COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 49 (2025) e70108

ISSN: 1551-6709 online
DOI: 10.1111/cogs.70108

Acquisition and Utilization of Recursive Rules in Motor Sequence Generation

Maurício D. Martins,[a,b,**] Zoe Bergmann,[a,**] Elena Leonova,[c]
Roberta Bianco,[d] Daniela Sammler,[b,e] Arno Villringer[b,f,g]

[a] *SCAN-Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna*
[b] *Max Planck Institute for Human Cognitive and Brain Sciences*
[c] *Institute for Cognitive Studies, Saint Petersburg State University*
[d] *Neuroscience of Perception & Action Laboratory, Italian Institute of Technology*
[e] *Research Group Neurocognition of Music and Language, Max Planck Institute for Empirical Aesthetics*
[f] *Berlin School of Mind and Brain, Humboldt Universität zu Berlin*
[g] *Clinic for Cognitive Neurology, University Hospital Leipzig*

Received 7 July 2024; received in revised form 14 July 2025; accepted 31 July 2025

Abstract

Recursive hierarchical embedding allows humans to generate multiple hierarchical levels using simple rules. We can acquire recursion from exposure to linguistic and visual examples, but only develop the ability to understand “multiple-level” structures like “[[second] red] ball]” after mastering “same-level” conjunctions like “[second] *and* [red] ball.” Whether we can also learn recursion in motor production remains unexplored. Here, we tested 40 adults’ ability to learn and generate sequences of finger movements using “multiple-level” recursion and “same-level” iteration rules (like linguistic conjunction). Rule order was counterbalanced. First, they learned the generative rules (without explicit rule instructions or feedback) by executing examples of motor sequences based on visual cues displayed on the screen (learning). Second, participants were asked to discriminate between correct and incorrect motor sequences beyond those to which they were previously exposed (discrimination). Finally, they were asked to use the rules to generate new hierarchical levels consistent with the

---

**Shared first authors.

Correspondence should be sent to Maurício D. Martins, SCAN-Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Liebiggasse 5, 1090 Wien, Austria. E-mail: mauricio.martins@univie.ac.at

previously given (generation). We repeated the procedure (all three phases) on 2 days, allowing for a night of sleep. We found that most participants could discriminate correct/incorrect sequences based on recursive rules and use recursive rules to generate new hierarchical levels in motor sequences, but mostly on the second day of testing, and when they had acquired iterative before recursive rules. This aligns with previous literature on vision and language and with literature showing that sleep is necessary to generate abstract knowledge of motor sequences. Lastly, we found that the ability to discriminate well-formed motor sequences using recursion was insufficient for motor generativity.

Keywords: Recursion; Iteration; Motor production; Procedural learning; Rule acquisition

1. Introduction

The capacity to represent and generate multilayered hierarchical structures is a fundamental aspect of human cognition available in language (Berwick, Friederici, Chomsky, & Bolhuis, 2013; Chomsky, 1957; Friederici, Chomsky, Berwick, Moro, & Bolhuis, 2017), music (Cheung, Meyer, Friederici, & Koelsch, 2018; Jackendoff & Lerdahl, 2006; Koelsch, Rohrmeier, Torrecuso, & Jentschke, 2013; Lerdahl & Jackendoff, 1983), complex action sequencing (Badre & D'Esposito, 2009; Fitch & Martins, 2014; Lashley, 1951), theory of mind (De Villiers, Hobbs, & Hollebrandse, 2014; De Villiers & De Villiers, 2014), and social organization (Ferguson, 2018; Redhead & Power, 2022; Seyfarth & Cheney, 2017). In these domains, hierarchies can be generated using recursive hierarchical embedding, which yields complex structures when applying simple embedding rules to its own output (Fitch, 2010; Martins, 2012).

Recursion is a term used in various fields with multiple definitions (Fitch, 2010; Lobina, 2011; Martins, 2012; Martins & Fitch, 2014). Recognizing this multiplicity, it is essential to specify how we are using it, even if other definitions are possible within cognitive sciences. In particular, we focus on recursive hierarchical embedding, which has two components: embedding and recursion. Hierarchical embedding is a process through which an element, or set of elements, is made "subordinate" to another "dominant" element. For instance, in English, when the word "film" is embedded in "committee" to form [[film] committee], it refers to a kind of committee, not a kind of film. Recursion is the process through which a function's output is used again as input to the same function. For instance, the natural numbers are described by the recursive function $N_i = N_{i-1} + 1$, which generates the infinite set $\{1, 2, 3,\ldots\}$. By combining these two properties—recursion and hierarchical embedding—we can generate hierarchies of unbounded depth. For instance, by using the recursive embedding rule NP→ [[NP] NP], we can add "student" to "film committee" and obtain [[[student] film] committee] and so on.

Also crucial is clarifying the level of analysis relevant for empirical investigation. As reviewed elsewhere, recursion has been discussed at the levels of (1) generative processes, (2) stimuli, and (3) cognitive representations (Lobina, 2011; Martins, 2012). Empirically, the first two levels are ambiguous because generative processes are often opaque, and embedded structures may also be generated without complex cognition, for example, tree branches and bacterial growth patterns. For this reason, we focus on whether participants can extract the regularities of hierarchical stimuli to generate new levels beyond the given (Martins, 2012).

Recursion has long been considered a core component of human cognition. While initially thought to be specific to linguistic syntax (Berwick & Chomsky, 2016; Chomsky, 1957; Friederici et al., 2017; Hauser et al., 2002), more and more evidence suggests that it is a widely available mechanism to parse symbolic structures, allowing mental programs to run across domains (Asano, Boeckx, & Seifert, 2021; Dehaene, Al Roumi, Lakretz, Planton, & Sablé-Meyer, 2022). For example, in language, recursion is not only present in syntax but also at the level of discourse (Everett, 2012; Levinson, 2013). Beyond language, recursive rules have been proposed to create intricate musical compositions (Lerdahl & Jackendoff, 1983; Rohrmeier, 2011) and complex visual structures (Fischmeister, Martins, Beisteiner, & Fitch, 2017; Martins, 2012; Martins et al., 2014; Martins, Laaha, Freiberger, Choi, & Fitch, 2014). Moreover, recursion has been proposed as a cognitive mechanism that enables humans to plan elaborate sequences of actions (Jackendoff, 2009, 2011). While recursion has been empirically studied in the context of language, vision, music, and action planning, its role in motor production remains relatively unexplored (Martins, Bianco, Sammler, & Villringer, 2019). This paper investigates the acquisition and utilization of recursive rules in producing motor sequences, extending our understanding of hierarchical cognition. It also offers a novel paradigm to the research program of determining whether recursion is domain-general or multi-domain-specific by systematically mapping the domains in which it is available and the commonalities and differences across domains.

Investigations into recursion have often utilized artificial grammar learning (AGL) paradigms (Reber, 1967) to explore the ability to learn and discriminate self-embedded structures in humans and other animals (Ferrigno, Cheyette, Piantadosi, & Cantlon, 2020; Fitch & Hauser, 2004; Gentner, Fenn, Margoliash, & Nusbaum, 2006; Liao, Brecht, Johnston, & Nieder, 2022). For example, parsing the string AABB as a pattern $A_1A_2B_2B_1$ requires representing long-distance dependencies $A_1 \ldots B_1$ that go beyond the adjacent level, crucially distinguishing the underlying representation from a simple sequence (Uddén, Martins, Zuidema, & Fitch, 2020). Furthermore, the ability to generalize the pattern A^nB^n to levels beyond the given—for example, $A_1A_2A_3B_3B_2B_1$—further suggests that the underlying generative grammar is being represented.

In the early work, only humans were shown to acquire the underlying center-embedded representations of the kind A^nB^n (Fitch & Hauser, 2004). However, recent studies involving monkeys (Ferrigno et al., 2020) and crows (Liao et al., 2022) have revealed that these species can perform well in these tasks, albeit after extensive training. The task developed in these experiments is particularly powerful, requiring participants to actively generate motor sequences with an underlying A^nB^n structure by tapping figures on a touch screen in the correct order.

The conclusion that animals can represent recursion, however, has been disputed. First, the acquisition patterns suggest that simpler cognitive mechanisms, such as associative learning, might better explain animal performance (Rey & Fagot, 2023). Second, Recurrent Neural Network models suggest that nonrecursive processes can also represent long-distance dependencies after extensive training (Lakretz & Dehaene, 2021). Third, AGL paradigms have additional limitations, as center-embedded structures are challenging to comprehend if reaching three to four embeddings (Christiansen & Chater, 1999) because they impose significant working memory demands. Crucially, the human ability to parse center-embedded structures correlates with working memory capacity (Ferrigno et al., 2020). Finally, both recursive and nonrecursive (iterative) computational processes (Lobina, 2011; M. D. Martins, Martins, & Fitch, 2016; van der Hulst, 2010) can formally generate center-embedded structures. This is understood as the competence-performance problem and posits that adequate *performance* in processing center-embedded structures is insufficient evidence for the *competence* of recursion, that is, of an underlying recursive computation (Lobina, 2011). However, the capacity to parse long-distance dependencies demonstrates at least the ability to form graph representations above simple sequences (Uddén et al., 2020).

To circumvent these limitations and focus on the competence for recursion, we developed a general approach based on generating hierarchical fractal structures (Martins, 2012; Martins et al., 2016). Fractals have similar structural properties across multiple hierarchical levels and can be generated by repeatedly applying the same recursive rules (Mandelbrot, 1977; Martins, 2012). Interestingly, they can also be generated using more straightforward "iterative" nonrecursive rules (Martins, 2012; Martins et al., 2016), which add multiple elements within the same hierarchical level without generating new levels. Our approach contrasts how identical fractal structures are represented when generated by different rules—recursion and iteration. Importantly, while fractals tend to infinity, we use the term here as a placeholder for a structure that exhibits self-similarity across hierarchical levels. While an approximation, this placeholder is useful for empirical investigation.

