.M. (2009). Neural representation of time in cortico-basal ganglia circuits. *Proc. Natl. Acad. Sci. USA* 106, 19156–19161.
- Jin, X., Tecuapetla, F., and Costa, R.M. (2014). Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. *Nat. Neurosci.* 17, 423–430.
- Johansson, F., Jirenhed, D.-A., Rasmussen, A., Zucca, R., and Hesselow, G. (2014). Memory trace and timing mechanism localized to cerebellar Purkinje cells. *Proc. Natl. Acad. Sci. USA* 111, 14930–14934.
- Jusczyk, P.W., Luce, P.A., and Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *J. Mem. Lang.* 33, 630–645.
- Karuz, E.A., Newport, E.L., Aslin, R.N., Starling, S.J., Tivarus, M.E., and Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain Lang.* 127, 46–54.
- Kheifets, A., and Gallistel, C.R. (2012). Mice take calculated risks. *Proc. Natl. Acad. Sci. USA* 109, 8776–8779.
- King, J.R., Faugeras, F., Gramfort, A., Schurger, A., El Karoui, I., Sitt, J.D., Rohaut, B., Wacongne, C., Labyt, E., Bekinschtein, T., et al. (2013). Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *Neuroimage* 83, 726–738.
- King, J.-R., Gramfort, A., Schurger, A., Naccache, L., and Dehaene, S. (2014). Two distinct dynamic modes subtend the detection of unexpected sounds. *PLoS ONE* 9, e85791.
- Koechlin, E., and Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron* 50, 963–974.
- Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. *Curr. Opin. Neurobiol.* 15, 207–212.
- Kraus, B.J., Robinson, R.J., 2nd, White, J.A., Eichenbaum, H., and Hasselmo, M.E. (2013). Hippocampal "time cells": time versus path integration. *Neuron* 78, 1090–1101.
- Kroch, A., and Joshi, A. (1985). The linguistic relevance of tree-adjointing grammar. Technical Reports (CIS), 671.
- Laje, R., and Buonomano, D.V. (2013). Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nat. Neurosci.* 16, 925–933.

- Lashley, K.S. (1951). The problem of serial order in behavior. In *Cerebral Mechanisms in Behavior: The Hixon Symposium*, L.A. Jeffress, ed. (Wiley), pp. 112–146.
- Leon, M.I., and Shadlen, M.N. (2003). Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron* 38, 317–327.
- Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503, 78–84.
- Mantini, D., Corbetta, M., Romani, G.L., Orban, G.A., and Vanduffel, W. (2013). Evolutionarily novel functional networks in the human brain? *J. Neurosci.* 33, 3259–3275.
- Mäntysalo, S., and Näätänen, R. (1987). The duration of a neuronal trace of an auditory stimulus as indicated by event-related potentials. *Biol. Psychol.* 24, 183–195.
- Marcus, G.F. (2001). *The Algebraic Mind - Integrating Connectionism & Cognitive Science* (MIT Press).
- Marcus, G.F. (2013). Evolution, Memory, and the Nature of Syntactic Representation. In *Birdsong, Speech, and Language* (MIT Press), pp. 27–44.
- Marcus, G.F., Vijayan, S., Bandi Rao, S., and Vishton, P.M. (1999). Rule learning by seven-month-old infants. *Science* 283, 77–80.
- Marcus, G.F., Fernandes, K.J., and Johnson, S.P. (2007). Infant rule learning facilitated by speech. *Psychol. Sci.* 18, 387–391.
- Maruyama, M., Pallier, C., Jobert, A., Sigman, M., and Dehaene, S. (2012). The cortical representation of simple mathematical expressions. *Neuroimage* 61, 1444–1460.
- May, P.J., and Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology* 47, 66–122.
- Mazoyer, B.M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Dehaene, S., Cohen, L., and Mehler, J. (1993). The cortical representation of speech. *J. Cogn. Neurosci.* 5, 467–479.
- McComb, K., Packer, C., and Pusey, A. (1994). Roaring and numerical assessment in contests between groups of female lions, *Panthera leo*. *Anim. Behav.* 47, 379–387.
