Cognitive and neural representations of fractals in vision, music and action.

Mauricio de Jesus Dias Martins

SCAN-Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria; email: mauricio.martins@univie.ac.at

Abstract

The concept of fractal was popularized by Mandelbrot as a tool to tame the geometrical structure of objects with infinite hierarchical depth. The key aspect of fractals is the use of simple parsimonious rules and initial conditions, which when applied recursively can generate unbounded complexity. Fractals are structures ubiquitous in nature, being present in coast lines, bacteria colonies, trees, and physiological time series. However, within the field of cognitive science the core question is not which phenomena can generate fractal structures, but whether human or animal minds can represent recursive processes, and if so in which domains. In this chapter we will explore the cognitive and neural mechanisms underlying the representation of recursive hierarchical embedding. Language is the domain in which this capacity is best studied. Humans can generate an infinite array of hierarchically structured sentences, and this capacity distinguishes us from other species. However, recent research suggests that humans can represent similar structures in the domains of music, vision and action and has provided additional cues as to how these capacities are cognitively implemented. Using a comparative approach, we will map the commonalities and differences across domains and offer a roadmap to understand the neurobiological implementation of fractal cognition.

Keywords

Recursion, fMRI, Cognition, Language

1 Introduction

Hierarchies, intuitively understood as structures with multiple levels of organization, are common in nature. The most paradigmatic example of a hierarchy is that of a tree, with a trunk, branches, sub-branches, and leaves. In a tree, several branches originate from the trunk, several sub-branches originate from each branch, and several leaves sprout from each sub-branch. This “branching” property, in which one item at level of organization L(i) connects with two or more items at a subordinate level L(i+1), is the key feature of all hierarchies (Fig. 1a) (Udden et al., 2019). It is therefore no surprise that the metaphor of tree is used to describe hierarchies in multiple domains.

Fractals are a kind of hierarchy which can have an unbounded number of structural levels. When trying to measure the length of coast lines, Mandelbrot understood that the measured length depended on the size of the ruler (Mandelbrot, 1977). When using smaller rulers, more and more infinitesimal segments of the coastline could be revealed increasing its length potentially to infinity. The crucial insight of Mandelbrot was that these structures with infinitesimal details often displayed the property of self-similarity, i.e., they often have regularities across levels of organization, which allow for parsimonious formal descriptions (Fig. 1b). In other words, Mandelbrot understood that we could describe natural objects with unbounded hierarchical depth using simple sets of rules. To do so we need: a description of initial conditions, certain transformation rules, and the ability to apply these rules recursively, i.e., to feed the output of the operation back to itself as input (Fig. 1c).

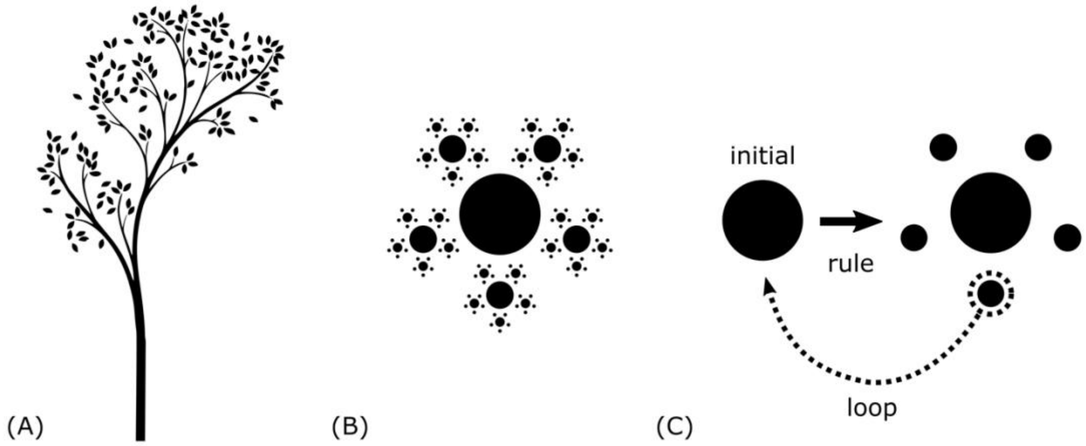

Fig. 1 (A) Hierarchies are tree-like structures in which an element at a level L(i) (e.g. trunk) connects with two or more elements at a subordinate level L(i+1) (e.g. branches); **(B)** Fractals are structures in which this "hierarchical branching" can extend to infinity by using transformation rules consistent across multiple levels L(1), L(2), L(3) … ; **(C)** To generate fractals we need an *initial* condition L(0), a transformation *rule* {L(i+1) → L(i) + some operation}, and a *loop*. In this case, the initial condition L(0) is a black circle, the transformation rule generates L(i+1) by "adding 5 circles at particular angles and distance relative each circle in L(i)", and then the same rule can applied to each newly added circle, recursively, to generate fractal (B).

Defined as such, we can find fractal structures not only in coast lines and trees, but also in bacteria colonies (Matsuyama & Matsushita, 1993), worm behavioral responses (Arata et al., 2022), tissue structure (e.g. human lungs and blood vessels) and physiological signals such as electroencephalogram (EEG) and electrocardiogram (ECG) time series (Ruiz-Padial & Ibáñez-Molina, 2018; Tiwari et al., 2019). In all these domains we can find complex multilayered signals with structural self-similarity. However, the fact that these signals can be generated by natural phenomena does not make them immediately relevant or transparent for the human cognitive level. Our fractal EEG or ECG time series are not immediately transparent to our perception without sophisticated technology. Similarly, the fact that bacteria colonies can grow in a fractal pattern does not mean that this structure is represented by the bacteria themselves. Hence, from the perspective of cognitive science the relevant question is whether human and animal minds can *represent* the kinds of recursive processes that generate fractals (Lobina, 2017; Martins, 2012). In the next section we will discuss how the representation of these generative processes can be experimentally investigated.

1.1 Recursion in human cognition

Humans are exceptional in their ability to generate complex and unbounded hierarchical structures, across a variety of domains, including in the social domain, visuo-spatial processing, action planning, and in language. In all these domains we are able not only to recognize underlying hierarchical structures but also to extend them beyond the given in a way that is consistent with the previous levels. Imagine the complex hierarchical structure denoting a national defense system composed of several armies, battalions etc. We can always go beyond the given and create supra-national command levels dominant over (but consistent with) the national structures (think of NATO). Similarly, in the language domain, we can create an infinite set of structured expressions from a finite set of words and hierarchical rules. For example, we could imagine an arbitrarily long sentence S, and embed it within the sentence “I think that [S]”. We could then embed the resulting sentence in “Peter thinks that [I think that [S]]”, and so on.