This approach has been promising for isolating the behavioral and neural processes specifically involved in representing recursive generative rules across domains (Martins, 2012; Martins et al., 2016). For example, principal component analysis has shown that the ability to learn and use recursive rules in music, vision, and action planning clusters together within a component orthogonal to domain-specific nonrecursive tasks (Martins, Gingras, Puig-Waldmueller, & Fitch, 2017). This suggests common underlying cognitive mechanisms for recursion across domains. Furthermore, we have observed a correlation between the ability to parse 2-level center-embedded linguistic sentences and visual recursion (Martins et al., 2019). Our neuroimaging data also suggest that the neural networks active in acquiring recursive and hierarchical rules in the visual and logic domains are similar to those involved in linguistic recursion (Martins et al., 2019; Scholz, Villringer, & Martins, 2023).

Finally, our research in the visual domain has revealed an interesting developmental trajectory that mimics language. In language, the acquisition of recursion in several linguistic forms (e.g., [[second] red] ball), including adjectives, possessives, verbal compounds, and sentence complements (Roeper, 2011; Yang, Hu, Fan, Dong, & Jeschull, 2022) crucially scaffolds the prior acquisition of conjunction—the understanding of words as at the same hierarchical level (e.g., [second] and [red] ball). In other words, prior experience with "same-level" structures is necessary for acquiring recursion and creating hierarchical dependencies. Similarly, in vision, we have found that children's acquisition of "multiple-level" visual recursion dramatically improves after the experience with "same-level" visual iteration, which adds elements within a fixed hierarchical level without generating new levels (Martins et al., 2014). Attempts to acquire recursion before iteration have shown a detrimental impact on recursion performance. Another salient observation is that while recursive rules in vision are more challenging to learn than iteration, once acquired, they seem to be more efficient in (1) reducing demands in visuospatial working memory (Martins et al., 2016), (2) enhancing the capacity to detect fine-grained mistakes (Martins et al., 2014), and (3) reducing neural activity in the frontoparietal regions associated with cognitive effort (Fischmeister et al., 2017).

We extend this work to the domain of motor recursion. While the hierarchical organization of goal-oriented action and its cognitive underpinnings has been previously characterized (Badre, 2008; Fazio et al., 2009; Koechlin & Jubault, 2006; Pulvermüller & Fadiga, 2010), it is unclear whether humans can also acquire the representation of hierarchically organized motor sequences when these are not embedded within goal hierarchies. Previous research shows that humans can utilize recursive rules to generate motor fractals when explicitly given the rules and after extensive training (Martins et al., 2019). This capacity correlates with activity in brain areas involved in motor imagery. Crucially, here, we test whether humans can also acquire those representations when the rules are not explicitly provided and no trial feedback is given. Furthermore, we characterize the underlying learning dynamics across 2 days as abstract representations of motor sequences are facilitated after sleep consolidation (Fischer, Nitschke, Melchert, Erdmann, & Born, 2005; Fogel et al., 2017).

As mentioned above, it is particularly challenging to devise convincing empirical approaches to differentiate between hierarchical representations and alternatives such as basic iteration or sequential processing (Uddén et al., 2020). It has recently been proposed that studies aiming at isolating cognitive processes associated with hierarchical cognition should include (1) comparing hierarchical and nonhierarchical models and (2) testing generative abilities beyond simple recognition tasks (Dedhe, Clatterbuck, Piantadosi, & Cantlon, 2023). Due to their perceptual nature, the visual and auditory domains do not easily lend themselves to a generative task beyond discrimination. Here, we can overcome this limitation with a motor recursion paradigm.

Based on the background above, this study aims: (1) to test whether human adults can acquire recursive rules in the motor domain from practicing examples without feedback; (2) to investigate their post-learning ability to discriminate between motor sequences that follow or violate the recursive rule; and (3) to assess whether participants can use the acquired recursive rule productively to generate motor sequences.

We hypothesize that humans can acquire the ability to generate motor sequences using recursion without prior explanation of the rules or trial feedback, similar to language, vision, and music findings. We employ a motor recursion task and compare it with an iteration task, both previously used in an functional magnetic resonnance imaging (fMRI) procedure where participants had been previously instructed and heavily trained on the rules (Martins et al., 2019). We predict that the recursion learning curves will be steeper than iteration, reflecting the cognitive demands of understanding and generating recursive structures. Additionally, we anticipate that acquiring iteration before recursion will facilitate the learning process. Lastly, we expect performance to improve on the second day of testing following a sleep period, capitalizing on the role of sleep in consolidating motoric representations.

2. Methods

2.1. Participants

We tested 40 healthy participants, including 17 males and 23 females, aged 18–38 ($M = 29$). All participants were nonmusicians, had normal or corrected-to-normal vision and audition, were right-handed, and had no neurological or psychiatric disease history. All participants provided written informed consent according to the local ethics committee guidelines. This work was approved under the ethics committee project number 016-15-26012015 at the Max Planck Institute for Human Cognitive and Brain Sciences.

2.2. Tasks and stimuli

Participants saw sequences of keypresses on a schematic keyboard on a computer screen. They were asked to repeat or complete these sequences on a 16-key MIDI piano keyboard (see Fig. 1 caption for video examples). The keyboard did not produce sound, but contained visual and tactile markers on specific keys (3, 5, 7, 10, 12, and 14) for spatial reference. Participants were instructed to produce the correct keypress sequences with the correct fingers and follow the correct temporal structure (see details in Fig. 1). The temporal structure of the sequences was given by an auditory metronome sounding at 60 bpm (4 beats per second).

2.3. Trial structure

The typical trial was composed of three steps (I, II, and III), and the relation between steps was determined by one of two rules: iteration or recursion (Fig. 1b,c). Participants were asked to press the keyboard buttons following the sequences they saw on the screen as they were presented (i.e., simultaneously). These sequences were visually depicted by overlaying colored circles onto the keys of a virtual keyboard. Different colors denoted the fingers participants used to press the keys (red/thumb, green/index, and blue/middle finger). Crucially, both physical and virtual keyboards were silent, generating no auditory tones.

The application of different rules always generated a final motor sequence of the kind $[[K - 2s, K - s, K], [K - s, K, K + s], [K, K + s, K + 2s]]$. To increase stimulus variability, we introduced the free parameters s and k_0. Parameter s could be a value within the set $\{-2, -1, 1, 2\}$. If s was positive (1 or 2), the sequences were *ascending*, meaning that they unfolded from left to right on the keyboard (e.g., $[K-1, K, K+1]$). If s was negative (-1 or -2), the sequence was *descending*, meaning that sequences unfolded from right to left on the keyboard (e.g., $[K+1, K, K-1]$). When $s = 1$ or -1, the sequence (within each cluster) was formed by *adjacent* keys ($[K+1, K, K-1]$), and when $s = 2$ and -2, the sequence was formed by *nonadjacent* keys, meaning that there was a space of one key between a pair of elements within the cluster (e.g., $[K-2, K, K+2]$). The initial key k_0 could be one of the middle four keys of the keyboard $\{7, 8, 9, 10\}$. Overall, these variations produced 16 different sequences balanced across conditions. Crucially, participants had to extract these free parameters during steps I and II to correctly generate the sequence in step III.

