Signatures of a "Language of Thought" and its cross-modal integration

Master thesis

submitted in partial fulfillment of the requirements for the degree of

Master in Cognitive Science

***

M2 CogMaster

at the *École Normale Supérieure*

Written by

Maxime Cauté

École Normale Supérieure de Rennes

Under the supervision of

Prof. Stanislas Dehaene

Collège de France

INSERM–CEA

Cognitive Neuroimaging Lab

Year 2021 – 2022

Abstract

The hypothesis of a Language of Thought (LoT) at the heart of human perception has received an increasing attention over the year. Many studies proposed to analyze human performance on various tasks along the lines of the complexity of the mental program required to process their related stimuli. However, these domains have until now mostly been studied in isolation. Although some work tackled the interactions between domains such as language and mathematics, behavior and brain imaging results do not agree over the existence of shared syntactic mechanisms. In this work, we proposed to investigate a potential domain-generality of LoT over two simpler domains: binary patterns and binary sequences. As binary sequences were already shown to display the signature of a LoT-like mechanism in the literature, we replicated these findings for binary patterns in Exp. 1. In Exp. 2, we tried to find a priming effect across the two domains, without success. Given our small sample size, it is impossible to know whether this failure is due to an absence of effect or to an insufficient statistical power. Our results are nonetheless encouraging and push for further investigations.

Acknowledgments

I'd like to thank first and foremost all the people who welcomed me at NeuroSpin, and offered me great times among them. In particular, thank you to Fosca Al Roumi, Caroline Bévalot, Lucas Benjamin, Lorenzo Ciccione, Marie Lubineau and Cassandra Pottier-Watkins for their everyday kindness, as well as their precious advice. Thank you similarly to Christophe Pallier, to whom I also still need to return a very nice book on parsing trees.

Thank you as well to my classmates Samuel Debray, Michael Goodale, Maëva l'Hôtellier, and Cyann Bernard for the great discussions on this topic with them, their support, and their friendship. In particular, thank you to Samuel for the template of the present Document

Last but definitely not least, thank you to Mathias Sablé-Meyer and Stanislas Dehaene for their precious supervision and guidance.

Contents

Declaration of Originality

This thesis investigates the domain-generality of the Language of Thought (LoT). Although some prior work did tackle LoT-based interaction for complex domains such as language and mathematics, no clear consensus emerged from these works. To limit the number of possible confounds, we focused on simple domains, which were never studied beyond isolation. We proposed a habituation-based priming paradigm to investigate the syntactic interactions across domains. Although time did not allow to run a full experiment using this paradigm, we reported encouraging results. Should this work be pursued, it may tip decisively the balance about whether syntactic structures are shared across different domains of cognition, one way or another.

Declaration of Contribution

Definition of the resarch question. Maxime Cauté, Mathias Sablé-Meyer, Stanislas Dehaene

Litterature review. Maxime Cauté, with suggested articles from Mathias Sablé-Meyer and Stanislas Dehaene.

Choice of methodology. Maxime Cauté, Mathias Sablé-Meyer, Stanislas Dehaene

Programming. Experiments were programmed by Maxime Cauté upon a basis from Mathias Sablé-Meyer; complexities for Geo were computed using a tool from Sergio Romano [Ama+17] .

Design of the experiments. Maxime Cauté, Mathias Sablé-Meyer, Stanislas Dehaene.

Subjects recruitment and testing. Mathias Sablé-Meyer for uploading Exp. 1 and downloading resulting data, Stanislas Dehaene for online recruitement (Exp. 1), Maxime Cauté for the whole Exp. 2.

Data analysis. Maxime Cauté, with occasional help from Mathias Sablé-Meyer

Interpretation of the results. Maxime Cauté, Mathias Sablé-Meyer & Stanislas Dehaene.

Redaction of the thesis. Maxime Cauté, using a template from Samuel Debray.

Proof-reading of the thesis. Mathias Sablé-Meyer.

Pre-registration

0.1 Administrative information

The internship will be conducted at NeuroSpin's Unicog team (INSERM – CEA Paris-Saclay) under the supervision of Stanislas Dehaene. Mathias Sablé-Meyer (PhD students) acted as a junior referee.

0.2 Introduction

Background and rationale. Syntactic ability is foundational for human language and the complex organisation of sentences. It notably plays an essential role in allowing for a wide expressibility of human languages and their learning. Karl Lashley's [Las+51] and Noam Chomsky's [Cho57] early proposals of a mental representation of syntax involving recursively nested structures were pursued by Jerry Fodor in 1975 [Fod75], who theorized that human syntax could be described in a "Language of Thought" (LoT). LoT-theoretic approaches were consequently proposed to describe items in further cognitive domains, such as mathematics, geometrical figures [Ama+17; Sab+21], or auditory sequences [Pla+21]. The latter notably used the same language of thought, Geo [Ama+17] (or a simplification in the case of binary sequences), which closely matched the "complexity gradients" that could be correlated both to human insight on inherent complexity and performance [Sab+21], [Ama+17]. Most interestingly, this gradient was found by Sablé-Meyer et al. [Sab+21] to hold across ages and cultures.

In parallel, recent studies have found syntactic interactions across these different cognitive domains. Scheepers et al [Sch+11] observed in 2011 that the positioning of relative clauses in sentences was primed by the structure of a prior arithmetic formula: upon being presented the ambiguous, incomplete sentence "The tourist guide mentioned the bells of the church that…", participants would be more likely to use a plural verb in the relative clause (i.e. attach "that" to "the bells"; high-attachment) than a singular verb (i.e. attach "that" to "the church"; low-attachment) if presented a high-attachment-like formula like "80 – (9 + 1) x 5" prior to this. Conversely, they were more prone to complete the sentence with a low-attachment clause if primed with a low-attachment-like formula such as "80 – 9 + 1 x 5". This mathematics-linguistic syntactic priming effect has since then been replicated numerous times [HCN18; NO18; SS14; PHS18; ZML18; ZMZ21]. Further works also highlighted syntactic interactions with other domains [VH16], especially music [VH16; FP14].