The cognitive and neural mechanisms underlying the capacity to recursively embed hierarchies within higher-order structures – or to add new levels beyond the given - are an exciting and very active topic of research (Martins, 2012, for a review). For example, within the domain of language, the generative syntax tradition proposes that there is a core recursive cognitive system which allows the generation of unbounded complexity from a finite set of primitives and rules (Berwick & Chomsky, 2015; Friederici et al., 2017; Hauser et al., 2002; Perfors et al., 2010, 2011). These syntactic rules govern how different syntactic categories (nouns, verbs, etc.) can be combined to generate increasingly complex sentences. Consider the following system to generate noun phrases (NP), verb phrases (VP) and sentences (S) using determinants (D), nouns (N) and verbs (V):

1. NP →(D) N
2. VP → V (NP)
3. S → NP VP

With these rules, we can generate the noun and verb phrases:

NP → [the$_D$ girl$_N$]$_{NP}$;
VP → [opened$_V$ the door$_{NP}$]$_{VP}$

and combine them to form a sentence:

S → [[the girl]$_{NP}$ [opened the door]$_{VP}$]$_S$

Unlike in the visual domain, in language we rarely have self-similarity at the surface level. While sentences like “[I think that [Peter thinks that [Susan thinks that [S]]]]” can exist, they are rare. Mostly, self-similarity exists at a deeper conceptual level. For instance, the compound noun “[[[student] film] committee]” can be translated as NPs embedded within NPs ([[[NP]NP]NP]) using the generative rule NP → [[NP]NP]. These rules, like fractals, involve *direct recursion*, in which both sides of the symbol “→” contain objects of the same category. However, even such direct recursion is rare in language (Arsenijević & Hinzen, 2012). Most commonly, sentences are generated using *indirect recursion*, i.e., via the combination of a set of syntactic rules, which do not necessarily have similar categories on both sides of “→” (Roeper, 2011).

Here is how it works in practice. We can use the rule system above to generate the sentence *S’*:

S’ → [[the girl]$_{NP}$ [kissed the boy]$_{VP}$]$_S$

If we modify rule 1 to allow embedding of *S’* within a NP, we create an indirect loop between NP and S because NP can now contain a S (NP → N *S’*) and S can contain a NP (S → NP VP).

With this *indirect recursive* loop can we can embed *S’* within S = “[[the girl]$_{NP}$ [opened the door]$_{VP}$]$_{S}$”, because the latter contains the NP “the girl”:

S = [[the girl [*S’*]]$_{NP}$ [opened the door]$_{VP}$]$_{S}$.

If we now re-write S including *S’*, and use the appropriate pronoun, we obtain the following sentence (see Fig.2 for a graphical description):

[[the girl [*who kissed the boy*]$_{S'}$]$_{NP}$ [opened the door]$_{VP}$]$_{S}$

This system can be applied to generate sentences of unbounded depth by allowing sentences to be embedded within other sentences.

The core point, made in a seminal paper by Hauser, Chomsky and Fitch is that “Language is based on a recursive generative procedure that takes elements from [...] the lexicon, and applies repeatedly to yield structured expressions, without bound” (Hauser et al., 2002). Crucially, this idea of recursion in language requires more than the mathematical notion of “definition by induction” which requires initial conditions and a recursive function (Odifreddi, 1989). For example, the set of natural numbers can be generated using the initial condition L(0) = 1, and the recursive function L(i+1) = L(i)+1. When applied recursively, the function generates the set {1, 2, 3, …, n}. However, natural numbers are not hierarchical, hence classical recursion is not sufficient to describe fractals. Recursion in language also requires that the rule allows the generation of new hierarchical levels via structural embedding (e.g NP → [[NP]NP]). Thus, when we talk about recursion in language, we are mainly referring to Recursive Hierarchical Embedding (henceforth RHE), which is the focus of our chapter.

RHE as a procedure to generate complex hierarchies from a finite set of rules and primitive elements allows us to model other phenomena present in human cognition. Beyond the domains of language and vision (Fitch et al., 2005; Hauser et al., 2002; Jackendoff & Pinker, 2005; Pinker & Jackendoff, 2005), also music and action planning contain complex hierarchical structures which can be generated and represented using RHE (Badre, 2008; Fadiga et al., 2009; Fitch & Martins, 2014; Jackendoff, 2003; Jackendoff & Lerdahl, 2006; Lerdahl & Jackendoff, 1977; Rohrmeier & Koelsch, 2012).

Below, we will discuss the state of the art on how the representation of RHE is instantiated in the cognitive and neurobiological context. We will start with the domain of linguistic syntax, and then discuss how we used musical and motor fractals to investigate the representation of RHE in these domains.

2 Neural mechanisms of recursive hierarchical embedding in language

Linguistic utterances are sequential. When we speak or write, we produce words one after the other. However, this sequential order is only superficial. To understand the intended meaning of a sentence, we also need to project the underlying hierarchical structure (Berwick & Chomsky, 2015). For instance, in the sentence “The girl who kissed the boy *opened* the door”, we know that it was *the girl* who opened the door and not *the boy*, despite *the boy* being closer to the verb *opened* in the word sequence. We can infer this relationship automatically because the phrase *the girl* is closer to the verb *opened* in the hierarchical structure (Fig. 2), i.e., it occupies a hierarchical level closer to the verb, in comparison with *the boy* which is more deeply nested.