- Melançon, A., and Shi, R. (2015). Representations of abstract grammatical feature agreement in young children. *J. Child Lang.* Published online January 30, 2015.
- Mello, G.B.M., Soares, S., and Paton, J.J. (2015). A scalable population code for time in the striatum. *Curr. Biol.* 25, 1113–1122.
- Merchant, H., Zarco, W., Pérez, O., Prado, L., and Bartolo, R. (2011). Measuring time with different neural chronometers during a synchronization-continuation task. *Proc. Natl. Acad. Sci. USA* 108, 19784–19789.
- Merchant, H., Bartolo, R., Pérez, O., Méndez, J.C., Mendoza, G., Gámez, J., Yc, K., and Prado, L. (2014). Neurophysiology of timing in the hundreds of milliseconds: multiple layers of neuronal clocks in the medial premotor areas. *Adv. Exp. Med. Biol.* 829, 143–154.
- Merritt, D.J., and Terrace, H.S. (2011). Mechanisms of inferential order judgments in humans (*Homo sapiens*) and rhesus monkeys (*Macaca mulatta*). *J. Comp. Psychol.* 125, 227–238.
- Meyer, T., and Olson, C.R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc. Natl. Acad. Sci. USA* 108, 19401–19406.
- Meyer, T., Ramachandran, S., and Olson, C.R. (2014). Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *J. Neurosci.* 34, 9332–9337.
- Minier, L., Fagot, J., and Rey, A. (2015). The Temporal Dynamics of Regularity Extraction in Non-Human Primates. *Cogn. Sci.* Published online August 25, 2015. <http://dx.doi.org/10.1111/cogs.12279>.
- Monti, M.M., Parsons, L.M., and Osherson, D.N. (2009). The boundaries of language and thought in deductive inference. *Proc. Natl. Acad. Sci. USA* 106, 12554–12559.
- Monti, M.M., Parsons, L.M., and Osherson, D.N. (2012). Thought beyond language: neural dissociation of algebra and natural language. *Psychol. Sci.* 23, 914–922.
- Moro, A. (1997). Dynamic Antisymmetry: Movement as a Symmetry-breaking Phenomenon. *Stud. Linguist.* 51, 50–76.
- Moro, A., Tettamanti, M., Perani, D., Donati, C., Cappa, S.F., and Fazio, F. (2001). Syntax and the brain: disentangling grammar by selective anomalies. *Neuroimage* 13, 110–118.
- Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., and Weiller, C. (2003). Broca's area and the language instinct. *Nat. Neurosci.* 6, 774–781.
- Nakai, T., and Sakai, K.L. (2014). Neural mechanisms underlying the computation of hierarchical tree structures in mathematics. *PLoS ONE* 9, e111439.
- Namoodiri, V.M.K., Huertas, M.A., Monk, K.J., Shouval, H.Z., and Hussain Shuler, M.G. (2015). Visually cued action timing in the primary visual cortex. *Neuron* 86, 319–330.
- Neubert, F.-X., Mars, R.B., Thomas, A.G., Sallet, J., and Rushworth, M.F.S. (2014). Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. *Neuron* 81, 700–713.
- Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. *Proc. Natl. Acad. Sci. USA* 109, 11860–11865.
- Nieder, A., Diester, I., and Tudusciuc, O. (2006). Temporal and spatial enumeration processes in the primate parietal cortex. *Science* 313, 1431–1435.
- Ninokura, Y., Mushiaki, H., and Tanji, J. (2004). Integration of temporal order and object information in the monkey lateral prefrontal cortex. *J. Neurophysiol.* 91, 555–560.
- Ohmae, S., Uematsu, A., and Tanaka, M. (2013). Temporally specific sensory signals for the detection of stimulus omission in the primate deep cerebellar nuclei. *J. Neurosci.* 33, 15432–15441.
- Orbán, G., Fiser, J., Aslin, R.N., and Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proc. Natl. Acad. Sci. USA* 105, 2745–2750.