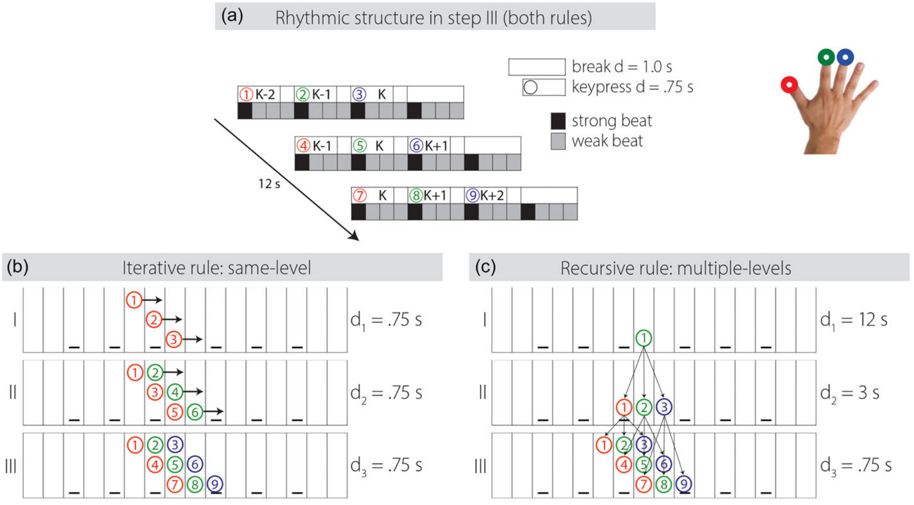

Fig. 1. Task principles. Participants were asked to generate sequences of keypresses on a keyboard with the thumb, index, and middle finger (denoted by red, green, and blue, respectively). Keypress order is denoted here as numbers, which were not shown to the participants. Sequences were formed in three steps (I, II, and III), which followed one of two rules: Iteration and Recursion. Participants executed the motor sequence displayed on the screen during steps I and II. In step III, they were asked to generate the final sequence without visual support. (a) Temporal structure in step III: In step III, both rules resulted in the same complete sequence of nine keypresses, here [[K−2, K−1, K], [K−1, K, K+1], [K, K+1, K+2]]. K is the key in the spatial center of the pattern. The sequence lasted 12 s and was aligned with a metronome with four beats per second (one strong and three weak). Keypresses started at the onset of the strong beat and were released at the offset of the third weak beat; thus, each keypress had the duration (d) = .75 s followed by a short pause of .25 s. Hierarchical clustering within the sequence (three clusters of three items) was marked by a 1-s break after each cluster, as well as by the finger pattern (red, green, and blue). (b) Iterative rule: Step I was composed of three keypresses executed with the thumb (red) on the first (strong) beat of each cluster [[K−2, _, _], [K−1, _, _], [K, _, _], each with d = .75 s. In step II, a second keypress with the index was added to each chunk: [[K−2, K−1, _], [K−1, K, _], [K, K+1, _]]. Thus, the iterative rule added elements within the same hierarchical level without generating new levels. Step III was simply the serial completion of the pattern with the middle finger [[K−2, K−1, K], [K−1, K, K+1], [K, K+1, K+2]]. Importantly, all sequences (in steps I, II, and III) had a total duration of 12 s. Here, __ denotes a period of four beats equivalent to (but without) a keypress. The horizontal arrow denotes that, in this example, within-level keypresses are added in ascending order from left to right. See https://osf.io/gbd8j for video trial examples of both ascending and descending order. (c) Recursive rule: Step I was a single keypress with the index finger (green) on key K. In the figure example, step II was a sequence of three keypresses [K−1, K, K+1] executed with the thumb (red), index (green), and middle finger (blue). Step III was a sequence of nine keypresses sequence [[K−2, K−1, K], [K−1, K, K+1], [K, K+1, K+2]]. As denoted by the vertical arrows, the underlying recursive rule was the substitution of each keypress K_n with a sequence of three keypresses $[K_{n+1} - n, K_{n+1}, K_{n+1} + n]$. In general, the recursive transformation rule was $K_n \rightarrow [K_{n+1} - s, K_{n+1}, K_{n+1} + s]$. K is the reference key pressed in step I, and $s = \{-2, -1, 1, 2\}$ denotes the distance and direction of keypresses relative to the reference K. In the example above, $s = 1$; thus, the sequence is composed of adjacent keypresses executed from left to right (see the text for details). In addition to the key, the recursion also operated in the dimension of rhythm to facilitate the representation that each cluster of three keypresses (and rhythmic breaks) was "nested" within a higher-level keypress of equal total duration (see Fig. 1A for rhythmic structure). In step I, the keypress duration $d_1 = 12$ s. In step II, each keypress had a duration $d_2 = 3$ s followed by a 1 s break. In step III, each keypress had the duration $d_3 = .75$ s followed by a break of .25 s *within* cluster, and a break of 1 s *between* clusters. Thus, the general temporal rule of key presses and within-level breaks was $d_{n+1} = d_n /4$. See https://osf.io/6q5jm for video trial examples.

2.4. *Procedure*

Broadly, tasks, stimuli, and trial structure were similar to those used in Martins et al. (2019). However, in this prior study, participants were explicitly given the task rules, shown illustrative examples, and trained with verbal feedback and corrections by the experimenter between sessions within a single day. In the current experiment, participants first executed correct examples without prior explanation of the rule or trial feedback. Then, rule acquisition was tested by (1) *discrimination* of correct/incorrect step III sequences and (2) uncued *motor generation* of the sequence in step III.

In this experiment, participants were tested on 2 separate days. During each testing day, they performed two tasks (recursion and iteration), each with a duration of 1 h and each composed of three phases: Training, Discrimination, and Generation. In the training phase, participants were asked to execute sequences of finger movements by following visual cues displayed on the screen and in sync with the metronome from step I to step III. In the discrimination phase, they were again instructed to press the buttons along with the visual cues synchronized with the beats from steps I to III, and afterward, they were asked to evaluate whether step III sequences were correct or incorrect. Finally, in the generation phase, they followed steps I and II synchronized with the metronome and visual cues, and then were asked to generate the correct step III without the support of visual cues.

The generation phase was the crucial test to evaluate whether participants had acquired the ability to use recursive and iterative rules to generate motor sequences. The training and discrimination phases allowed us to map further whether a potential inability to do so was due to (1) motor learning impairment (failure in training), (2) rule representation impairment (discrimination phase), or (3) active generativity impairment (generation phase). Importantly, during the training phase, participants were only exposed to sequences with adjacent keypresses, meaning with $s = \{-1, 1\}$ (e.g., https://osf.io/ekb7n; https://osf.io/awuqr). In the other two phases, however, they were asked to go beyond the training sample and discriminate/generate sequences also containing nonadjacent keypresses, that is, $s = \{-2, 2\}$ (e.g., https://osf.io/fs8nr; https://osf.io/dqrvn).

Participants were randomly divided into two groups of 20 each. The first group started with iteration followed by recursion (I-R), and the second started with recursion followed by iteration (R-I). Each group followed the same order on both days of testing.

2.4.1. *Training phase*

Video depictions of Training trials can be found for Iteration (https://osf.io/gbd8j) and Recursion (https://osf.io/6q5jm). Participants were asked to press the correct keys with the correct fingers, as demonstrated on the screen, synchronized both with the visual cues (colored circles) and with the metronome beats. Crucially, keypresses did not generate auditory tones. The training phase comprised 16 trials, in which there was always visual assistance in producing steps I, II, and III. Participants were instructed to understand the rules underlying the sequences' formation. They were told there was only one general rule for all sequences during the recursive task and only one rule for all sequences during the iterative

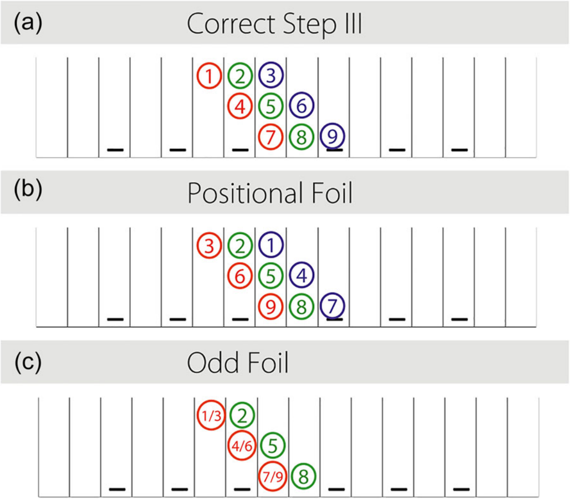

Fig. 2. Foil categories. (a) Correct step III. (b) In the positional foil trials, step III was correct relative to the position of the keys but incorrect relative to the keypress direction. Step II could unfold from left to right at the between-cluster level—[[_, K−1, _], [_, K, _], [_, K+1, _]]—but the third step would unfold from right to left at the within-cluster level—[[K, K−1, K−2], [K+1, K, K−1], [K+2, K+1, K]]. This foil was introduced to check whether participants were sensitive to the coherence of directionality between local and global levels, a hallmark of the “multiple-level” recursive representations. (c) In the odd foil trials, the third keypress of each cluster was a repetition of the first. Crucially, [[K−2, K−1, K−2], [K−1, K, K−1], [K, K+1, K]] could be a plausible continuation of [[K−2, K−1, _], [K−1, K, _], [K, K+1, _]], but it violated the underlying iterative rule. Due to these foil specificities, we expected participants to perform worse in the positional foils in the recursive trials and worse in the odd foils in the iteration trials.

task. Crucially, participants were not told about the concepts of recursion and iteration, nor were these terms presented during testing.

The first goal of this phase was to train participants in executing the motor sequences. The second goal was to expose them to the underlying generative rules without explicit rule instructions or trial feedback and test whether motor performance (keypress accuracy and synchrony with the metronome) improved with time. We tested whether accuracy and synchrony improved across trials and between day 2 and day 1. We also tested whether these measures differed between tasks (recursion vs. iteration) and task-order conditions (R-I vs. I-R).

2.4.2. *Discrimination phase*

Video depictions of Discrimination trials are available for Iteration (https://osf.io/83nwq) and Recursion (https://osf.io/2gd4k). The procedure was similar to the training phase, except that step III could either be a correct continuation of step II by applying the recursive or iterative rule or be a violation (foil). There were two foil categories: odd and positional (Fig. 2). While both were wrong, odd foils were a more plausible continuation of step II in iteration, and position foils were a more plausible continuation of step II in recursion. In the case of odd foils, step III started with the same sequence of keys (within each cluster) pressed in step II in the iteration. Correct discrimination required the understanding that cluster completion should follow the same within-cluster direction as step II. In the case of positional foils, step III followed the same between-cluster direction as step II in the recursive condition. In contrast, the within-cluster direction was incoherent with that of the between-cluster. Thus, correct discrimination required an understanding of between- and within-cluster level coherence and, consequently, of the hierarchical relationship between levels.

Overall, there were 32 discrimination trials. Step III was well-formed in 16 trials; in the remaining 16, step III was a violation (eight odd and eight positional foils). Similar to the training phase, participants were instructed to execute the sequence presented on the screen and synch with the metronome. At the end of the trial, they were asked to determine whether step III was a correct continuation of step II, according to the previously trained rules (iteration or recursion). After step III, the screen presented the question: "Was the last step of the sequence correct or incorrect?". Participants answered by pressing one of two keys on the keyboard.

The goal of the discrimination phase was to test whether participants formed a sufficient understanding of the underlying rules after the training phase, which would allow them to classify foils versus well-formed structures above chance. We also tested for differences in discrimination accuracy between days (1 vs. 2), tasks (recursion vs. iteration), and task order (I-R vs. R-I).

2.4.3. Generation phase

Video depictions of Generation trials are available for Iteration (https://osf.io/7w5ps) and Recursion (https://osf.io/qh93n). The procedure was similar to the training phase, except that in this phase, performance on step III was not guided by visual cues. Thus, after steps I and II, participants had to generate step III using the correct rule (recursion or iteration) and remaining in synch with the metronome (Fig. 1a). In step III, the screen depicted an empty keyboard, and participants were given a 1-s visual warning cue before the metronome started. There was a total of 16 generation trials. This phase evaluated whether participants could use the underlying rules to generate well-formed structures. We also tested for accuracy differences between days (1 vs. 2) tasks (recursion vs. iteration), and task orders (I-R vs. R-I).

2.5. Analyses

2.5.1. Training, discrimination, and generation

All analyses were performed in Rstudio (R 1.79). Model details can be found in the Supplementary Materials. Code and datasets are available in the OSF repository https://osf.io/dbvrn. For the training and generation phases, we coded each keypress as "correct" if the correct key was pressed within a window of [−.25; 1] seconds relative to the metronome beat. Otherwise, the keypress was coded as "incorrect." We also measured the temporal latency between the metronome beat and the keypress for the generation phase. For the discrimination phase, we recorded whether participants correctly discriminated between correct versus incorrect trials. For all analyses, only step III data were included.