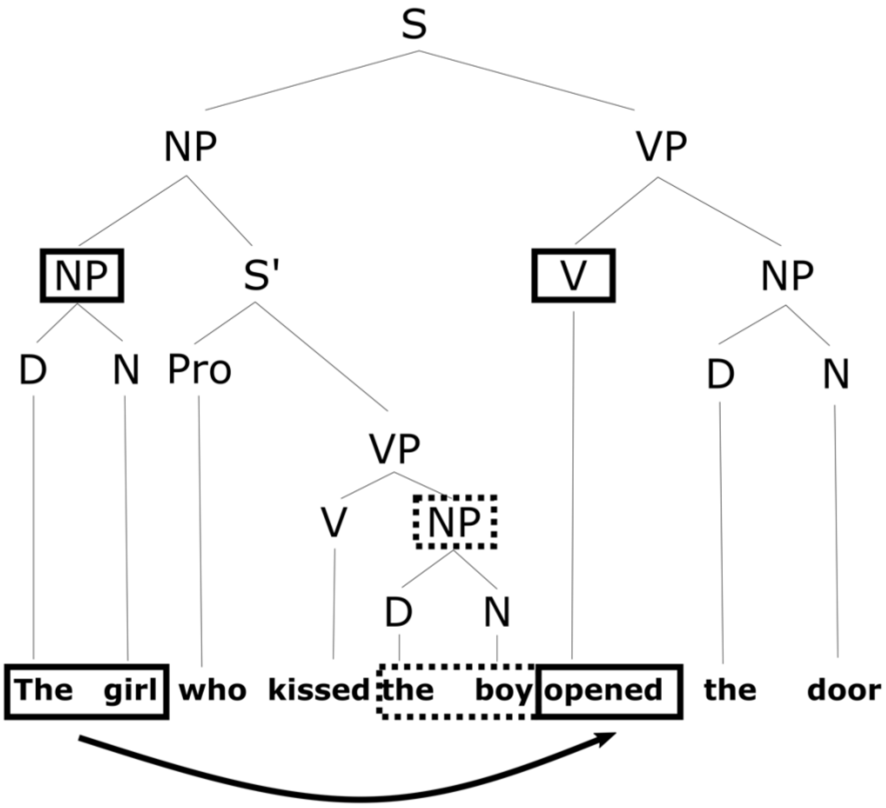

Fig. 2. Language is hierarchical. In the sentence **"**the girl *who kissed the boy* opened the door", the "boy" is closer to the verb "opened" than the "girl". However, competent English speakers can understand that it was "the girl" who "opened the door" and not "the boy". This is because the NP "the girl" is closer to the verb in the hierarchical structure (solid boxes) than the NP "the boy", which is deeply nested in the bottom of the hierarchy (dashed boxes).

When competent adult speakers process language, the hierarchical interpretation takes precedence over the sequential interpretation. Consequently, one of the most distinctive signatures of hierarchical processing in language is the capacity to process dependencies between words which are non-adjacent in the sequential structure. This behavioral signature has been extensively used in empirical research (Udden et al., 2019, for a review).

Interestingly, children younger than 7 y.o. might not always be able to inhibit the sequential representation, and recursive hierarchical structure seems to develop slowly and requires extensive experience (Roeper, 2007, 2009, 2011). For instance, when asked to pick the "second red ball" from an array of colored circles, children often chose the ball in the second position which is also red. This corresponds to the conjunctive interpretation "second AND red ball" which has no hierarchical depth. Conversely, adults can project the structure [second [red ball]], and choose (for instance) the fourth circle in the array, but the second one that is red. The capacity to represent this nested structure is also a crucial signature used in empirical research (Roeper, 2007, 2009, 2011).

When applied to neuroimaging research, these empirical signatures have yielded consistent results. The processing of linguistic syntax, hierarchical depth, and non-adjacent dependencies strongly correlates with activity in a left lateralized network comprising inferior frontal gyrus (IFG) and posterior superior temporal sulcus (pSTS) (extending to middle and superior temporal gyri)(Friederici et al., 2017; Hagoort & Indefrey, 2014; Makuuchi et al., 2009;

Zaccarella et al., 2015). This pattern of activity is present in both natural and artificial languages but is modulated by the developmental stage and the amount of experience with hierarchical structure (Jeon & Friederici, 2013, 2015; Skeide et al., 2014; Skeide & Friederici, 2016). For instance, in children, both syntactic and semantic processing seem to initially correlate with activity in left pSTS, but syntax-related activity in IFG becomes more robust later, around 9 years old (Skeide et al., 2014). Research in adults investigating the learning of artificial languages also shows that IFG activity increases with the degree of training and automatization (Jeon & Friederici, 2015). These results suggest that there is an optimal neural machinery specialized in the processing of hierarchical structure in language. However, stable recruitment of this specialized circuitry emerges throughout ontogeny and necessitates a certain volume of experience and exposure. In other words, the pattern of brain activity during the acquisition of RHE rules in language is different to that found after extensive training and automatization.

This pattern of activity is particularly interesting because IFG and pSTS are among the brain structures which underwent rapid expansion in the human lineage in comparison with other primates (Buckner & Krienen, 2013; Rilling et al., 2008). Most notably, the fiber tract connecting them – the arcuate fasciculus – also underwent dramatic expansion in humans and is underdeveloped in human children until the age of 7-8 y.o (Rilling et al., 2008). This could explain why the capacity to represent hierarchies in non-human animals is limited in scope - being available perhaps in spatial navigation and social structures (Buzsáki & Moser, 2013; Seyfarth & Cheney, 2014) - and depth - not going beyond a few hierarchical levels. The necessity of a mature arcuate fasciculus to support the capacity to represent hierarchies could also explain why RHE is not present in human children until relatively late.

This set of findings raises interesting questions about human evolution, but it is limited to the domain of language. Is RHE available in other domains? If so, does it recruit similar brain networks? Is there evidence for domain general acquisition principles; and do we find similar domain specialization with experience? Answering these questions requires the extension of this research program to other domains. In the next sections we will summarize how we used fractals in music (Martins, Gingras, et al., 2017; Martins et al., 2020), action (Martins, Bianco, et al., 2019), and vision (Martins, Fischmeister, et al., 2014; Martins et al., 2016) to map the cognitive and neural mechanisms supporting RHE in different domains. Using a comparative approach, we will analyze the commonalities and differences between domains, and draw the larger picture concerning how RHE is implemented in neurocomputational systems.

3 Recursive hierarchical embedding in the visual, musical, and motor domains – behavioral research

As we discussed above, words are presented sequentially in language. However, in addition to this sequential structure there is an underlying syntactic hierarchy which is represented by competent language speakers (Udden et al., 2019). This kind of dual nature – sequential and hierarchical – is also present in domains like music and action. Music is composed both by a sequence of tones unfolding in time one after the other, but also by an underlying hierarchy of tonal tension which goes beyond adjacent relationships (Jackendoff & Lerdahl, 2006; Koelsch et al., 2013; Lerdahl & Jackendoff, 1977). Experienced musicians and music listeners can detect these tension dynamics along musical phrases and react with surprise when there are violations (Koelsch et al., 2013). Similarly, in action, we execute motor commands sequentially, one after the other. For instance, in the process of making coffee we must turn on the machine, add coffee to the container, add water to the container, place the cup under the dispenser and press the button to dispense the coffee. However, in case there is already coffee and water in the respective containers, we can skip those steps without getting lost in the sequence. This is because we can represent the hierarchical structure underlying the goal of making coffee which goes beyond adjacent connections between motor commands one after the other (Fitch & Martins, 2014; Jackendoff, 2013; Lashley, 1951).

fMRI research on the musical and motor domains has mostly investigated the responses elicited by the processing of tonal and action violations (Fadiga et al., 2009; Fitch & Martins, 2014, for reviews). These violations consistently activate IFG across domains in addition to other regions more suggestive of domain specialization (e.g. sensorimotor brain areas in the motor domain and auditory areas in music) (Bianco et al., 2015, 2016). The common IFG activation between language, action and music suggests computational commonalities in the implementation of hierarchical representations across these domains (Fadiga et al., 2009; Fitch & Martins, 2014, for reviews). However, the exact activity loci can vary. In music and action, IFG activity is stronger in the right hemisphere, and precise comparisons reveal little commonalities between domains within left IFG, hinting on a language specialized circuitry (Amunts & Zilles, 2012; Fedorenko et al., 2011; Fedorenko & Shain, 2021). This circuit micro-specialization is to be expected in the development of expert automatized processing (Asano et al., 2022). However, a robust arcuate fasciculus connecting IFG and STS might facilitate similar computations across domains, even if the exact loci of activity differ from domain to domain.