Research question Is the complexity of binary sequences, as computed by the Language-of-Thought Geo (from Amalric et al. [Ama+17]), a factor in human proficiency with them in the visual modality?

Experimental hypotheses We hypothesized that binary sequences are dealt with in humans by a geometrical language of thought similar to Geo, and that more complex sequences would result in a lesser efficiency of human subjects in tasks involving them.

0.3 Methods

Participants 100 participants will be recruited by an on-line advertisement on Twitter. Since we test for an arguably universal capacity (syntax, required for language), no demographic criterion will be implemented, with the sole exception of neurological deficits or current psychiatric medication. We will also reject participants whose performance is too low: namely whose response time is above median plus 3 standard errors, or whose accuracy was not significantly better ($p < 0.05$) than chance, as measured by a binomial test against a balanced coin ($P(1) = 0.5$) on all their trials. Responses whose reaction time is above 3200ms will also be specifically excluded.

Procedure and stimuli We will use a match-to-sample paradigm, in which we will present 16-item binary sequences to participants as both sample and query. Sequences will be displayed vertically at the center of the screen; items will be black dots shifted either to the left or to the right. The sample sequence will be presented 1200ms, followed by a mask flashed during 500ms; the query sequence will be presented in the same fashion. The masks will consist in a display of all possible dot positions, that is of two vertical trails of dots shifted to the left and the right from the center.

Our original sequences will be the 10 size-16 ones used in Planton et al., and will present various complexities (as per Planton et al.'s measure; see Fig. 2.2). Deviant sequences will be generated by shifting a single dot to the other side (from left to right, or right to left), with a uniform random choice over said dot. These will then be broken down between participants, as we will only present each one of them with 4 of these 16 possible unique deviants.

We will then balance the trials for each participant with three controlled parameters, by evenly splitting them between matching trials (both sample and query are the same) and non-matching trials (one point was shifted between sample and query); this would be the match condition. For non-matching trials, deviants will be presented as sample in half the trials, the other half having the original as sample instead (and, consequently, the deviant as query); this would be the order condition. Finally, we will make it so that sequences were presented with their first dot to the left (normal display) or to the right (negative display); this would be the negativity condition. In total, each participant will thus pass $2^3 = 8$ test trials matched over these 3 conditions per sequence, or 80 in total. Participants also had 4 training trials prior to test, with 2 very simple and unbalanced sequences (with 1 or 2 B-items only), in both matching and non-matching conditions. We also introduced 10 control trials, with one popping every 10 test trials, one per test sequence, in which the query sequence was the negated sample, to test whether participants could distinguish left-right inversion. There were thus 94 trials in total, including training and control.

Measures We will measure success rate and reaction time as dependent measures.

Predictions Our hypotheses predict that success rate will be negatively correlated with complexity, and that reaction time will similarly display a positive correlation with complexity.

Analyses Reaction time will be converted to logarithmic values in order to reduce the correlation effect between its values and its standard deviation.

Analyses will be run by fitting a Generalized Linear Model (GLM), for both success rate and reaction time in parallel, with LoT-complexity of the sequence as a fixed factor, and participants as a random factor. Analyses will be separate for matching and non-matching conditions. Non matching conditions will also have the distance of the outlier on the sequence, and the outlier position relative to the sequence template as further fixed factors. We will also add Shannon surprise as a fixed factor for all conditions, to account for effects from transition probabilities. These analyses should find a significant, positive (resp. negative) effect of complexity on success rate (resp. reaction time).

Interpretation Should we find the significant results presented above, we will conclude that our hypothesis is verified. If not, the hypothesis will be rejected

0.4 Expected contributions

All my work will be mainly done by myself. Mathias Sablé-Meyer (MSM) and Stanislas Dehaene (SD) will provide guidance and feedback. MSM will provide templates and help for experiment code and analysis code.

CHAPTER 1

Introduction

Human language is perhaps one of our most characterizing feature as a species. Many other animals display communicative systems (like monkeys [Cäs+13; Coy+15] or songbirds [Ber+11]), yet none of these systems is considered equivalent to language. A key difference lies in the existence of "long-range dependencies" in human languages, where an arbitrarily long relative clause may be inserted between two words without affecting their relationship [Deh+15]. E.g. in English, the sentence "The car behind the truck is red", is still read as "The car [...] is red", and not, "[...] the truck is red". These dependencies are well highlighted by syntactic processes such as agreement: "The cars behind the truck are red". No similar observation was reported in animal productions [HCF02].

Generative linguists, Lashley [Las+51] and Chomsky [Cho56] ahead, offered a formal account for this intriguing discrepancy. They proposed that the syntax of human language uniquely relies on recursively nested structures [Cho57; Deh+15]: a sentence like above can be derived from the simple sentence "[The car]$_{DP}$ is$_V$ red$_{AP}$" with the substitution rule "$DP \rightarrow DP$ behind DP". Here this rule is applied to the determiner phrase (DP) "[the car]$_{DP}$", which is replaced by "[[the car]$_{DP}$ behind [the truck]$_{DP}$]$_{DP}$". The operation can be repeated at will, e.g. to get "The car behind the truck behind the bike is red".