We compared keypress accuracy between days (1 vs. 2), tasks (recursion vs. iteration), and task orders (I-R vs. R-I) by running generalized linear mixed models using the function glmer() from the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015). The dependent variables were the correctness of each keypress (correct vs. incorrect) for the training and generation phases and the correctness of each discrimination trial (correct vs. incorrect) for the discrimination phase. We used a logit link function (binomial family) for both analyses. Because there were nine keypresses per trial (16 trials), there were 144 data points per participant for training and generation. In the discrimination phase, there were only 32 data points per subject (one per trial). We also included the variable "trial" (1−16) in the training phase models to quantify the motor learning slopes across trials. Since this was not the focus of the other phases, and there was a ceiling effect toward the end of the training, the variable "trial" was omitted from the discrimination and generation models. The discrimination models included the variable foil type (correct, odd, and positional) to ensure that participants could reject different foils correctly.

The full models for training, discrimination, and generation were, respectively:

$$\text{Accuracy} = \left(\text{Task} \times \text{Day} \times \text{Order} \times \text{Trial}\right) + (1 + \text{Task} + \text{Day} + \text{Trial}|\text{Participant})$$

$$\text{Accuracy} = \left(\text{Task} \times \text{Day} \times \text{Order} \times \text{Foil}\right) + (1 + \text{Task} + \text{Day} + \text{Foil}|\text{Participant})$$

$$\text{Accuracy} = \left(\text{Task} \times \text{Day} \times \text{Order}\right) + (1 + \text{Task} + \text{Day}|\text{Participant})$$

We ran the full models up to the highest possible interactions for all analyses, then performed model selection with the function drop1() from the package *stats* (R Core Team, 2023). In all cases, the best fit was the model without the highest-order interaction (see results and Supplementary Materials). To analyze the significant effects, we computed pairwise comparisons of means and slopes using the package emmeans() (Russell, 2018). All pairwise comparisons were Bonferroni corrected.

We also measured the proportion of trials for which participants correctly pressed 9/9 keys. Considering four free parameter s combinations $\{-2, -1, 1, 2\}$ (ascending vs. descending; adjacent vs. nonadjacent keys), the probability that participants applied the correct one was 25%. With 16 trials, participants who correctly pressed 9/9 keys in eight trials performed statistically above chance (*Binomial test with a probability of 25%*: $z = 2.02$, $p = .03$).

2.5.2. *Beat-keypress latency*

We performed exploratory analyses on the beat-keypress latency differences across the nine keypresses composing the generation trials. Previous research suggests a higher latency of keypresses at the beginning of each cluster in motor sequences, indicating the retrieval process of a hierarchical motor schema from long-term memory (Moss, Zhang, & Mayr, 2023). To replicate this analysis (Moss et al., 2023), we tested the effects of "chunk transition," that is, the time latency difference between the 1st key of each cluster versus the other keys within the cluster (2nd or 3rd). To do so, we encoded the first key of each cluster as a dummy category "1st" and the other keys as a second category "2nd | 3rd." We then ran a generalized linear mixed model testing for the time latency difference between these two categories ("1st" vs.

"2nd | 3rd") and its interaction with task differences (recursion vs. iteration). Similar to Moss et al. (2023), our model included task random slopes and participant random intercepts. In other words, we ran the model:

$$\text{Beat - keypress latency} = (\text{Task} \times \text{Key}[1\text{stvs.}''2\text{nd}|3\text{rd}'']) + (1 + \text{Task}|\text{Participant})$$

2.5.3. *Post-experiment questionnaire*

Participants filled in a post-task questionnaire to gauge their cognitive and motor strategies. This 16-item questionnaire included 7-choice Likert scale questions about their ability to: (1) correctly hear the metronome; (2) remember the rule in step III; (3) understand the rule in steps I and II; (4) press the correct keys; (5) using the correct fingers; (6) following the correct rhythm; (7) rely on step I as a reference; (8) relying on step II as a reference; (9) imagining the fingers touching the keyboard before step III; (10) imagining colored dots moving in the keyboard before step III; (11) imagining their fingers moving on the keyboard immediately before step III; (12) preparing motorically for the sequence immediately before step III; (13) preparing visually for the sequence immediately before step III; (14) preparing spatially for the sequence immediately before step III; (15) using an intuitive strategy; and (16) thinking explicitly about the rule. At the end of each day (1 and 2), participants filled out the same questionnaire for each task (recursion and iteration).

Raw data is shown in Fig. S1. We first performed an exploratory factor analysis using Oblimin rotation and Principal Axis Extraction to test for participants' general cognitive and motor strategies. The number of factors was determined by parallel analysis, which was performed in Jamovi (The Jamovi Project, 2021). The questions clustered along three main factors (Barlett's test of Sphericity: $c2\ (105) = 990$, $p < .001$, *Overall KMO* $= .76$, see Table S1).

The first factor included Questions 2–6 and 8, which related to *using task rules in sequence generation*. Factor 2 included Questions 9–12 and 14, which related to *sensorimotor and visuospatial imagery*. Factor 3 included items related to the use of *explicit versus intuitive strategies*, such as positive loadings of Questions 7, 13, and 16 and negative loadings of Question 15.

We then tested (1) the correlation between keypress accuracy and factor scores and (2) whether the strategies varied across tasks, days, and task orders:

$$\text{Factor score} = \text{Day} \times \text{Task} \times \text{Order} + (1|\text{Participant})$$

3. Results

3.1. *Training phase*

Training results are summarized in Table 1 and Fig. 3.

The best model fit for accuracy during training (Table S2) revealed significant interactions of Task×Day×Order ($OR = 2.55$, 95% CI [1.76, 3.69], $p < .001$), Task×Day×Trial ($OR = 0.93$, 95% CI [0.89, 0.96], $p < .001$), and Task×Order×Trial ($OR = 1.16$, 95% CI [1.12, 1.20], $p < .001$). Full pairwise contrasts are depicted in Tables S3 and S4.

Table 1
Summary statistics

		Day 1		Day 2	
		Iteration	Recursion	Iteration	Recursion
Training	Order RI	90±30	76±43	94±24	90±29
% Correct *(M±SD)*	Order IR	79±40	78±42	90±31	86±34
Discrimination	Order RI	76±43	71±45	85±36	86±35
% Correct *(M±SD)*	Order IR	68±47	83±37	84±36	87±33
Generation	Order RI	63±48	26±44	72±45	48±50
% Correct (*M±SD)*	Order IR	58±49	53±50	69±46	70±46
Generation	Order RI	5/20	2/20	8/20	6/20
(Ps above chance)	Order IR	7/20	7/20	10/20	11/20

Note. Means (M) and Standard Deviations (SD) of the % of keys pressed correctly per trial during Training and Generation and the % of correct Discrimination trials. The bottom panel depicts the number of participants (Ps) with above-chance performance in the generation phase, defined as >50% trials with all keypresses correct (9/9).

Regarding the effects of Task×Day×Order (Table S3), we found that participants generally improved their performance from day 1 to day 2 regardless of task or order (I-R iteration: $z = -7.122$, $p < .0001$; I-R recursion: $z = -5.226$, $p < .0001$; R-I iteration: $z = -4.276$, $p < .0001$; R-I recursion: $z = -8.200$, $p < .0001$). Moreover, accuracy was higher in iteration versus recursion on both days for the R-I order (day 1: $z = 7.428$, $p < .0001$; day 2: $z = 3.834$, $p = .0001$), but only on day 2 for the order I-R ($z = 2.608$, $p = .0091$). Task order R-I versus I-R increased accuracy for iteration on day 1 ($z = -3.639$, $p = .0003$) but did neither impact day 2 nor recursion (ps > .05).

Regarding the effect of Task×Day×Trial (Table S4), we found that learning slopes were *steeper* in recursion versus iteration on day 1 ($z = -4.0$, $p < .0001$) but not on day 2 ($z = 1.7$, $p = .1$). This effect was caused by an initial lower accuracy for recursion on day 1. Following the effect of Task×Order×Trial, we found that learning slopes were *steeper* for the task with which participants initiated the procedure: iteration versus recursion in the I-R order ($z = 4.8$, $p < .0001$) and in recursion versus iteration in the R-I order ($z = -5.7$, $p < .0001$).

3.2. Discrimination phase

Discrimination results are summarized in Table 1 and Fig. 4.

The best model fit for discrimination accuracy (Table S5) revealed significant interactions of Task×Day×Order (*OR* = 4.3, 95% CI [1.97, 9.29], $p < .001$), Task×Order×Foil odd (*OR* = 0.16, 95% CI [0.06, 0.45], $p < .001$), and Day×Order×Foil positional *(OR* = 6.58, 95% CI [2.58, 16.80], $p < .001$). Full pairwise contrasts are depicted in Table S6.