This research based on the detection of violations is interesting but presents some confounds. The patterns of activity might reflect attention or surprise effects rather than computations specific for recursive hierarchical embedding. Moreover, it is not clear whether participants must represent the underlying generative processes in these violation paradigms. To investigate the latter more specifically we need to tap on the capacity to extend hierarchies beyond the given. To resolve this gap, my colleagues and I used the principles of fractal geometry to create novel paradigms in the visual, music and action domains (Fig. 3):

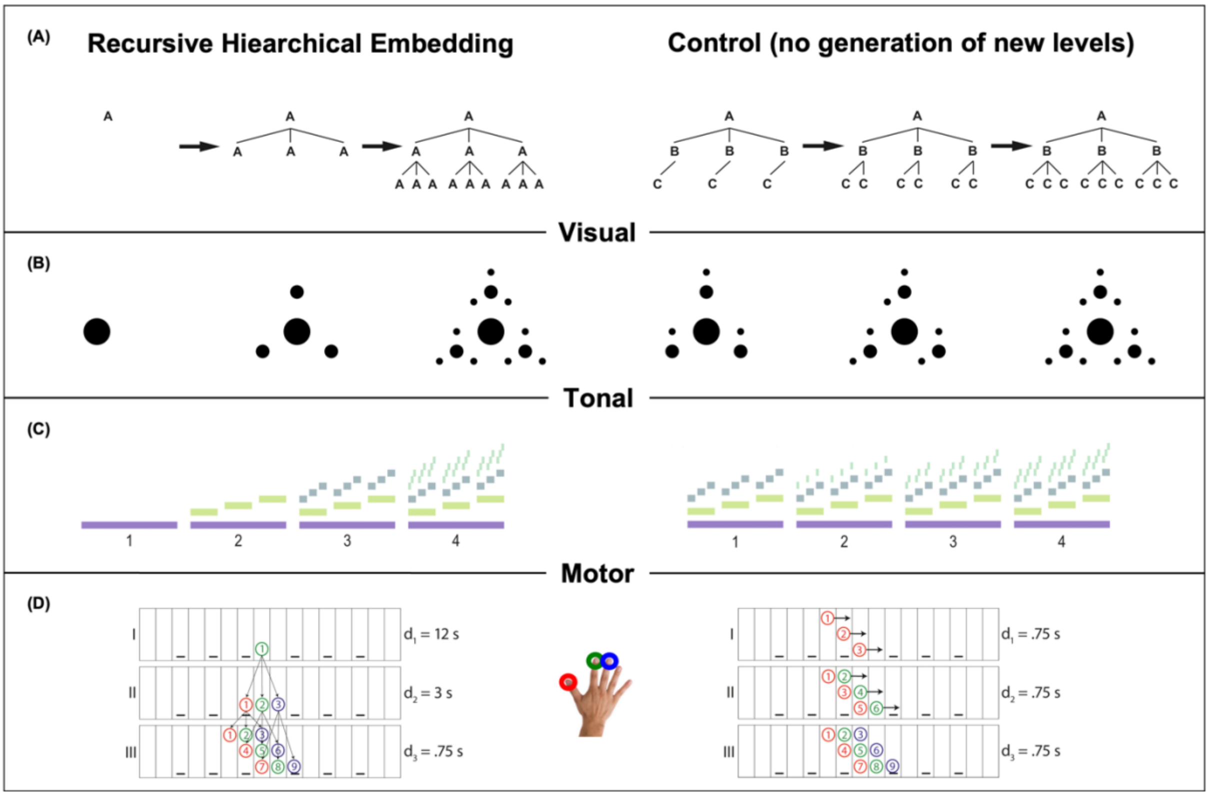

Fig. 3. (A) General principles of Recursive Hierarchical Embedding (RHE; left) and a control task which does not generate new levels of the hierarchy (right). At surface, both rules generate the same final structure. However, in the RHE task, each step generates a new hierarchical level, while in the control task each step adds elements within a fixed hierarchical level without generating new. **(B) Visual fractals.** These are generated using the rules described in A. **(C) Tonal fractals.** Here the horizontal axis denotes tone duration, and the vertical axis denotes pitch. Tonal fractals are generated stepwise from an initial long-duration and low-pitch tone. Each step adds a sequence of (higher pitch and shorter duration) tones with a particular contour, and pitch relations consistent with the previous level. **(D) Motor fractals.** Each step adds a sequence of finger tapping movements on a silent keyboard (with the thumb, index, and middle finger) with contour and key distances consistent to those of the previous level. The duration of each finger tap at level d_n is shorter than those of level d_{n-1}.

In this research program, we focus on the capacity to represent recursive hierarchical embedding, i.e. the ability to extract cross-level regularities and use them productively to generate new levels (Martins, 2012). Crucially, we do not focus on how participants represent fractal structures *per se* but rather how participants represent their generative process. Indeed, in our control condition, the rule also generates a visual fractal, but without adding new hierarchical levels. This controls for stimuli-related features and focuses the representation of the underlying RHE process. We expose participants to the first three iterations of the generative process and ask them to imagine, and then select (or execute), the next correct continuation (Fig. 4b).

We first utilized this paradigm in the visual domain (Martins et al., 2016). We learnt that adult humans can easily represent RHE in vision, and that this capacity is neither well predicted by general intelligence nor by visual-spatial working memory. However, we found that the capacity to process RHE in vision correlates strongly with recursive action planning as measured by the Tower of Hanoi, hinting on shared cognitive resources across domains (Martins, Gingras, et al., 2017). We also found that this capacity matures slowly, and it is available in children older than 8 y.o., but not before (Martins, Laaha, et al., 2014). Similar to language, mastering RHE requires experience with simpler conjunctive visual processes, before RHE becomes available. These data suggest that there might be some cognitive specialization for RHE and shared resources across domains.