Although we only gave one substitution rule in the above example, languages possess many, as even a simple sentence like "[[Mary]$_N P$ [likes$_V$ [John]$_{NP}$]$_{VP}$]$_S$" features (at least) two : $S \rightarrow NP\ VP$ and $VP \rightarrow V\ NP$. Crucially, the set of all substitution rules in a language, called its grammar, needs to be parsimonious. In fact, the sentence above may be obtained with a grammar such as $\{S \rightarrow NP\ NP;\ NP \rightarrow V\ NP\}$; however this grammar may also produce invalid sentences like "[[Mary]$_{NP}$ [likes$_V$ [likes$_V$ [John]$_{NP}$]$_{NP}$]$_{NP}$]$_S$". The constraint of parsimony heavily restricts the space of possible grammars for human languages. It has been mathematically demonstrated that, to account for all natural sentences and only them, a supra-regular grammar was needed [Shi88]. Supra-regular grammars rank high in Chomsky's hierarchy [Cho57], given that they can not be generated using simple mechanisms such as finite-state automata.

Since only human languages appear to use (or at least require) supra-regular grammars, a major question arose: are they accessible to other animals, or is it a human specificity?

Comparative studies tackled the question using artificial grammar learning (AGL) paradigms (for a review, see [FF12] (2012) or [Deh+15] (2015)). In AGL paradigm, subjects are taught a pattern involving categories of items (e.g. symbols or sounds), their ability to generalize the known pattern to novel categories being tested afterwards. The results of comparative AGL experiments are heavily disputed [Deh+15]. Recent research nonetheless suggests that some monkeys are able to learn and generalize at least one supra-regular grammar[1] [Jia+18; MDF20; Fer+20]. Although these studies challenge the view of supra-regularity being human-specific, they also highlight the huge advantage of humans in this field. Monkeys' generalization of the grammar comes at the cost of hundreds of trials, while humans, including preschoolers, only require a few of them [Jia+18]. Furthermore, despite their success, monkeys failed in generalizing another, more complex grammar, that humans could generalize [MDF20].

If anything, humans thus appear to have an edge in grammar manipulation and processing that may go well beyond natural languages. In 1975 already, Fodor [Fod75] proposed that human thoughts were structured along rules similar to natural languages: mental states (e.g. of belief) could be considered recursive, along the grammar $\{MP \rightarrow P; MP \rightarrow O\ R\ MP\}$, where P is a proposition (e.g. "the Earth is flat"), O an organism ("Mary", or "Philip"), and R a relation from organisms to mental representations MP ("believes that"). This may create mental states such as "Mary believes that the Earth is flat", but also "Philip believes that Mary believes that the Earth is flat".

Fodor's hypothesis of a "Language of Though" (LoT) was generalized in by Fitch [Fit14], who argued that many domains of human cognition actually featured recursive syntactic structures. Musical phrases display long-distance key agreement, even though a local, subdominant part may have a different key [Koe+13]; in mathematics, the formula 2x3+4 should be interpreted as the nested structure [[2x3] + 4]. This "dendrophilia" hypothesis, named after the tree-like structures involved in Chomsky's generative grammar [Cho57], received a significant attention over the years [PTG12; RSF13; Ama+17; Dav18; Sab+21; Pia21]. In the modern version of the LoT Hypothesis (LoTH), concepts are often thought of as the output of a "program", especially for domains with some abstract formalism, such as geometrical figures [Sab+21; Tra+21], sequences [YJ15; Ama+17; Pla+21], and numbers [PTG12; RSF13]. In particular, it has been proposed that subject performance on a given stimulus was directly related to the complexity of the program required to generate it [PTG12; RSF13; YJ15; Ama+17]. However, these domains have mostly been studied in isolation. Although Yildirim et al. [YJ15] observed that LoT-based knowledge could be transferred across modalities, they did not test it across two different domains.

In this work, we investigated the domain-generality of a potential LoT, by focusing on two simple domains: binary sequences of sounds and binary patterns. In Exp. 1, we look for LoT effects in an experimental design involving binary patterns; the case of binary sequences has already been covered by prior work from Planton et al. [Pla+21]. In Exp. 2, we investigate cross domain effects using in a priming-oriented experimental setup.

[1] More precisely, the mirror of language patterns, e.g. ABCCBA.

CHAPTER 2

State of the art

2.1 Interactions between mathematics and language

According to Fitch's [Fit14] dendrophilia hypothesis, the human brain processes many domains with recursively embedded tree structures. This is precisely how mathematical formulas are accounted for in programming languages, using Chomsky's formal grammars [Cho57]. As a result, the highly abstract domain of mathematics appears a typical candidate for dendrophilia [SS14; AD19]. Here, we propose a quick overview of the interactions between mathematics and music.

Syntactic priming Priming is a classical way to highlight a conceptual relationship between two stimuli. Kintsch & Mross [KM85] observed in 1984 that recognizing an existing word, such as "steel" was easier when presented before with a semantically associated word, such as "iron". This priming effect heavily relies on conceptual proximity, since the effect was much stronger than with a mere thematic relationship (e.g. "gate" was barely primed by "plane", even though they were linked by the context of an airport). A similar effect, called syntactic priming (or structural priming) exists across syntactic structures. People are more likely to utter a sentence in the passive form if they were themselves presented with a passive before that [Boc86]. This effect is known as syntactic priming, or structural priming.