Regarding the effects of Task×Day×Order, we found that participants were better in discriminating correct sequences (vs. foils) on day 2 (vs. day 1), regardless of task or order (I-R iteration: $z = -6.34$, $p < .0001$; I-R recursion: $z = -2.88$, $p = .0003$; R-I iteration: $z = -3.69$, $p = .0002$; R-I recursion: $z = -4.83$, $p < .0001$). Participants were also more accurate on recursion versus iteration (in both days) when the order was I-R (day 1: $z = -5.88$, $p < .0001$; day 2: $z = -2.36$, $p = .02$) but not when the order was R-I (*ps* > .4). Finally,

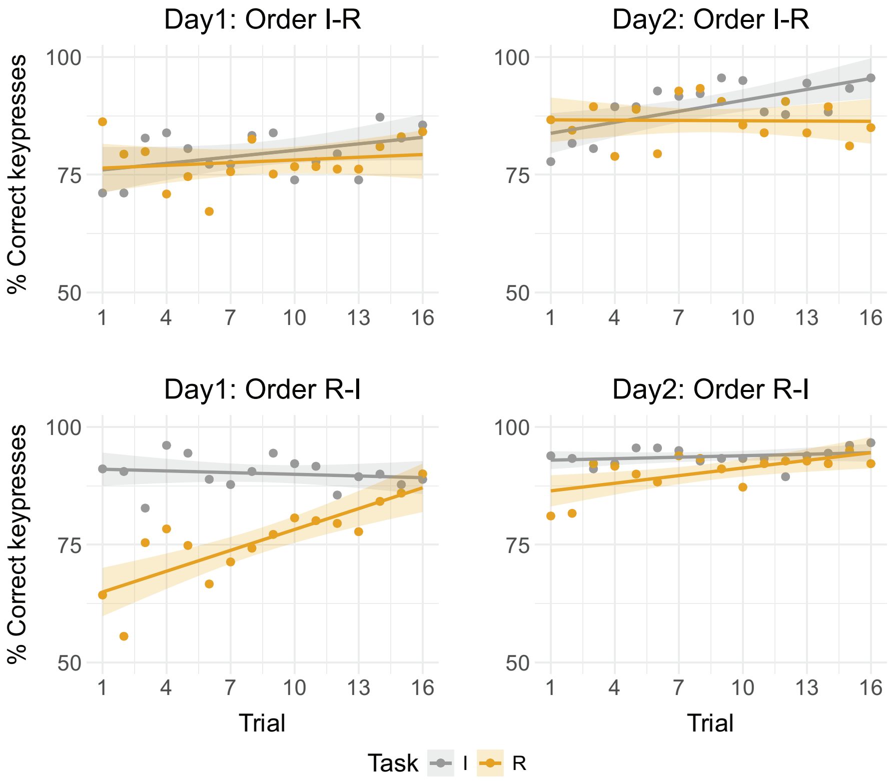

Fig. 3. Training phase: mean keypress accuracy across trials. The goal of the training phase was to expose participants (all nonmusicians) to the process of generating motor fractals and to measure their ability to execute the sequences of nine keypresses motorically. Participants were asked to execute sequences in three consecutive steps based on visual cues and to try to extract the underlying generative rule. Participants performed iterative (I) and recursive (R) tasks on 2 days. Half of the participants ($n = 20$) started the procedure with iteration (order I-R; upper row), and half ($n = 20$) with recursion (order R-I; lower row). The training tasks were composed of 16 trials. Shaded areas around regression lines denote confidence intervals; dots represent participants' averaged keypress accuracy in step III. Globally, keypress accuracy was high, indicating adequate motor performance, especially on the second day. Accuracy was particularly low, and learning rates were steeper for recursion versus iteration on the first day, especially when the procedure started with recursion.

participants were more accurate in a certain task within day 1, when they performed that task in the second position of the procedure (order R-I vs. I-R for iteration $z = -2.12$, $p = .03$; order I-R vs. R-I for recursion: $z = 2.67$, $p = .008$). These differences were not significant on day 2 ($ps > .1$).

Notably, while participants were only trained with sequences containing adjacent keypresses (i.e., with $s = \{-1, 1\}$ in the transformation rule $K_n \rightarrow [K_{n+1} - s, K_{n+1}, K_{n+1} + s]$), they were equally able to discriminate well-formed sequences with nonadjacent keypresses ($s = \{-2, 2\}$) from day 1 (Fig. S3).

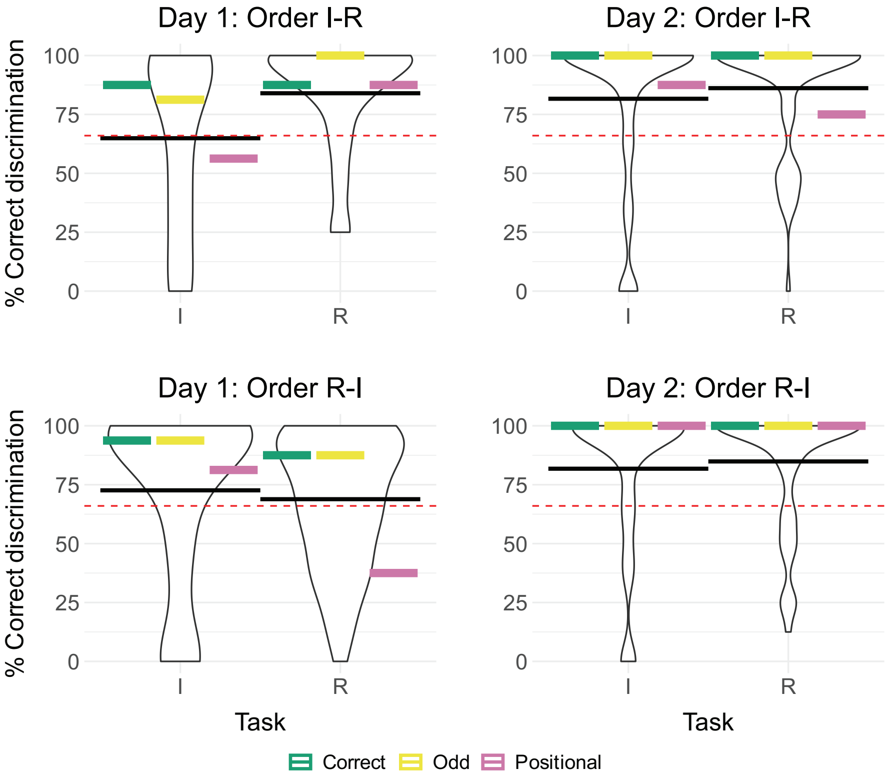

Fig. 4. Discrimination phase: discrimination accuracy of correct sequences and foils (step III). The discrimination phase aimed to evaluate whether participants could correctly extract the underlying generative rule and use it to discriminate between well-formed sequences and violations. Participants performed 32 discrimination trials. Sixteen consisted of correct sequences (green) and 16 incorrect sequences, which could consist of odd foils (yellow, $n = 8$) or positional foils (pink, $n = 8$). The red dashed horizontal line depicts the global accuracy above which performance accuracy was statistically above chance (22 out of 32 trials, *Binomial test*: $z = 1.94$, $p = .03$). Black horizontal bars depict mean accuracy. Colored horizontal bars depict the median for each trial type (correct, odd, positional). Generally: (1) participants could discriminate between correct and incorrect trials across all conditions and days; (2) this capacity improved on the second day; (3) some participants struggled in rejecting positional Foils on the first day, especially in the task they started with, indicating that the expectation of hierarchical self-similarity might not yet been formed; and (4) most participants could correctly classify all trial types on the second day.

Regarding the Foil effects, we found that some participants struggled with rejecting positional foils, especially on day 1 and in the task with which they started the procedure (Fig. 4, left). Importantly, most participants performed above chance in all foil categories on day 2 (Fig. 4, right).

3.3. Generation phase

Generation results are summarized in Table 1 and Fig. 5. Individual-level dynamics across days are shown in Fig. S4.

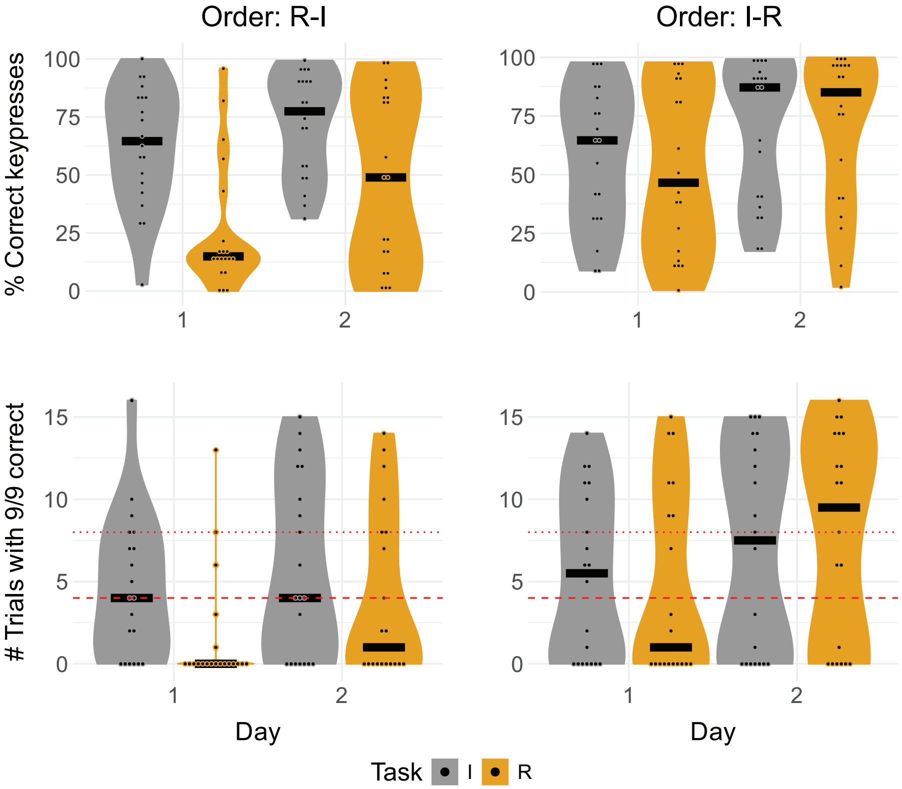

Fig. 5. Generation phase: keypress accuracy. The goal of the generation phase was to evaluate whether participants could use the rules acquired during the training and discrimination phases to extend motor sequences beyond the given hierarchical level. After being shown steps I and II, they were asked to generate step III following a metronome without visual support. Top row: mean % correct keypresses per participant. Each dot represents individual means, and black horizontal bars depict median accuracy. Bottom row: Number of trials (out of 16) in which participants pressed all nine keys correctly. The probability of using the correct parameter combination given in step II (out of 4: ascending vs. descending; adjacent vs. nonadjacent keys) is 25%, corresponding to 4 out of 16 trials (red dashed line); a proportion of 50% correct trials is significantly above chance (red dotted line). Broadly, participants improved significantly from day 1 to day 2, and this effect was more pronounced for recursion. Performance in recursion significantly improved by starting the procedure with iteration. Most participants of the order I-R could use recursive rules productively to generate new hierarchical levels, although many struggled.