- Orlov, T., Yakovlev, V., Hochstein, S., and Zohary, E. (2000). Macaque monkeys categorize images by their ordinal number. *Nature* 404, 77–80.
- Pallier, C., Devauchelle, A.D., and Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci. USA* 108, 2522–2527.
- Pegado, F., Bekinschtein, T., Chausson, N., Dehaene, S., Cohen, L., and Naccache, L. (2010). Probing the lifetimes of auditory novelty detection processes. *Neuropsychologia* 48, 3145–3154.
- Penn, D.C., Holyoak, K.J., and Povinelli, D.J. (2008). Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.* 31, 109–130, discussion 130–178.
- Pinker, S., and Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition* 95, 201–236.
- Plate, T.A. (1995). Holographic reduced representations. *IEEE Trans. Neural Netw.* 6, 623–641.
- Raij, T., McEvoy, L., Mäkelä, J.P., and Hari, R. (1997). Human auditory cortex is activated by omissions of auditory stimuli. *Brain Res.* 745, 134–143.
- Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychol. Rev.* 77, 481–495.
- Restle, F., and Brown, E.R. (1970). Serial pattern learning. *J. Exp. Psychol.* 83, 120–125.
- Rilling, J.K., Glasser, M.F., Preuss, T.M., Ma, X., Zhao, T., Hu, X., and Behrens, T.E. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.* 11, 426–428.
- Rolheiser, T., Stamatakis, E.A., and Tyler, L.K. (2011). Dynamic processing in the human language system: synergy between the arcuate fascicle and extreme capsule. *J. Neurosci.* 31, 16949–16957.

- Saffran, J., and Wilson, D. (2003). From Syllables to Syntax: Multilevel Statistical Learning by 12-Month-Old Infants. *Infancy* 4, 273–284.
- Saffran, J.R., Aslin, R.N., and Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.
- Saffran, J., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., and Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition* 107, 479–500.
- Saur, D., Schelter, B., Schnell, S., Kratochvil, D., Küpper, H., Kellmeyer, P., Kümmerer, D., Klöppel, S., Glauche, V., Lange, R., et al. (2010). Combining functional and anatomical connectivity reveals brain networks for auditory language comprehension. *Neuroimage* 49, 3187–3197.
- Schapiro, A.C., Rogers, T.T., Cordova, N.I., Turk-Browne, N.B., and Botvinick, M.M. (2013). Neural representations of events arise from temporal community structure. *Nat. Neurosci.* 16, 486–492.
- Schoenemann, P.T., Sheehan, M.J., and Glotzer, L.D. (2005). Prefrontal white matter volume is disproportionately larger in humans than in other primates. *Nat. Neurosci.* 8, 242–252.
- Sella, F., Berteletti, I., Lucangeli, D., and Zorzi, M. (2015). Spontaneous non-verbal counting in toddlers. *Dev. Sci.* Published online March 5, 2015. <http://dx.doi.org/10.1111/desc.12299>.
- Shetreet, E., and Friedmann, N. (2014). The processing of different syntactic structures: fMRI investigation of the linguistic distinction between wh-movement and verb movement. *J. Neurolinguist.* 27, 1–17.
- Shetreet, E., Friedmann, N., and Hadar, U. (2009). An fMRI study of syntactic layers: sentential and lexical aspects of embedding. *Neuroimage* 48, 707–716.
- Shima, K., Isoda, M., Mushiake, H., and Tanji, J. (2007). Categorization of behavioural sequences in the prefrontal cortex. *Nature* 445, 315–318.
- Smith, K.S., and Graybiel, A.M. (2013). A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron* 79, 361–374.
- Smolensky, P. (1990). Tensor Product Variable Binding and the Representation of Symbolic Structures in Connectionist Systems. *Artif. Intell.* 46, 159–216.
- Smolensky, P., and Legendre, G. (2006). *The harmonic mind* (MIT Press).