Syntactic priming was used to highlight structural interactions between mathematics and music. Scheepers et al [Sch+11] observed in 2011 that the positioning of relative clauses in sentences was primed by the structure of a prior arithmetic formula: upon being presented the ambiguous, incomplete sentence "The tourist guide mentioned the bells of the church that…", participants would be more likely to use a plural verb in the relative clause (i.e. attach "that" to "the bells"; high-attachment) than a singular verb (i.e. attach "that" to "the church"; low-attachment) if presented a high-attachment-like formula like "80 – (9 + 1) x 5" prior to this. Conversely, they were more prone to complete the sentence with a low-attachment clause if primed with a low-attachment-like formula such as "80 – 9 + 1 x 5". his mathematics-linguistic syntactic priming effect has since then been replicated numerous times [SS14; NO18; HCW20] including in other languages than English [PHS18; ZML18]. As a note, further works highlighted syntactic interactions across other domains [VH16], especially music [FP14; VH16].

Brain imaging Another way to probe relationship between concepts is to observe the common networks involved in their processing. Amalric & Dehaene [AD19] scanned the brain of mathematicians under fMRI while they were processing simple assertions of either general knowledge or mathematics. Participants had to label these assertions as true or false in a very short time (2.5s). For mathematics, assertions ranged from rotely memorized formula to sentences regarding properties of geometrical figures or complex numbers. General knowledge was used as a proxy for language.

Mathematical statements were associated with the activation of a distinct language from general knowledge. This network was centered on bilateral IntraParietal Sulci (IPS; notably involved in number perception [Pia+04; Fia+07]) and Inferior Temporal (IT) regions. The regions were not only significantly more activated when presented with mathematical statements; they also had a significant drop of activity when general knowledge statements were presented instead. On the contrary, language-specific activation appeared around the temporal pole and the pars orbitalis of the Inferior Frontal Gyrus (IFG): only geometry- and complex-numbers-related statement did trigger activation along with general knowledge; other mathematical statements were presented as formulas. Quite notably, this held true even when formulas where heard and not read by the participant, refuting low-level interpretations. Other regions (pars triangularis of the IFG and the Superior Temporal Sulcus (STS)) displayed significantly greater activity for language, but were still activated by mathematics.

A second experiment dropped formulas and focused on natural-language assertions of similar grammatical complexity, still covering either mathematical or general knowledge. The set of stimuli also now included negated and quantified statements (e.g. "some oceans are warm"). Under these knew conditions, maths-specific activation was still found in bilateral IPS and IT. Similarly, greater activation for language was still observed in the STS and the left pars orbitalis of the IFG.

Although these results provide strong evidence for dissociated networks in mathematics and language, it must be noted that the study focused on semantic differences between the two. This is especially true for the second experiment, which specifically controlled for syntactic discrepancies! As a result, the possibility for a shared syntactic network between maths and language two domains remains open.

In the following sections, we will focus on simpler domains which are less likely to bear higher-order effects. In particular, we will cover previous research on LoT-interactions pertaining to sequences of items.

2.2 A first take on cross-modality

Yildirim & Jacobs' experiment Yildirim & Jacobs [YJ15] provided a first experimental take on LoT interactions across modalities. They investigated transfer learning, that is, whether knowledge acquired in the auditory modality could be applied to the visual modality, and vice-versa. In Yildirim & Jacobs' experiment, this knowledge was a set of categorization rules. Participants were in fact presented with sequences of items generated by 4 possible rules, resulting in 4 categories they had to learn. The sequences had 7 possible different items, ordered on an imaginary circle, resulting in a succession relationship between them: item B was the successor of item A; item C was of B; up to A which was the successor to item G. The 4 possible rules defined the relationship between the item of the sequence. As an example, the first one, [+1], imposed that every item in the sequence was obtained from the previous one by a clockwise shift on the imaginary circle (e.g. “BCDEFGABC...”).

Participants learned the categories through trial and error. They took a training session (with auditory feedback) of 7 blocks of 36 trials, with 9 exemplars of each of the 4 categories. Once the sequences learned, they took 56 further trials in a subsequent test, this time without feedback. Testing trials included both sequences from training (n=9), and new sequences from each categories (n=6). Crucially, training and testing did not occur in the same modality.

For half the participants, training sequences were visually displayed. The display used 7 LEDs placed regularly on a circle. Each LED corresponded to a sequence item, so that links between LEDs matched those of the items (e.g., shifting the LED for item A by a 1/7th clockwise rotation gave the LED for item B). Sequences were then played using a flash of the associated LED. For the other half of the participants, training sequences were displayed with audio sounds. This display used audio emitters instead of LEDs, organized in an identical way. Instead of flashing a light, the emitters produced a small beep to play their items. The beep was common across emitters: participants perceived sequences as a series of sound locations, and not of different sounds.

Results After training, participants were tested on their converse modality. What was tested was thus their ability to translate the rules for visually-perceived spatial locations into rules for the same locations, this time perceived auditorily (and vice-versa). Yildirim & Jacobs. found that their participants were able to do so, since their 18 participants (9 for each modality) did significantly better than chance on training.

Three further observations were of interest to us. First, participants trained in the visual modality achieved better performances than those trained in the auditory one. The authors argued that this was most likely due to differing reliability between the two modalities. A second observation was that participants did not fare significantly better on familiar sequences than new ones, showing some degree of spontaneous generalization. The third, last observation was that classification performance varied across categories. As an example, ~ 90% of sequences from category 1 (following a simple [+1], clockwise turn rule) were correctly classified, while only ~ 80% of sequences from category 4 were. Category 4 followed a [+2 -1] rule which was to turn two times clockwise to get the second item from the first, and then one time counterclockwise to get the third item from the second, and repeating this pattern. One example of such sequence would be “ACBDCED. . . ”. The differences in performance across categories occurred in both training and testing.