The best model fit for accuracy during generation (Table S7 revealed significant interactions between Task×Day ($OR = 2.16$, 95% CI [1.86, 2.51], $p < .001$) and Task×Order ($OR = 0.13$, 95% CI [0.06, 0.32], $p < .001$). Full pairwise contrasts are depicted in Table S8.

Regarding the effects of Task×Day, we found that participants were more accurate in iteration versus recursion on both days (Day 1: $z = 6.10$, $p < .0001$; Day 2: $z = 2.68$, $p = .0074$) and more accurate on day 2 versus day 1 on both tasks (iteration: $z = -3.29$, $p = .001$; recursion: $z = -6.81$, $p < .0001$).

Regarding the effects of Task×Order, we found that participants were less accurate in recursion (vs. iteration) when they started the procedure with recursion (order R-I: $z = 6.35$, $p < .0001$). However, when the task order was reversed, there was no difference between tasks (order I-R: $z = -0.06$, $p = .9$). Similarly, participants were more accurate in the order I-R versus R-I for recursion ($z = 2.89$, $p = .004$) but not iteration ($z = -0.40$, $p = .68$).

Notably, while participants were only trained with sequences containing adjacent keypresses (i.e., with $s = \{-1, 1\}$ in the transformation rule $K_n \rightarrow [K_{n+1} - s, K_{n+1}, K_{n+1} + s]$), they were equally able to generate well-formed sequences with nonadjacent keypresses ($s = \{-2, 2\}$) from day 1 (Fig. S5).

Interestingly, keypress accuracy in the generation phase was correlated only moderately with discrimination ($r(38) = .32$, $p = .05$) and keypress accuracy in the training phase ($r(38) = .38$, $p = .01$) (all on the second day).

We also measured the proportion of trials for which participants correctly pressed 9/9 keys (Fig. 5, bottom row). Considering four parameter combinations $\{-2, -1, 1, 2\}$ (ascending vs. descending; adjacent vs. nonadjacent keys), the probability that participants applied the correct one was 25%. However, this estimate is extremely conservative as the probability of executing a correct sequence of nine keypresses given a 16 key-keyboard is exceedingly low. With 16 trials, participants who correctly pressed 9/9 keys in eight trials performed statistically above chance (*Binomial test with a probability of 25%*: $z = 2.02$, $p = .03$). The proportion of participants who had more than eight correct trials on the second day was 10/20 for iteration and 11/20 for recursion in the order I-R, and 8/20 for iteration and 6/20 for recursion in the order R-I.

3.4. Beat-keypress latency

In addition to general keypress accuracy, we measured beat-keypress latencies in the generation phase and analyzed the patterns across the nine keypresses (Fig. 6). Unsurprisingly, latency was highest in the first key of the sequence. Following Moss et al. (2023), we computed a chunk transition measure corresponding to the latency difference between each cluster's "1st" key versus the other keys "2nd | 3rd." We found an interaction between Chunk Transition × Task ($OR = 0.97$, 95% CI [0.96, 0.98], $p < .001$). On average, the first key of each cluster had the *highest* latency for iteration ($z = -5.27$, $p < .0001$) but not for recursion ($z = 1.51$, $p = .13$). And the latencies for the "2nd | 3rd" keys were *higher* in recursion versus iteration ($z = 6.04$, $p < .0001$), while the task differences were not significant for the 1st key ($z = 1.56$, $p = .12$).

3.5. Post-experiment questionnaire analysis

Our factor analysis suggests that the post-experiment questions cluster along three factors (Table S1). Factor 1 included items related to *task rules in the sequence generation*. Factor 2 included items related to *sensorimotor and visuospatial imagery*. Factor 3 included items related to the use of *explicit versus intuitive strategies*.

In general, we found that keypress accuracy in the generation phase correlated with the scores in Factor 1 ($r = .349$, $p < .001$) and Factor 2 ($r = .211$, $p = .008$) but not Factor 3

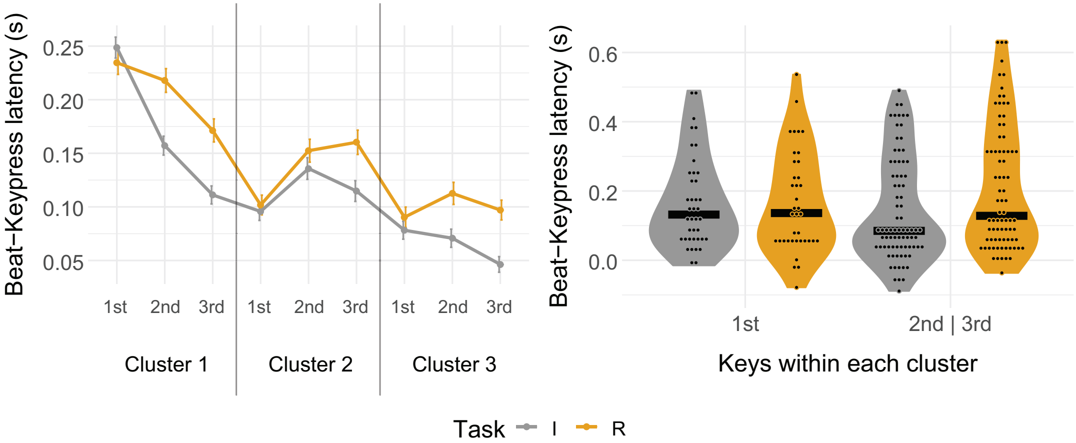

Fig. 6. Generation phase: beat-keypress latency. Keypresses were valid if occurring within a window of [−.25; 1] seconds relative to the metronome beat. Negative latency values represent keypresses preceding the metronome. Here, we depict only the latency data from correct keypresses in step III (including both days and task orders). Left: Latencies across the nine keypresses (three keypresses per cluster—1st, 2nd, 3rd). Error bars represent the standard error. Right: 1st and 2nd | 3rd keypress latencies averaged across clusters (Chunk transition analysis). Previous research suggests that latency is higher in the first item of a cluster relative to the others due to the initiation of a motor schema (Moss et al., 2023). Here, this pattern is found (on average) for iteration but not for recursion. Latencies are higher for recursion than iteration for “2nd | 3rd” keys, which indicates greater cognitive effort in completing each cluster. Each data point is the average keypress latency per participant for the 1st, 2nd, and 3rd keys (40 data points per keypress). Black horizontal bars depict median values.

($r = .108$, $p = .180$), suggesting that adequate performance relied on both the understanding and use of task rules and sensorimotor and visuospatial imagery. We also found that Factor 1 scores were higher on day 2 versus day 1 ($b = 0.8$, 95% CI [0.62, 1.02], $p < .001$), and for iteration versus recursion when the task order was R-I ($b = -0.53$, [−0.92, −0.14], $p = .009$). This suggests task rules were less clear on day 1 and for recursion in the order R-I. We did not find significant task, day, and order effects for Factors 2 and 3 (Tables S9–S11).

We also explored the data on individual questions (Figs. S1 and S2). Participants reported using explicit rule representation strategies ($M = 5.16$ on a 1–7 Likert scale) more often than intuitive strategies ($M = 3.7$). However, neither strategy significantly correlated with recursive generation accuracy (explicit: $r = .17$, $p = .13$; intuitive: $r = -.04$, $p = .87$). Importantly, recursion generation accuracy was significantly correlated with motor ($r = .22$, $p = .05$) and spatial ($r = .34$, $p = .002$) imagery before step III, while visual preparation ($r = .00$, $p = .98$) and imagining the sequence of colored dots on the keyboard ($r = .09$, $p = .44$) were not. This suggests that visual imagery played a less significant role than motor-spatial imagery.

4. Discussion

We have shown for the first time that humans can (1) acquire recursive rules in the motor domain from practicing examples without trial feedback or prior rule explanation; (2) use these rules to discriminate between well-formed hierarchical sequences versus violations; and (3) generate new hierarchical levels consistent with the previously given. Furthermore, with our novel design and procedure, we revealed interesting learning patterns as follows: (1) participants were more accurate on the second day of testing in all phases; (2) the ability to use recursion was greatly enhanced by the prior acquisition of iteration, mimicking findings from other domains; and (3) a strong ability to discriminate well-formed recursive structures was not sufficient for above-chance novel sequence generativity. Finally, using recursive rules was associated with higher effort in cluster completion relative to iteration. We now discuss these findings in detail.

First, more than one test session was needed for participants to generate new hierarchical levels following the recursive rule, while participants successfully utilized the iterative rule already on day 1. Broadly, participants performed better in both rules on the second day in all phases, that is, training, discrimination, and generation. These results align with previous work suggesting that abstract representations of motor sequences are facilitated by sleep consolidation (Fischer et al., 2005; Fogel et al., 2017). However, our experiment was not designed to test for the causal role of sleep. As we did not explicitly test for sleep as a factor in acquiring hierarchical levels in motor sequences, the increase in performance may also be due to the effect of a prior session. Post-experiment questionnaires showed, for example, that the understanding and utilization of task rules increased on the second day and correlated with keypress accuracy in the generation phase. In particular, on the second day, participants reported (1) that it was easier to extract the trial-relevant information from step II to apply the rule in step III, and (2) that they were more confident to have used the correct fingers, pressed the correct keys, and followed the correct rhythm. Notably, most participants seemed to think about the rules explicitly rather than use intuitive strategies.