To investigate these questions, we developed music and motor tasks (Martins, Bianco, et al., 2017; Martins, Gingras, et al., 2017) with the same underlying principles (Fig 4c&d). Here, each step of the process generated an additional hierarchical level of musical tones, or of finger movements presented on a silent keyboard. Broadly, our behavioral research suggests that the ability to represent RHE shares cognitive resources across domains, when controlling for domain-specific effects. Factor analyses including musical and visual RHE tasks, their respective non-recursive controls, and the recursive action planning Tower of Hanoi, show that all recursive tasks cluster together within a factor orthogonal to general cognitive capacity (Martins, Gingras, et al., 2017).

Recent research in other domains has further hinted on a specialized capacity to the processing of hierarchical structures. For example, accuracy in the visual RHE task was found to specifically correlate with the understanding of sentences with 2-levels of hierarchical embedding, such as “[the bird [who the frog [who is red] washes] laughs]” (Martins, Krause, et al., 2019). In this sentence, participants need to parse the non-adjacent relationship between *bird* and *laughs*. The same visual RHE was found to specifically correlate with the ability to parse 2-level if-then logical hierarchies, again when controlling for domain specific variance (Scholz, 2020). In this logical task, participants are asked to determine the numeric value of a pair of objects based on their relationship. For example, in the hierarchical task, if the first object is red and if the shape of the second object is *similar* to the first one, then the numeric value is “3”; if the first object is red and if the shape of the second object is *different* to the first one, then the numeric value is “1”, etc. If the first object is green, numeric contingencies are different. In the control – conjunctive - task the value of the pair is simply the sum of the individual values without any higher order relation.

In summary, behavioral research suggests that there are specific cognitive resources dedicated to the processing of RHE across domains. In the next section we will examine the neural systems that support these capacities.

4 Recursive hierarchical embedding in the visual, musical, and motor domains – fMRI research

To investigate the neural bases supporting the representation of RHE, we applied similar paradigms across domains (Fig. 3). In general, we present participants with the first three steps generating a fractal structure and ask participants to imagine the correct continuation of the same process. In the visual task, they chose the correct choice from two alternatives (Martins, Fischmeister, et al., 2014); in the musical task they decide whether the 4th step is correct or incorrect (Martins et al., 2020); and in the motor task they execute the correct 4th step in a keyboard inside of the scanner (Martins, Bianco, et al., 2019). In all cases, the control task also depicted a control process generating a fractal structure in three steps. However, the latter does not entail the generation of new hierarchical levels. This contrast, by keeping the choice and execution stimuli constant across conditions, is optimal to isolate the internal representations of RHE. Crucially, we focused our fMRI contrasts on the period in which participants are imagining the 4th step, after attending to the 3rd (thus a period with internal *simulation* but without external *stimulation*).

With this design, we found that different domains activate specialized regions. In particular, we find bilateral activity in the visual central stream, hippocampus and default mode network in the visual task (Fischmeister et al., 2017; Martins, Fischmeister, et al., 2014); bilateral activity in STG for the musical domain (Martins et al., 2020); and bilateral activity in sensorimotor and premotor areas, basal ganglia and cerebellum in the motor domain (Martins, Bianco, et al., 2019)

These findings create an interesting puzzle. Our behavioral studies clearly suggest that the implementation of RHE across domains is supported by common cognitive machinery. However, our neuroimaging results clearly show that these tasks rely of separate, and domain-specialized neural circuits. To solve this apparent paradox, we recall our observation that the cognitive and neural processes necessary to acquire RHE rules might differ from those used for automatic processing after extensive training.

4.1 Acquisition vs. Automatic processing of RHE structures

Traditionally in fMRI experiments, participants are very well trained with the paradigm or task that they execute in the scanner. To increase the signal to noise ratio, it is sensible to reduce the cognitive variability among participants by excluding those who perform poorly, and by bringing participants’ performance close to ceiling in preliminary sessions. Unfortunately, this procedure introduces a bias. As we reviewed above for language, the patterns of neural activity change with training and experience (Jeon & Friederici, 2015; Skeide & Friederici, 2016). After extensive training, neural activity might reflect fast and automatic retrieval of previously formed visual, musical, or motor schemas. This access to procedural or declarative long-term memory information might be supported by brain networks crucially distinct from those supporting the *acquisition* of RHE.

Following this line of reasoning, we devised paradigms to test which brain regions might be critically involved in the *acquisition* of RHE, by using participants without prior experience with our paradigms. In the first study, we tested patients with brain lesions in the left hemisphere and mapped the brain lesions which disturbed the process of acquisition of RHE in the visual domain (Martins, Krause, et al., 2019). Because lesions can affect both accuracy and response times, we used a drift diffusion model to account for behavioral performance (Ratcliff & McKoon, 2008; Wiecki et al., 2013). With this model we can combine individual accuracy and response time to generate two sensible measures of processing. The first is called drift and is the speed with which participants can accumulate information from the stimuli. The second is called boundary separation. It measures how much information each patient collects before providing an answer and can be seen as a measure of impulsivity or cognitive control.

Using these methods, our main finding was that lesions in the posterior middle temporal gyrus negatively decreased the speed of information processing specifically associated with RHE (drift) (Martins, Krause, et al., 2019). In addition, we found that lesions in IFG negatively impacted cognitive control (boundary separation). This specific brain network is reminiscent to that used in the processing of language during the processes of acquisition (posterior temporal cortex) (Skeide et al., 2014) and expert processing (IFG) (Jeon & Friederici, 2015; Skeide et al., 2014), respectively. Crucially, this experiment also provided behavioral evidence for shared resources between visual and linguistic RHE, when controlling for general attention and working memory. In other words, when testing untrained participants, we find evidence for similar cognitive and neural resources used in the implementation of RHE in different domains.

In a second experiment, we tested participants on their ability to parse 2-level if-then logical hierarchies as described above (Scholz, 2020; Scholz et al., 2022). We contrasted the hierarchical condition (in which the first object determines how the second is interpreted) and the flat condition (in which the two objects are interpreted independently). Crucially, our participants were given very little training before the scanning session. Under these conditions, we found that the processing of hierarchies was supported by a left lateralized network comprising IFG and MFG, and pSTS (Scholz et al., 2022). Again, this (left lateralized) pattern of activity is reminiscent of syntactic processing in language. Interestingly, the higher the number of trails that participants completed, the higher the activity in IFG. This confirms that this region plays an essential role in expert processing of complex hierarchical stimuli. However, in this experiment, IFG expertise-related activity was confined to the right hemisphere.

Altogether, these findings suggest that the acquisition of RHE rules might be supported by cognitive and neural systems which are used across domains. However, the longer the training with these kinds of structures in each domain, the wider the separation of the neural systems supporting their automatized processing. This specialization of neural resources after training is consistent with what we know from the neuroscience of learning, and is essential to the capacity to retrieve previous experiences from long-term memory to solve similar problems in the present. Domain-specialized circuitry leads to faster and more efficient retrieval of information, and this is what distinguishes experts from non-experts.