Model Variations in difficulty are typically interpreted as signatures of a hypothesized LoT [AD19; Pla+21]. Here, Yildirim & Jacobs propose a LoT generating possible rules as programs. Programs combined LoT-primitives, including probabilistic ones, so that their output was not entirely deterministic. One program could thus generate several sequences. The assignment of one specific program (i.e. rule) to a category used a model balancing the program’s precision (its ability to generate the attested examples, no more no less) and its simplicity. The details of this model are beyond the scope of this thesis; see Piantadosi et al. [PTG12] for more information on probabilistic LoT-grammars, which Yildirim & Jacobs [YJ15] adapted for their model. This LoT-based model was able to account well for participants' data, showing similar patterns of performance on the same trials set. Namely, once categories ranked by performance, the order was the same for human participants and the model.

Limitations The results from Yildirim & Jacobs brought strong evidence that some common knowledge was used in both visual and auditory modality. It also demonstrated that the learning and use of this knowledge could be well accounted for by a LoT-based approach. However, this work undergoes three limitations.

First, although perception occurs in two different modalities, both target the same domain of spatial locations. From a LoT-perspective, this experiment may thus be explained by multisensory semantics (relative to spatial locations) rather than a shared syntactic module across domains. However, LoT is hypothesized to occur even across a variety of domains [VH16], linked by syntactic effects, such as priming [SS14].

Another limitation of Yildirim & Jacobs is that participants were aware of the existence of a common rule across modalities: they were told that visual categories were the same as auditory ones. Therefore, this experiment does not provide any evidence as to how spontaneous the observed transfer-learning is. This also opens the way to convoluted conversion strategies across modalities.

A final limitation of this experiment is that, although its design is rooted within geometry (with regular positions on a circle), it does not investigate geometric properties such as symmetry. Although their heptagonal design precisely limited the possibilities for symmetries to play a role, their proposed LoT consequently may not be able to properly capture operations run by the human brain. In the following section 2.3, we will present an alternative LoT proposed by Amalric et al. [Ama+17], aimed to better cover geometric abilities. The other two limitations are related to cross-modality and will be covered in Exp. 2 of the present thesis.

2.3 Geo, a language for patterns in the brain

2.3.1 Formal definition of Geo

Amalric et al. [Ama+17] proposed a formal language, Geo, dedicated to describing sequences of movement over a regular octagon. Regular octagons are very convenient polygons for pattern perception, given that they display a number of interesting symmetries. When drawn so that they have one horizontal segment (see Fig. 2.1), they display a horizontal symmetry H, a vertical symmetry V, two diagonals symmetry A and B, and even a point symmetry P. These regularities allow to describe movement on octagons in a very compressed manner.

The language Geo thus contains two types of movement descriptors: rotation-based descriptors, and symmetry-based ones. Rotation-based descriptors are similar to those used in Yildirim & Jacobs [YJ15], as they correspond to a number of rotations clockwise or counterclockwise. As an example, 0 denotates staying at the same location, +1 denotates a movement to the next element clockwise, while −2 denotates a movement to the second element counterclockwise. Symmetry-based descriptors are specific to Geo and describe movement along the symmetry axes and center presented above. As an example, point-symmetry movement P is equivalent to a +4. See Fig. 2.1 for all descriptors and a movement example.

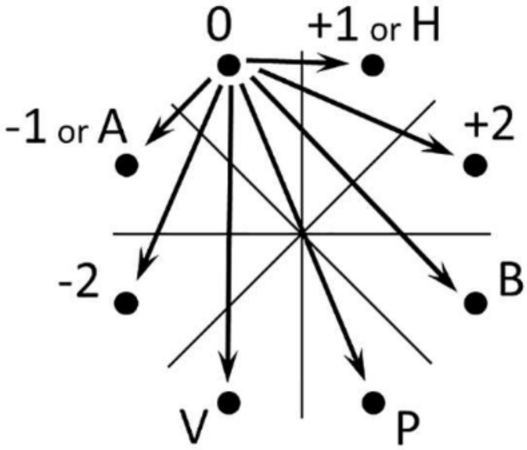

Fig. 2.1. Representation of Amalric et al.'s stimuli octogon, and the possible primitive instructions of Geo.

Using these elementary descriptors, as primitives for Geo, one can already describe any sequence on the octagon. As an example, a sequence of movements in a square could be expressed as +2 +2 +2 +2. However, Geo provides a way to compress these repeating sequences, through the $[X]^n$ operation, where X denotes an expression and n the number of times it is repeated. As such, +2 +2 +2 +2 can also be equivalently expressed as $[+2]^4$. Repetitions can also include some variation between them. Envision one wants to draw a square over 4 points of the octagon, and then draw another square over the four remaining points. Without repetition, it could be expressed as +2 +2 +2 +2 (first square) +1 (offset) +2 +2 +2 +2 (second, offset square). With simple repetition, it could be simplified as $[+2]^4$ +1 $[+2]^4$. Variations allow to express this under the form $[X]^n < I >$, where I is a single instruction (movement) offsetting every repetitions. Our expression can thus be expressed as $[[+2]^4]^2 < +1 >$: there are two squares ($[+2]^4$), offset by +1 from one another[1].