Second, we found a significant effect of the task order. Prior experience with iteration benefits the acquisition of recursion. Interestingly, this effect was already visible during training while identical motor sequences were copied. However, it was particularly pronounced in the generation phase and crucially asymmetric between recursion and iteration. For instance, on day 2, task order significantly impacted the accuracy of recursion (I-R: 70% vs. R-I: 48%) but not iteration (I-R: 69% vs. R-I: 72%). Out of 20 participants, 11 could use recursion to generate the correct sequences in most trials on the second day when starting the procedure with iteration and then recursion. In the converse order, only 6/20 did so. The effects were even more significant on the first day, in which 7/20 participants could perform when the task order was iteration-recursion, while only 2/20 did so in the converse order. This (asymmetrical) enhancing effect of iteration on recursion replicates our findings in the visual domain (Martins et al., 2014). It is also consistent with the literature on language acquisition, showing that learning "same-level" conjunctive constructions (e.g., the [second] and [red] ball) paves the way to the acquisition of recursive rules adding information across "multiple-levels" (e.g., the [[second] red] ball) (Matthei, 1982; Roeper, 2011; Yang et al., 2022). The overall proportion of participants with a majority of correct trials in our learning task without explicit rule instructions or feedback was similar to previous findings with the same tasks after 1 h of training with explicit rule explanation and verbal feedback (Martins et al., 2019).

Third, we found evidence that high-performing motor execution in training and above-chance discrimination of well-formed structures were not sufficient to ensure the capacity to generate new hierarchical levels beyond the given. In particular, while the ability to discriminate well-formed recursive structures in the task order R-I was above chance on days 1 and 2 (mean accuracies 71% and 85%, respectively), generation accuracy was poor (26% and 48%). Consequently, the correlation between discrimination and generation performance was only weak ($r(38) = .32$, $p = .05$). These findings underscore recent evidence (Dedhe et al., 2023) that discrimination tasks are insufficient to determine whether participants can represent recursion. For instance, associative learning might suffice to parse long-distance dependencies in AGL paradigms (Lakretz & Dehaene, 2021; Rey, Perruchet, & Fagot, 2012; Rey & Fagot, 2023). Moreover, discrimination could have been solved by a heuristic representation that did not involve the recursive rule. For example, rejecting the positional foil (intended to test recursive rule understanding) required the understanding of coherence between hierarchical levels, or in other words, the coherence between and within keypress clusters. This coherence is a prerequisite for recursive hierarchical embedding. However, without analyzing participants' error behaviors across a wide variety of foils, it is difficult to exclude the possibility that they may have used different heuristics. Conversely, the generation task cannot be solved without extracting the trial-specific free parameters k (initial key) and s (transformation) to assess where the sequence starts, its directionality, and whether it generates sequences of adjacent or nonadjacent keys (the latter of which were not included in the training). In addition to these abstract, free parameters, participants must be able to execute the sequence within the correct rhythmic structure. In other words, generation reduces participants' likelihood of relying on simple heuristics to solve the task. Importantly, the ability to motorically execute the sequences during training does not guarantee adequate performance during generation, as the latter requires the retrieval of the appropriate sequence. While training performance was relatively high for most participants, many struggled in the generation phase. Furthermore, the correlation between training and generation was .38.

Finally, beat-keypress latency patterns revealed higher cognitive effort in generating motor sequences when applying recursive rules. Interestingly, latencies were identical for iteration and recursion in the first key of each cluster, while they were higher for recursion in the second and third keys. Following Moss et al. (2023), this suggests that cluster completion (but not initiation) was more challenging with recursion, perhaps because the latter required adding two keypresses per cluster instead of one. The requirement of adding more elements to the sequence is likely to increase the working memory load (Kessler & Oberauer, 2014). These results are also consistent with previous neuroimaging findings that recursion is more demanding to the motor control system than iteration (Martins et al., 2019).

4.1. Theoretical and empirical contributions

Our work is the first demonstration that human adults can acquire and use recursive rules to generate fractal structures in the motor domain. We have also demonstrated similar recursive abilities in the visual and musical domains (Martins et al., 2016, 2017), including in children and stroke patients (Martins et al., 2014; Martins et al., 2019). Overall, our results support the claim that recursion is available in several cognitive domains beyond language, thus falsifying the hypothesis that recursion is language-domain-specific (Hauser et al., 2002).

This is also the first demonstration, beyond language, that humans can acquire recursion in the sensorimotor domain from practicing examples without prior rule explanation or trial feedback. In our work with vision and music, we tested for the ability to discriminate between correct and incorrect stimuli perceptually. As shown here and discussed above (also here, Dedhe et al., 2023), discrimination per se might be insufficient evidence for a robust generative capacity. Previous research has shown that humans and nonhuman animals (birds and primates) can touch images in the correct sequence on a touchscreen following an underlying hierarchical structure (Ferrigno et al., 2020; Liao et al., 2022). However, in these experiments, the underlying hierarchy was cued visually (using brackets of different shapes and colors), and feedback was provided. In our learning setup, participants learned and generalized by copying examples without feedback, which is more naturalistic and prevents associative learning (Lakretz & Dehaene, 2021; Rey & Fagot, 2023). Furthermore, participants were only trained with well-formed examples of motor sequences that included adjacent keypresses ($s = \{-1, 1\}$). However, already on the first day, they were equally accurate in discriminating and generating sequences with nonadjacent keypresses ($s = \{-2, 2\}$), demonstrating that they could generalize beyond the training set.

Importantly, neuroimaging research suggests that hierarchical cognition in language is supported by neural systems different from those of the motor domain (Friederici, 2023; Zaccarella, Papitto, & Friederici, 2021). Thus, motor recursion is likely independent of language, in line with the idea that distinct biological systems can implement similar computations (Asano et al., 2021; Dehaene et al., 2022; Fedorenko & Shain, 2021; Martins et al., 2014; Martins, Muršič, Oh, & Fitch, 2015; Martins et al., 2019; Martins et al., 2020).

It is important to note that the mechanisms involved in acquiring recursive rules are poorly understood, and it is unclear whether these require domain-general or domain-specific cognitive and neural systems. This topic is discussed extensively elsewhere (Martins, 2017, 2024; Martins et al., 2015; Martins & Fitch, 2014). One hypothesis is that while domain-specialized cognitive and neural systems are involved in applying recursive rules when these are *well-trained* (Martins et al., 2019; Martins et al., 2020; Martins et al., 2014), the *acquisition* of recursive rules might show cognitive and neural similarities between music, action, vision, and language (Martins et al., 2017; Martins et al., 2019). Under this framework, participants may use domain-general cognition to extract a recursive rule that they then instantiate motorically. Importantly, this general cognition does not seem to activate the classic fronto-parietal brain networks associated with cognitive effort and intelligence (Fischmeister et al., 2017; Martins et al., 2019).

In the current task, participants' performance correlated with increased confidence in pressing the correct keys, using the correct fingers, and following the correct rhythm. Moreover, most participants reported preparing motorically, imagining finger movements, and touching the keys. We also know that when participants train extensively with the task, recursion in the motor domain activates classic motor imagery areas, including basal ganglia, cerebellum, and sensorimotor and premotor regions (Martins et al., 2019). However, with the current behavioral setup, it is impossible to determine whether motor domain-specific systems are involved in acquiring the rule or whether the motor domain interfaces with a domain-general recursive system. Future neuroimaging studies focusing on acquiring motor recursion will be essential to disambiguate these hypotheses.

Finally, our results offer design suggestions for future experiments. First, we found a gap between discrimination and generation performance, supporting the idea that generation is a more robust test of the underlying representations (Dedhe et al., 2023). Second, more than 1 day may be necessary to consolidate recursive rules from sensorimotor sequences (Fischer et al., 2005; Fogel et al., 2017), especially if feedback is not provided. Third, since the knowledge of recursion scaffolds on iteration, fair tests of recursion in clinical and nonhuman samples should start the procedure with iteration, as a failure to pass the task might not represent an inability to develop the knowledge of recursion, but rather the lack of relevant prior knowledge with (simpler) iterative rules (Martins et al., 2014; Roeper, 2011; Yang et al., 2022).

Funding

This work was supported by the Max Planck Society. R.B. is funded by the European Union (MSCA, PHYLOMUSIC, 101064334).

Conflicts of interests

The authors have no competing interests to declare relevant to this article’s content.

Ethics approval statement

This work was approved under the ethics committee project number 016-15-26012015 at the Max Planck Institute for Human Cognitive and Brain Sciences.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent for publication

The authors affirm that human research participants provided informed consent for publication.

Availability of data and materials

The datasets generated during and/or analyzed during the current study are available in the OSF repository, https://osf.io/dbvrn

Code availability statement

The code used for statistical analysis in the current study is available in the OSF repository, https://osf.io/dbvrn

References

Asano, R., Boeckx, C., & Seifert, U. (2021). Hierarchical control as a shared neurocognitive mechanism for language and music. *Cognition*, *216*, 104847. https://doi.org/10.1016/j.cognition.2021.104847

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

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience*, *10*(9), 659–669. https://doi.org/10.1038/nrn2667

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Berwick, R. C., & Chomsky, N. (2016). *Why only us: Language and evolution*. MIT Press.