5 Conclusion – cognitive and neural bases of fractal cognition

Fractals are complex structures with an unbounded number of hierarchical levels. However, they can be built with simple generative rules, that once mastered, allow us to tame complexity. The capacity to represent recursive hierarchical embedding (RHE) is key to our ability to represent fractals and other complex structures. As we reviewed, human adults can use RHE to generate fractal-like structures in language, vision, music and in the motor domains.

From the evolutionary point of view, this capacity is interesting because it seems uniquely human. Other species can represent simple hierarchies to navigate their social rank and for spatial navigation (Buzsáki & Moser, 2013; Seyfarth & Cheney, 2014), but these are limited in scope and depth. While a few non-human primates might learn to represent two hierarchical levels also in artificial languages, these few individuals require an extensive amount of training (Ferrigno et al., 2020). For humans, the acquisition of this capacity is also difficult, as it comes late in ontogenetic development. However, once it is available, it enables the representation of hierarchies of unlimited depth, and can be used in a multiplicity of domains.

Our review from the literature and our own results point to distinct cognitive and neural systems involved in the processing of RHE in humans (Fig. 4).

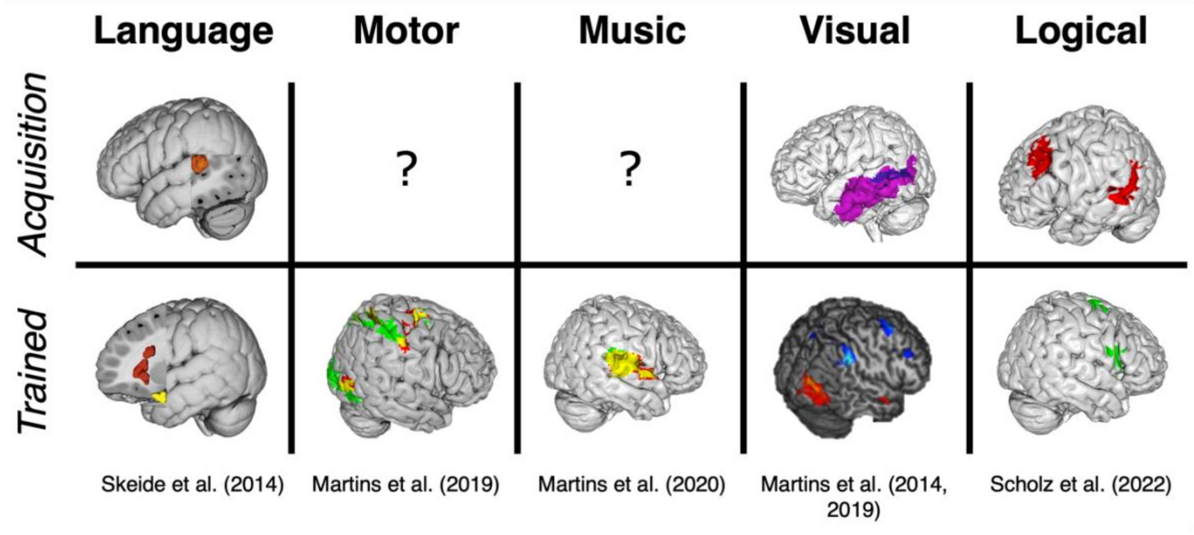

Fig. 4. Commonalities and differences in the processing of recursive hierarchical embedding across domains. The left posterior temporal cortex, spanning pSTS and adjacent areas, is critically involved in the acquisition of RHE across domains. After extensive training, RHE processing becomes supported by increasingly domain-specialized neural regions.

First, both in language and in the visual domains, mastery of RHE becomes available after 7-8 y.o. and it requires previous experience with conjunctive structures (Martins, Laaha, et al., 2014; Roeper, 2011). Interestingly, this is the age at which the accurate fasciculus matures to its distinctively human connectivity pattern (Rilling et al., 2008). Second, we found that the two main regions connected via this pathway – IFG and pSTS – were somewhat involved in the acquisition of RHE. Left posterior temporal cortex is clearly recruited in the acquisition of linguistic syntax (Skeide & Friederici, 2016), logical rules (Scholz et al., 2022), and visual fractals (Martins, Krause, et al., 2019), but also in the processing of musical fractals (Martins et al., 2020). On the other hand, activity in lateral frontal cortex is less robust and varies depending on the level of expertise (Jeon & Friederici, 2015; Martins, Krause, et al., 2019; Skeide et al., 2014). These neuroimaging findings fit well with the behavioral research suggesting strong co-variance across domains in the capacity to acquire RHE (Martins, Gingras, et al., 2017; Martins, Krause, et al., 2019; Scholz, 2020).

On the other hand, and in line with previous literature, we find that expert processing of RHE recruits increasingly specialized neural systems which including specific regions within left IFG for language (Fedorenko et al., 2011; Jeon & Friederici, 2015; Skeide et al., 2014), right IFG for logical hierarchies (Scholz et al., 2022), sensorimotor areas for motor fractals (Martins, Bianco, et al., 2019), bilateral STG areas related to melody processing for music fractals (Martins et al., 2020), and the visual ventral stream for visual fractals (Martins, Fischmeister, et al., 2014). These distinct patterns might reflect domain general mechanisms operating over increasingly domain specialized representations (Matchin et al., 2017), better suited for fast, automatized and modular processing (Asano et al., 2022).

Finally, in this chapter we only reviewed the role of cortical regions in the implementation of RHE representations. However, an increasing body of evidence has implicated subcortical regions in both the acquisition and expert representation of these structures. For instance, basal ganglia activity is associated with expert processing of RHE in language (Jeon et al., 2014), action (Martins, Bianco, et al., 2017), and logical reasoning (Scholz et al., 2022). On the other hand, the hippocampus is active in the processing of visual-spatial fractals (Martins, Fischmeister, et al., 2014), and especially involved in the processing of higher order elements in compositional structures (Scholz et al., 2022). These findings are consistent with the roles of the basal ganglia in expert processing hinging on the procedural memory system, and often connected with and mirroring IFG activity (Jeon et al., 2014); and with the role of the hippocampus in establishing novel compositional structures of the kind items-in-context (McKenzie et al., 2016; Ranganath, 2010).

Taken together, these findings suggest that the acquisition and expert processing of fractal structures results from the interplay between diverse neural structures and cognitive systems, such as those generally involved in procedural and declarative learning, conceptual reasoning, and cognitive control, but also regions encoding specialized representations. Mapping the precise roles and division of labor between these neural structures is an exciting endeavor for future research.