Geo thus propose a way to represent sequences over an octagon in a recursive way. It combines 10 primitive instructions (rotation[2]: 0, +1, +2, −1, −2; symmetry: H, V, A, B, P) with 3 primitive operators (concatenation [,], repetition $[]^n$; and repetition with variations $[]^n <>$). These representations can be numerous for one single sequence, as seen above for the 2-squares example. However, some are more compact than others: this compactness is measured by description length. Description length is computed as a measure of the number of primitive used, potentially weighted by primitives. In Geo, the description length of an expression is defined recursively as follows:

1. the length of any instruction is 2[3];

[1] Crucially, the offset instructions apply with respect to the repeated sequences start, and not their end. The fact that $[[+2]^4]^2 < +1 >$ is equivalent to $[+2]^4$ +1 $[+2]^4$ is a coincidence only due to the fact that $[+2]^4$ both starts and ends on the same point. $[[+2]^2]^2 < +1 >$ is, as an example, not equivalent to $[+2]^2$ +1 $[+2]^2$, but to $[+2]^2$ −3 $[+2]^2$.

[2] Rotations beyond 2 were not used as primitives by Amalric et al. Note that these absent rotations can be obtained from ±1, ±2, and P (e.g. +3 is [P; -1]; -3 is [P; +1]).

[3] The choice for instructions being of length 2 was made so that +2 +2 was stricly longer than $[+2]^2$. This

2. the length of a consecutive expressions is the sum of each expression's length;
3. the length of an expression iterated n times is the length of this expression plus $log(n)$;
4. the length of a repetition with variations is the length of the repetition plus the length of the variation instruction.

Given a sequence of movements, there is at least one expression whose length is minimal, that is, so that there is no other expression for the sequence with a smaller length. This minimal length is called the *Minimum Description Length* (MDL) of the sequence. As they hypothesized LoT to use the most parsimonious representation, Amalric et al. assumed MDL to be a good descriptor of how complex the LoT-representation of the sequence was. From now on, we will call MDL the *LoT-complexity* of a sequence.

2.3.2 Geo across populations

Amalric et al. [Ama+17] proceeded to test whether Geo could account well for human performance on geometrical tasks. French adults were asked to predict sequences of flashing lights on the corners of an octagon. Sequences were chosen so that their LoT-Complexities (according to Geo) spanned a wide range of possibilities, from irregular, incompressible sequences, to simple series of 1-element rotations. For each sequence, its two first items were flashed with possible variations on the starting point. Participants had to predict the following items one by one, being corrected for wrong guesses (so that they could predict the rest).

Amalric et al. found that participants made much more errors for sequences with a high LoT-complexity than for more regular ones. Overall, the complexity computed by Geo correlated very well with error rate. Crucially, some sequences' regularities relied heavily on symmetry. The better performance on those sequences, when compared to their irregular counterparts, demonstrated how important symmetries were to the hypothesized LoT. In fact a degraded version of Geo, without symmetries, performed much worse on these sequences, while the overall prediction was still decent. Amalric et al. also ran item-by-item analyses, showing that errors in sequences were higher for movements (=instructions) nested deeper in a Geo-generated expression for the sequence. Geo thus appeared as a good LoT-account for performance of French adults on this geometrical task.

Given the geometrical nature of the task, a potential LoT effect could however be heavily driven by formal education in mathematics. To investigate this possibility, Amalric et al. studied preschoolers and Munduruku adults on a simplified version of the task. As Mundurukus are people from an isolated Amazonian tribe, with little to no formal education, any shared difference could only be explained by education, and not age or culture.

Amalric et al. found that Geo was able to once again predict performances fairly well for these populations, although less than for French adults. The difference was greater for sequences involving point symmetry P, and higher level of nesting. Although the performance improved when P and nesting were made more complex (had greater length), it dropped when they were forbidden. These results thus hint that there is a difference between educated and non-educated difference, lying in how complex some LoT-operations are.

made compression optimal even on simpler repetitions

2.3.3 Geo in the brain

FMRI observations Beyond behavior, for which Geo accounted well, this language was also tested using brain imagery. Wang et al. [Wan+19] first investigated participant's brain responses on Amalric et al.'s [Ama+17] task when they simply had to follow with their eyes items as they moved. Wang et al. observed that the main hierarchical processing areas involved were the Inferior Frontal Gyri (IFG), the right Dorsolateral Prefrontal Cortex (DLPFC), and bilateral anterior caudate. Although the ventral part of IFG is often considered a core language network, the activation here was more dorsal, replicating observations from Amalric & Dehaene [AD19] on mathematics versus language. The IFG was found to correlate mostly with the overall sequence complexity, while the caudate and DLPFC were particularly responsive to the degree of embedding of the representation structure predicted by Geo.

A MEG-oriented experiment Al Roumi et al. [Al +21] adapted Amalric et al.'s [Ama+17] paradigm, in which participants had to indicate their predictions for the upcoming movements on an octagon. Participants instead had to report violations of a sequence they learned prior to the test. This adaptation allowed to smooth imagery results by simplifying participant's action: pressing a button (violation report) instead of 8 (prediction). It also separated the learning part from the testing part, a welcome modification when analyzing brain imagery.

The learning part consisted in 12 repetitions of a 8-items sequence, flashed over the octagon as in Amalric et al. [Ama+17]. If participants felt they learned the sequence before the end of this part, they could stop early by pressing a button. This offered a first, dependent measure of the task complexity as "encoding time". Once the learning part over, either by the participant or by the 12-repetitions limit, the sequence was played 12 times again, this time with violations. Participants had to report violations by pressing the button again, as soon as they spotted them. This provided two further dependent measures of reaction time and wrongness of the answer.

Beyond sequences, Al Roumi et al. also tested possible primitives, that is single movement instructions. These were taken from Geo's instruction primitives (minus the stay one, as none of their tested sequences included twice the same position), with the additional ± 3 rotations. Primitives were tested in a way analogous to sequences: participants had to learn them during 32 repetitions (which they could still stop), and then spot violations again on 32 repetitions. Unlike for sequences, the starting point for the primitive movement changed across repetitions, to prevent rote memorization.