Berwick, R. C., Friederici, A. D., Chomsky, N., & Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends in Cognitive Sciences*, *17*(2), 89–98. https://doi.org/10.1016/j.tics.2012.12.002

Cheung, V. K. M., Meyer, L., Friederici, A. D., & Koelsch, S. (2018). The right inferior frontal gyrus processes nested non-local dependencies in music. *Scientific Reports*, *8*(1), 3822. https://doi.org/10.1038/s41598-018-22144-9

Chomsky, N. (1957). *Syntactic structures*. De Gruyter Mouton. https://doi.org/10.1515/9783112316009

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, *23*(2), 157–205. https://doi.org/10.1207/s15516709cog2302_2

De Villiers, J., & De Villiers, P. A. (2014). The role of language in theory of mind development. *Topics in Language Disorders*, *34*(4), 313–328. https://doi.org/10.1097/TLD.0000000000000037

De Villiers, J., Hobbs, K., & Hollebrandse, B. (2014). Recursive complements and propositional attitudes. In T. Roeper & M. Speas (Eds.), *Recursion: Complexity in cognition* (pp. 221–242, Vol. 43). Springer International Publishing. https://doi.org/10.1007/978-3-319-05086-7_10

Dedhe, A. M., Clatterbuck, H., Piantadosi, S. T., & Cantlon, J. F. (2023). Origins of hierarchical logical reasoning. *Cognitive Science*, *47*(2), 13250. https://doi.org/10.1111/cogs.13250

Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S., & Sablé-Meyer, M. (2022). Symbols and mental programs: A hypothesis about human singularity. *Trends in Cognitive Sciences*, *26*(9), 751–766. https://doi.org/10.1016/j.tics.2022.06.010

Everett, D. L. (2012). What does Pirahã grammar have to teach us about human language and the mind? *WIREs Cognitive Science*, *3*(6), 555–563. https://doi.org/10.1002/wcs.1195

Fazio, P., Cantagallo, A., Craighero, L., D'Ausilio, A., Roy, A. C., Pozzo, T., Calzolari, F., Granieri, E., & Fadiga, L. (2009). Encoding of human action in Broca's area. *Brain*, *132*(7), 1980–1988. https://doi.org/10.1093/brain/awp118

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

Ferguson, N. (2018). *The square and the tower: Networks and power, from the Freemasons to Facebook*. Penguin Books.

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), eaaz1002. https://doi.org/10.1126/sciadv.aaz1002

Fischer, S., Nitschke, M. F., Melchert, U. H., Erdmann, C., & Born, J. (2005). Motor memory consolidation in sleep shapes more effective neuronal representations. *Journal of Neuroscience*, *25*(49), 11248–11255. https://doi.org/10.1523/JNEUROSCI.1743-05.2005

Fischmeister, F. P., Martins, M. D., Beisteiner, R., & Fitch, W. T. (2017). Self-similarity and recursion as default modes in human cognition. *Cortex*, *97*, 183–201. https://doi.org/10.1016/j.cortex.2016.08.016

Fitch, W. T. (2010). Three meanings of "recursion": Key distinctions for biolinguistics. In R. K. Larson, V. Déprez, & H. Yamakido (Eds.), *The evolution of human language* (1st ed., pp. 73–90). Cambridge University Press. https://doi.org/10.1017/CBO9780511817755.005

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *303*(5656), 377–380. https://doi.org/10.1126/science.1089401

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

Fogel, S., Albouy, G., King, B. R., Lungu, O., Vien, C., Bore, A., Pinsard, B., Benali, H., Carrier, J., & Doyon, J. (2017). Reactivation or transformation? Motor memory consolidation associated with cerebral activation time-locked to sleep spindles. *PLoS One*, *12*(4), e0174755. https://doi.org/10.1371/journal.pone.0174755

Friederici, A. D. (2023). Evolutionary neuroanatomical expansion of Broca's region serving a human-specific function. *Trends in Neurosciences*, *46*(10), 786–796. https://doi.org/10.1016/j.tins.2023.07.004

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

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, *440*(7088), 7088. https://doi.org/10.1038/nature04675

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). *The faculty of language: What is it, who has it, and how did it evolve?*

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Perception*, *26*(3), 195–204. https://doi.org/10.1525/mp.2009.26.3.195

Jackendoff, R. (2011). What is the human language faculty?: Two views. *Language*, *87*(3), 586–624. https://doi.org/10.1353/lan.2011.0063

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

Kessler, Y., & Oberauer, K. (2014). Working memory updating latency reflects the cost of switching between maintenance and updating modes of operation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 738–754. https://doi.org/10.1037/a0035545

Koechlin, E., & Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron*, *50*(6), 963–974. https://doi.org/10.1016/j.neuron.2006.05.017

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

Lakretz, Y., & Dehaene, S. (2021). *Recursive processing of nested structures in monkeys? Two alternative accounts* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/k8vws

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior; The Hixon Symposium* (pp. 112–146). Wiley.

Lerdahl, F., & Jackendoff, R. (1983). An overview of hierarchical structure in music. *Music Perception*, *1*(2), 229–252. https://doi.org/10.2307/40285257

Levinson, S. C. (2013). Recursion in pragmatics. *Language*, *89*(1), 149–162.

Liao, D. A., Brecht, K. F., Johnston, M., & Nieder, A. (2022). Recursive sequence generation in crows. *Science Advances*, *8*(44), eabq3356. https://doi.org/10.1126/sciadv.abq3356

Lobina, D. J. (2011). Recursion and the competence/performance distinction in AGL tasks. *Language and Cognitive Processes*, *26*(10), 1563–1586. https://doi.org/10.1080/01690965.2011.560006

Mandelbrot, B. B. (1977). *Fractals: Form, chance, and dimension*. Freeman.

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

Martins, M. D. (2017). The cognitive architecture of recursion: Behavioral and fMRI evidence from the visual, musical and motor domains. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*.

Martins, M. D. (2024). Cognitive and neural representations of fractals in vision, music, and action. In A. Di Ieva (Ed.), *The fractal geometry of the brain* (pp. 935–951). Springer International Publishing. https://doi.org/10.1007/978-3-031-47606-8_46

Martins, M. D., 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), 2623–2638. https://doi.org/10.1002/hbm.24549

Martins, M. D., Fischmeister, F. P., Puig-Waldmüller, E., Oh, J., Geißler, 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. D., 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 225*, 1997–2015. https://doi.org/10.1007/s00429-020-02105-7

Martins, M. D., & Fitch, W. T. (2014). Investigating recursion within a domain-general framework. In F. Lowenthal & L. Lefebre (Eds.), *Language and recursion* (pp. 15–26). Springer. Retrieved from http://link.springer.com.ezp-prod1.hul.harvard.edu/chapter/10.1007/978-1-4614-9414-0%7B_%7D2

Martins, M. D., Gingras, B., Puig-Waldmueller, E., & Fitch, W. T. (2017). Cognitive representation of "musical fractals": Processing hierarchy and recursion in the auditory domain. *Cognition 161*, 31–45. https://doi.org/10.1016/j.cognition.2017.01.001

Martins, M. D., Krause, C., Neville, D. A., Pino, D., Villringer, A., & Obrig, H. (2019). Recursive hierarchical embedding in vision is impaired by posterior middle temporal gyrus lesions. *Brain*, *142*(10), 3217–3229. https://doi.org/10.1093/brain/awz242

Martins, M. D., Laaha, S., Freiberger, 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. D., Martins, I. P., & Fitch, W. T. (2016). A novel approach to investigate recursion and iteration in visual hierarchical processing. *Behavior Research Methods*, *48*(4), 1421–1442. https://doi.org/10.3758/s13428-015-0657-1

Martins, M. D., Muršič, Z., Oh, J., & Fitch, W. T. (2015). Representing visual recursion does not require verbal or motor resources. *Cognitive Psychology*, *77*, 20–41. https://doi.org/10.1016/j.cogpsych.2015.01.004

Matthei, E. H. (1982). The acquisition of prenominal modifier sequences. *Cognition*, *11*(3), 301–332. https://doi.org/10.1016/0010-0277(82)90018-X

Moss, M. E., Zhang, M., & Mayr, U. (2023). The effect of abstract inter-chunk relationships on serial-order control. *Cognition*, *239*, 105578. https://doi.org/10.1016/j.cognition.2023.105578

Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351–360. https://doi.org/10.1038/nrn2811

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing [Computer software]. Retrieved from https://www.R-project.org/

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

Redhead, D., & Power, E. A. (2022). Social hierarchies and social networks in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*, 20200440. https://doi.org/10.1098/rstb.2020.0440

Rey, A., & Fagot, J. (2023). Associative learning accounts for recursive-structure generation in crows. *Learning & Behavior*, *51*, 347–348. https://doi.org/10.3758/s13420-022-00564-y

Rey, A., Perruchet, P., & Fagot, J. (2012). Centre-embedded structures are a by-product of associative learning and working memory constraints: Evidence from baboons (*Papio Papio*). *Cognition*, *123*(1), 180–184. https://doi.org/10.1016/j.cognition.2011.12.005

Roeper, T. (2011). The acquisition of recursion: How formalism articulates the child's path. *Biolinguistics*, *5*(1–2), 057–086. https://doi.org/10.5964/bioling.8831

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, *5*(1), 35–53. https://doi.org/10.1080/17459737.2011.573676

Russell, L. (2018). *emmeans: Estimated marginal means, aka least-squares means*. Retrieved from https://cran.r-project.org/package=emmeans

Scholz, R., Villringer, A., & Martins, M. D. (2023). Distinct hippocampal and cortical contributions in the representation of hierarchies. *eLife 12*:RP87075. https://doi.org/10.7554/eLife.87075

Seyfarth, R. M., & Cheney, D. L. (2017). Precursors to language: Social cognition and pragmatic inference in primates. *Psychonomic Bulletin & Review*, *24*(1), 79–84. https://doi.org/10.3758/s13423-016-1059-9

The jamovi project (Version 1.6). (2021). [Computer software]. Retrieved from https://www.jamovi.org

Uddén, J., Martins, M. D., Zuidema, W., & Tecumseh Fitch, W. (2020). Hierarchical structure in sequence processing: How to measure it and determine its neural implementation. *Topics in Cognitive Science*, *12*(3), 910–924. https://doi.org/10.1111/tops.12442

van der Hulst, H. (2010). *Recursion and human language*. De Gruyter Mouton. https://doi.org/10.1515/9783110219258

Yang, C., Hu, Y., Fan, J., Dong, X., & Jeschull, L. (2022). An experimental study of Mandarin-speaking children's acquisition of recursion under a formal definition/classification system for recursion. *Lingua*, *271*, 103228. https://doi.org/10.1016/j.lingua.2021.103228

Zaccarella, E., Papitto, G., & Friederici, A. D. (2021). Language and action in Broca's area: Computational differentiation and cortical segregation. *Brain and Cognition*, *147*, 105651. https://doi.org/10.1016/j.bandc.2020.105651

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary Information