References

Amunts, K., & Zilles, K. (2012). Architecture and organizational principles of Broca's region. *Trends in Cognitive Sciences*, *16*(8), 418–426. https://doi.org/10.1016/j.tics.2012.06.005

Arata, Y., Shiga, I., Ikeda, Y., Jurica, P., Kimura, H., Kiyono, K., & Sako, Y. (2022). Insulin signaling shapes fractal scaling of C. elegans behavior. *Scientific Reports*, *12*(1), 10481. https://doi.org/10.1038/s41598-022-13022-6

Arsenijević, B., & Hinzen, W. (2012). On the Absence of X-within-X Recursion in Human Grammar. *Linguistic Inquiry*, *43*(3), 423–440. https://doi.org/10.1162/LING_a_00095

Asano, R., Boeckx, C., & Fujita, K. (2022). Moving beyond domain-specific versus domain-general options in cognitive neuroscience. *Cortex; a Journal Devoted to the Study of the*

Nervous System and Behavior, *154*, 259–268. https://doi.org/10.1016/j.cortex.2022.05.004

Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, *12*(5), 193–200. https://doi.org/10.1016/j.tics.2008.02.004

Berwick, R. C., & Chomsky, N. (2015). *Why Only Us: Language and Evolution*. The MIT Press.

Bianco, R., Novembre, G., Keller, P. E. E., Kim, S.-G. G., Scharf, F., Friederici, A. D., Villringer, A., & Sammler, D. (2016). Neural networks for harmonic structure in music perception and action. *NeuroImage*, *142*, 454–464. https://doi.org/10.1016/j.neuroimage.2016.08.025

Bianco, R., Novembre, G., Keller, P. E., Scharf, F., Friederici, A. D., Villringer, A., & Sammler, D. (2015). Syntax in action has priority over movement selection in piano playing: An erp study. *Journal of Cognitive Neuroscience*, *28*(1), 41–54. https://doi.org/10.1162/jocn_a_00873

Buckner, R. L., & Krienen, F. M. (2013). The evolution of distributed association networks in the human brain. In *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2013.09.017

Buzsáki, G., & Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience*, *16*(2), 130–138. https://doi.org/10.1038/nn.3304

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Annals of the New York Academy of Sciences*, *1169*(1), 448–458. https://doi.org/10.1111/j.1749-6632.2009.04582.x

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(39), 16428–16433. https://doi.org/10.1073/pnas.1112937108

Fedorenko, E., & Shain, C. (2021). Similarity of Computations Across Domains Does Not Imply Shared Implementation: The Case of Language Comprehension. *Current Directions in Psychological Science*, *30*(6), 526–534. https://doi.org/10.1177/09637214211046955

Ferrigno, S., Cheyette, S. J., Piantadosi, S. T., & Cantlon, J. F. (2020). Recursive sequence generation in monkeys, children, U.S. adults, and native Amazonians. *Science Advances*, *6*(26), 1–11. https://doi.org/10.1126/sciadv.aaz1002

Fischmeister, F. P., Martins, M. J., Beisteiner, R., & Fitch, W. T. (2017). Self-similarity and Recursion as Default Modes in Human Cognition. *Cortex*, *97*, 183–201.

Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition*, *97*(2), 179–210. https://doi.org/10.1016/j.cognition.2005.02.005

Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*, *1316*(1), 87–104. https://doi.org/10.1111/nyas.12406

Friederici, A. D., Chomsky, N., Berwick, R. C., Moro, A., & Bolhuis, J. J. (2017). Language, mind and brain. *Nature Human Behaviour*, *1*(10), 713–722. https://doi.org/10.1038/s41562-017-0184-4

Hagoort, P., & Indefrey, P. (2014). The Neurobiology of Language Beyond Single Words. *Annual Review of Neuroscience*, *37*, 347–362. https://doi.org/10.1146/annurev-neuro-071013-013847

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Jackendoff, R. (2003). Pr{é}cis of Foundations of Language: Brain, Meaning, Grammar, Evolution. *Behavioral and Brain Sciences*, *26*, 651–707.

Jackendoff, R. (2013). Foundations of Laguage. *Brain, Meaning, Grammer, Evolution*, *53*(9), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and What's specical about it? *Cognition*, *100*(1), 33–72. https://doi.org/10.1016/j.cognition.2005.11.005

Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language. *Cognition*, *97*(2), 211–225. https://doi.org/doi: 10.1016/j.cognition.2005.04.006

Jeon, H. A., Anwander, A., & Friederici, A. D. (2014). Functional Network Mirrored in the Prefrontal Cortex, Caudate Nucleus, and Thalamus: High-Resolution Functional Imaging and Structural Connectivity. *Journal of Neuroscience*, *34*(28), 9202–9212. https://doi.org/10.1523/JNEUROSCI.0228-14.2014

Jeon, H. A., & Friederici, A. D. (2013). Two principles of organization in the prefrontal cortex are cognitive hierarchy and degree of automaticity. *Nature Communications*. https://doi.org/10.1038/ncomms3041

Jeon, H. A., & Friederici, A. D. (2015). Degree of automaticity and the prefrontal cortex. *Trends in Cognitive Sciences*, *19*(5), 244–250. https://doi.org/10.1016/j.tics.2015.03.003

Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), 15443–15448. https://doi.org/10.1073/pnas.1300272110

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (Issue 7, pp. 112–131). Wiley. https://doi.org/10.1093/rfs/hhq153

Lerdahl, F., & Jackendoff, R. (1977). Toward a Formal Theory of Tonal Music. *Journal of Music Theory*, *21*(1), 111–171.

Lobina, D. J. (2017). *Recursion: A Computational Investigation into the Representation and Processing of Language*. Oxford University Press. https://doi.org/10.1093/oso/9780198785156.001.0001

Makuuchi, M., Bahlmann, J., Anwander, A., & Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(20), 8362–8367. https://doi.org/10.1073/pnas.0810928106

Mandelbrot, B. (1977). *The fractal geometry of nature*. Freeman.