Behavioral results Al Roumi et al. observed that the three dependent variables nicely predicted one another, and that they correlated well with LoT-complexity on sequences. Still, the inclusion of ± 3 rotations and a distance-between-positions effect improved the fit of the model. Primitives were also significantly easier than a control condition, where movements could only be rote memorized from the starting point. These result corroborated Amalric et al.'s [Ama+17] finding with Geo, but pushed for the integration of two ± 3 additional primitives and visual distance into the language. These results were furthered by a later computational approach comparing different languages [Rom+18], which nonetheless favored Geo and this amended version over languages with arbitrary primitives such as following the digits of π.

MEG decoding In addition to behavior, the essence of Al Roumi et al.'s study was the brain imagery of the participant under Magneto-Encephalo-Graphy (MEG). Using decoders, they were able to predict better than chance which position was being flashed from brain activity. LoT was not correlated with how accurate the prediction was, but participant performance was. Beyond the currently flashed position, they were able to predict the upcoming flash position better than chance as well. This time, the accuracy of the prediction was influenced by LoT-complexity: prediction was less accurate for more complex sequences. This result matches the error patterns of the participants, with higher error rate on complex sequences, and hints at a LoT-driven anticipation in participants.

Al Roumi et al found further indirect evidence of a LoT like Geo in the brain. They trained a decoder for which primitive was displayed in the part where participants had to detect violations of primitives. The decoder was able to classify them better than chance. This classification was not only sensitive to visual features. Visually, primitives amounted to movements on the screen; yet two primitives could, from the same starting position, fall onto the same ending position. This is the case for +1 and horizontal symmetry H when starting from the topmost left corner of the octagon: both will end on its right, horizontally aligned counterpart (see Fig. 2.1 for illustration). The only difference between the two is their context, as one occurs within a block of +1 movements, and the other in a block of H movements. Brain activity still differentiated such trials, as the decoder was able to infer the correct category. Therefore, Al Roumi et al. provided here evidence that the brain encoded the abstract geometrical symmetries to an extent, well beyond the only visual features.

Al Roumi et al. were not only able to decode primitives from brain activity. Within a movement predicted by Geo to be in a loop, they succeeded in decoding the number of the iteration step better than chance. Overall, their imagery work provided numerous, yet still indirect, evidence of LoT in the brain, strengthening the Language of Thought Hypothesis (LoTH) but not testing it decisively.

2.4 Adaptation to binary sequences

In the wake of the difficulties faced to test LoTH, studies tried to investigate the most minimally structured case of patterns: binary sequences. In the following parts, we will refer to sequences as series of items A & B.

2.4.1 Investigating the signature of LoT

Binary-Geo Planton et al. [Pla+21] amended Geo for cases where only two items were possible: among instructions, only 0 (staying on the same item) and P (point symmetry; i.e. swapping to the other item) were relevant. They kept concatenation [], repetition $[]^n$ and repetition with variation $[]^n$ operators as primitives. The individual cost (length) of primitives were kept from the original language.

Experimental setup Planton et al. then tested how LoT-complexity would account for participant performance on a task. As Al Roumi et al. [Al +21], they used a violation detection paradigm. Upon being habituated to a sequence of sounds (with only two possible different sounds), participants had to detect violations in subsequent repetitions of this sequence.

Violations consisted in having one single item of the sequence being changed: in most cases, the wrong sound out of the possible two was played (creating a deviant); however, it also happened that a third, unrelated sound was played instead (super deviant).

Planton et al. tested participants on a variety of sequences, across 5 independent experiments. Each experiment only displayed sequences of same length, but chosen so that they spanned a wide range of LoT-complexity as computed by Geo. The first 4 experiments involved auditory sequences of length 16, 12, 8, 6. The 5th experiment included not only auditory sequences of length 8, but also visual sequences of same length to test if effects held across modalities. Length-16 sequences, which we reused in our experiments, can be seen in Fig. 2.2.

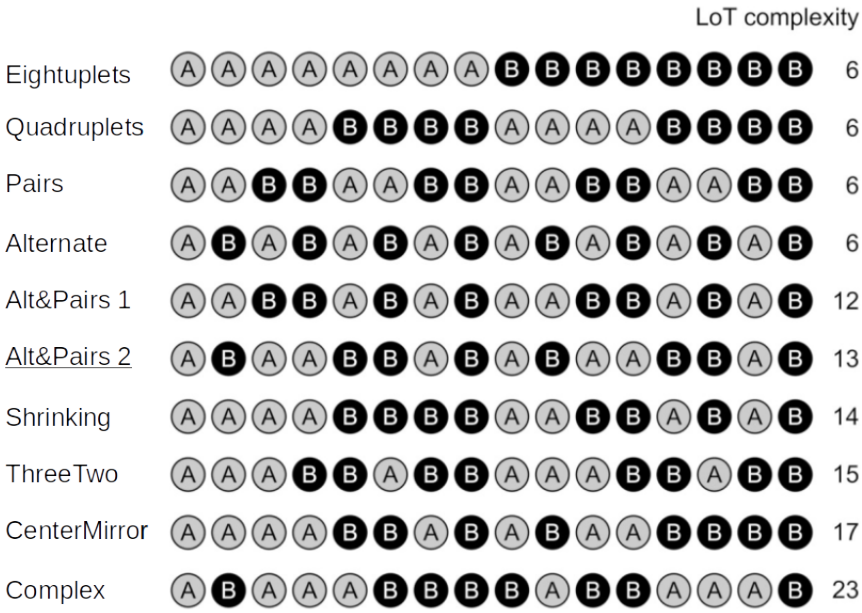

Fig. 2.2. Sequences from Planton et al., with their name and LoT complexity. All sequences were reused in Exp. 1, except for the underlined sequence (Alt&Pairs2), which had its later item swapped due to a coding error. Resulting complexity was 19.