- Snijders, T.M., Vosse, T., Kempen, G., Van Berkum, J.J.A., Petersson, K.M., and Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension: an FMRI study using word-category ambiguity. *Cereb. Cortex* 19, 1493–1503.
- Sohn, J.-W., and Lee, D. (2007). Order-dependent modulation of directional signals in the supplementary and presupplementary motor areas. *J. Neurosci.* 27, 13655–13666.
- Sportiche, D., Koopman, H., and Stabler, E. (2013). *An Introduction to Syntactic Analysis and Theory* (Hoboken: Wiley-Blackwell).
- Strauss, M., Sitt, J.D., King, J.-R., Elbaz, M., Azizi, L., Buiatti, M., Naccache, L., van Wassenhove, V., and Dehaene, S. (2015). Disruption of hierarchical predictive coding during sleep. *Proc. Natl. Acad. Sci. USA* 112, E1353–E1362.
- Summerfield, C., Trittschuh, E.H., Monti, J.M., Mesulam, M.M., and Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* 11, 1004–1006.
- ten Cate, C., and Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1984–1994.
- Terrace, H.S., Son, L.K., and Brannon, E.M. (2003). Serial expertise of rhesus macaques. *Psychol. Sci.* 14, 66–73.
- Todorovic, A., and de Lange, F.P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *J. Neurosci.* 32, 13389–13395.
- Tremblay, P., Baroni, M., and Hasson, U. (2013). Processing of speech and non-speech sounds in the supratemporal plane: auditory input preference does not predict sensitivity to statistical structure. *Neuroimage* 66, 318–332.
- Tyler, L.K., Marslen-Wilson, W.D., Randall, B., Wright, P., Devereux, B.J., Zhuang, J., Papoutsis, M., and Stamatakis, E.A. (2011). Left inferior frontal cortex and syntax: function, structure and behaviour in patients with left hemisphere damage. *Brain* 134, 415–431.
- Uhrig, L., Dehaene, S., and Jarraya, B. (2014). A hierarchy of responses to auditory regularities in the macaque brain. *J. Neurosci.* 34, 1127–1132.
- Vagharchakian, L., Dehaene-Lambertz, G., Pallier, C., and Dehaene, S. (2012). A temporal bottleneck in the language comprehension network. *J. Neurosci.* 32, 9089–9102.
- van der Velde, F., and de Kamps, M. (2006). Neural blackboard architectures of combinatorial structures in cognition. *Behav. Brain Sci.* 29, 37–70, discussion 70–108.
- van Heijningen, C.A.A., de Visser, J., Zuidema, W., and ten Cate, C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc. Natl. Acad. Sci. USA* 106, 20538–20543.
- Veit, L., and Nieder, A. (2013). Abstract rule neurons in the endbrain support intelligent behaviour in corvid songbirds. *Nat. Commun.* 4, 2878.
- Viswanathan, P., and Nieder, A. (2013). Neuronal correlates of a visual “sense of number” in primate parietal and prefrontal cortices. *Proc. Natl. Acad. Sci. USA* 110, 11187–11192.
- Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., and Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc. Natl. Acad. Sci. USA* 108, 20754–20759.
- Wacongne, C., Changeux, J.P., and Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *J. Neurosci.* 32, 3665–3678.
- Wang, L., Uhrig, L., Jarraya, B., and Dehaene, S. (2015). Representation of Numerical and Sequential Patterns in Macaque and Human Brains. *Curr. Biol.* 25, 1966–1974.
- Wilson, B., Slater, H., Kikuchi, Y., Milne, A.E., Marslen-Wilson, W.D., Smith, K., and Petkov, C.I. (2013). Auditory artificial grammar learning in macaque and marmoset monkeys. *J. Neurosci.* 33, 18825–18835.
- Xiang, H.D., Fonteijn, H.M., Norris, D.G., and Hagoort, P. (2009). Topographical functional connectivity pattern in the perisylvian language networks. *Cereb. Cortex* 20, 549–560.
- Yang, C. (2013). Ontogeny and phylogeny of language. *Proc. Natl. Acad. Sci. USA* 110, 6324–6327.