Martins, M. J. (2012). Distinctive signatures of recursion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 2055–2064. https://doi.org/10.1098/rstb.2012.0097

Martins, M. J., Bianco, R., Sammler, D., & Villringer, A. (2017). Fractals in Action: An fMRI study on the Generation of new Hierarchical Levels in Motor Sequences. *BioRxiv*. http://biorxiv.org/content/early/2017/12/14/232942.abstract

Martins, M. J., Bianco, R., Sammler, D., & Villringer, A. (2019). Recursion in action: An fMRI study on the generation of new hierarchical levels in motor sequences. *Human Brain Mapping*, *40*(9). https://doi.org/10.1002/hbm.24549

Martins, M. J., Fischmeister, F. P., Puig-Waldm<unicode>252</unicode>ller, E., Oh, J., Gei<unicode>223</unicode>ler, A., Robinson, S., Fitch, W. T., & Beisteiner, R. (2014). Fractal image perception provides novel insights into hierarchical cognition. *NeuroImage*, *96*, 300–308. https://doi.org/10.1016/j.neuroimage.2014.03.064

Martins, M. J., Fischmeister, F. P. S., Gingras, B., Bianco, R., Puig-Waldmueller, E., Villringer, A., Fitch, W. T., & Beisteiner, R. (2020). Recursive music elucidates neural mechanisms supporting the generation and detection of melodic hierarchies. *Brain Structure and Function*, *0123456789*. https://doi.org/10.1007/s00429-020-02105-7

Martins, M. J., Gingras, B., Puig-Waldmueller, E., & Fitch, W. T. (2017). Cognitive representation of “musical fractals”: Processing hierarchy and recursion in the auditory domain. *Cognition*, *161*. https://doi.org/10.1016/j.cognition.2017.01.001

Martins, M. J., Krause, C., Neville, D., Pino, D., Villringer, A., & Obrig, H. (2019). Recursive Hierarchical Embedding in vision is impaired by posterior Middle Temporal Gyrus lesions. *Brain*, *142*(10).

Martins, M. J., Laaha, S., Freiberger, E. M. E. M., Choi, S., & Fitch, W. T. (2014). How children perceive fractals: Hierarchical self-similarity and cognitive development. *Cognition*, *133*(1), 10–24. https://doi.org/10.1016/j.cognition.2014.05.010

Martins, M. J., Martins, I. P., & Fitch, W. T. (2016). A novel approach to investigate recursion and iteration in visual hierarchical processing. *Behavior Research Methods*, *48*(4). https://doi.org/10.3758/s13428-015-0657-1

Matchin, W. G., Hammerly, C., & Lau, E. (2017). The role of the {IFG} and {pSTS} in syntactic prediction: Evidence from a parametric study of hierarchical structure in {fMRI}. *Cortex*, *88*, 106–123. https://doi.org/10.1016/j.cortex.2016.12.010

Matsuyama, T., & Matsushita, M. (1993). Fractal morphogenesis by a bacterial cell population. *Critical Reviews in Microbiology*, *19*(2), 117–135. https://doi.org/10.3109/10408419309113526

McKenzie, S., Keene, C. S., Farovik, A., Bladon, J., Place, R., Komorowski, R., & Eichenbaum, H. (2016). Representation of memories in the cortical–hippocampal system: Results from the application of population similarity analyses. In *Neurobiology of Learning and Memory*. https://doi.org/10.1016/j.nlm.2015.12.008

Odifreddi, P. (1989). Classical Recursion Theory. In *Classical Recursion Theory* (Vol. 125, p. xviii+668). Elsevier. https://doi.org/10.2307/2687828

Perfors, A., Tenenbaum, J. B., Gibson, E., & Regier, T. (2010). How recursive is language? A Bayesian exploration. In H. van der Hulst (Ed.), *Recursion and Human Language* (pp. 159–175). de Gruyter Mouton.

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, *118*(3), 306–338.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what’s special about it? *Cognition*, *95*(2), 201–236. http://www.sciencedirect.com/science/article/pii/S0010027704001763

Ranganath, C. (2010). A unified framework for the functional organization of the medial temporal lobes and the phenomenology of episodic memory. *Hippocampus*, *20*(11), 1263–1290. https://doi.org/10.1002/hipo.20852

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. In *Neural Computation*. https://doi.org/10.1162/neco.2008.12-06-420

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., & Behrens, T. E. J. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience*, *11*(4), 426–428. https://doi.org/10.1038/nn2072

Roeper, T. (2007). *The Prism of Grammar: How Child Language Illuminates Humanism*. MIT Press.

Roeper, T. (2009). The minimalist microscope: How and where interface principles guide acquisition. *Proceedings of the 33rd Annual Boston University Conference on Language Development*. http://people.umass.edu/roeper/online%7B_%7Dpapers/MIcroscopic

Roeper, T. (2011). The Acquisition of Recursion: How Formalism Articulates the Child's Path. *Biolinguistics*, *5*(1–2), 57–86. http://www.biolinguistics.eu/index.php/biolinguistics/article/view/178

Rohrmeier, M., & Koelsch, S. (2012). Predictive information processing in music cognition. A critical review. *International Journal of Psychophysiology*, *83*(2), 164–175. https://doi.org/10.1016/j.ijpsycho.2011.12.010

Ruiz-Padial, E., & Ibáñez-Molina, A. J. (2018). Fractal dimension of EEG signals and heart dynamics in discrete emotional states. *Biological Psychology*, *137*, 42–48. https://doi.org/https://doi.org/10.1016/j.biopsycho.2018.06.008

Scholz, R. (2020). *A novel paradigm for testing the initial coding of hierarchical relationships within the medial temporal lobe in a circuit specific manner*. Humboldt University.

Scholz, R., Villringer, A., & Martins, M. J. D. (2022). Distinct hippocampal and cortical contributions in the representation of hierarchies. *BioRxiv*. https://doi.org/10.1101/2022.06.15.496057

Seyfarth, R. M., & Cheney, D. (2014). The evolution of language from social cognition. In *Current Opinion in Neurobiology* (Vol. 28, pp. 5–9). Elsevier Ltd. https://doi.org/10.1016/j.conb.2014.04.003

Skeide, M. A., Brauer, J., & Friederici, A. D. (2014). Syntax gradually segregates from semantics in the developing brain. *NeuroImage*, *100*. https://doi.org/10.1016/j.neuroimage.2014.05.080

Skeide, M. A., & Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nature Reviews Neuroscience*, *17*(5), 323–332. https://doi.org/10.1038/nrn.2016.23

Tiwari, A., Villatte, J. L., Narayanan, S., & Falk, T. H. (2019). Prediction of Psychological Flexibility with multi-scale Heart Rate Variability and Breathing Features in an "in-the-wild" Setting. *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, 297–303. https://doi.org/10.1109/ACIIW.2019.8925281

Udden, J., Martins, M. J., Zuidema, W., & Fitch, W. T. (2019). Hierarchical structure in sequence processing: how do we measure it and what's the neural implementation? *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12442

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*(August), 1–10. https://doi.org/10.3389/fninf.2013.00014

Zaccarella, E., Meyer, L., Makuuchi, M., & Friederici, A. D. (2015). Building by Syntax: The Neural Basis of Minimal Linguistic Structures. *Cerebral Cortex*, bhv234--. https://doi.org/10.1093/cercor/bhv234