LISAS measure Participant performance was computed using the *Linear Integrated Speed-Accuracy Scor (LISAS)* measure [Van17]. This measures combines both accuracy and reaction time over a series of trials into a unique measure for each participant, along the following formula.

$$LISAS = RT_j + S_{RT} \times \frac{ER}{S_{ER}}$$

Formally, the LISAS performance of a participant j is the sum of their reaction time RT_j and the standard deviation of reaction times across participants S_{RT}, pondered by the error rate ER_j, reduced with the standard deviation of error rates S_{ER} across participants.

Results For all lengths and modalities, Planton et al. found a correlation of participant performance with LoT-complexity. Within auditory modality, LoT-complexity was even a very good predictor of performance across different sequences length. Furthermore, as participants were asked to rate sequences by complexity, it was found that LoT-complexity closely matched this subjective metrics.

2.4.2 Beyond the Language of Thought

Binary sequences are a well explored domain of human cognition ([MMD22]), but it has also been abundantly considered by computer science and information theory [Sha48]. As a result, many studies have bridged the two disciplines: information-theoretic measures were proposed as predictors of human behavior when facing (among others) binary sequences.

Shannon surprise Shannon surprise [Sha48] (hereafter surprise) is a measure of how unexpected a signal is, given a set of statistics. Formally, the surprise associated to an event E is defined as the negative logarithm of its probability p_E of occurrence.

$$S(E) = -log(p_E)$$

In Planton et al. [Pla+21] , and likewise in our subsequent experiments, the surprise considered is the one related to transition probabilities (TPs). In fact, humans (but also non-human animals) have been shown to track to transitions between items in the form of probabilities [Deh+15]. To account for this in their models, Planton et al. computed the surprise associated with the deviant D given the preceding point X:

$$s_{X \to D} = -log(p(D|X)) = -log(p_{X \to D})$$

Entropy Entropy is a measure of how uncertain a sequence of events is [Sha48]. If only one event can occur, the sequence is deterministic; if all events are equiprobable, no prediction is possible whatsoever. In other words, a sequence is as uncertain as its elements are individually surprising. To capture this, the entropy H of a series of events S is defined as the average surprise of the events; by considering the proportion p_E of event E in S, H can be formulated as:

$$H(S) = -\sum_E p_E log(p_E)$$

To integrate the tracking of both TPs and individual items occurrence probabilities, Planton et al. considered the entropy of sequences on the basis of their item pairs. In fact, given a pair of item X, Y (possibly identical), the probability P_{XY} of the pair XY is defined as $P_X.P_{X \to Y}$. As a result, entropy in our case was defined as:

$$H = -(p_{AA} log(p_{AA}) + p_{AB} log(p_{AB}) + p_{BA} log(p_{BA}) + p_{BB} log(p_{BB})$$

Chunk complexity Chunk complexity, proposed by Mathy & Feldman [MF12] is a measure of how much a sequence can be chunked into subgroups with some constant regularity.

Planton et al. defined these subgroups as repetitions of a single item [Deh+15] (e.g. BBBB)[4]; note that the chunking was assumed to be maximal, that is, there could be no consecutive chunks repeating both the same item. Following Mathy & Feldman, the chunk-complexity of a sequence is defined as:

$$CC = \sum_i log(1 + L_i)$$

where L_i is the length of chunk i. As an example, the sequence $AABBBA$ would have a chunk complexity of $log(3) + log(4) + log(2)$. Note that, for a given sequence length, chunk complexity is maximal for alternations; more generally, chunk-complexity could be understood as how fragmented the sequence is.

Other predictors Planton et al. also introduced further predictors from information theory. However they did not perform better than LoT-complexity or the previously described predictors, neither in their experiment nor in ours. For the full description of their predictors, see the study by Planton et al. [Pla+21].

Lot-chunk variant The last predictor introduced by Planton et al. actually amended Geo. While it didn't change its primitives, it added the constraint that valid expressions could not divide chunks. As an example, the sequence $ABABBB$ couldn't, under this constraint, be interpreted as $(ABAB)BB$. To borrow a term from formal linguistics, this means that chunks are constituents of the sequence. Since this only constrains expressions considered as valid, using LoT-chunk instead of LoT will not decrease the complexity of sequences; however, some might see their complexity increase if their minimal expression was violating the added constraint.

Empirical comparison of predictors The predictors introduced above were compared to LoT-complexity in their prediction of participant performance on auditory sequences. Surprise was found to be a significant predictor, independently of LoT. However, prediction was oftentimes better when the two were combined.

The other predictors were analyzed in competition with LoT. Both the original LoT scored far better than its competitors, only being outmatched by its LoT-chunk variant. This held true both with and without surprise as an additional predictor.

Overall, Planton et al. brought indirect existence for a mechanism akin to a "Language of Thought" when processing binary sequences of sound. This evidence relied on the identification of complexity signatures in participant performance for a range of stimuli. In the following Exp. 1, we adapt this paradigm to another, close domain: binary patterns.

[4]Note however that different definitions could be used: Mathy & Feldman, who used a task of digits sequences memorization, defined chunks as groups of digits with constant increment or decrement (e.g. 34567 or 8642).

CHAPTER 3

Experiment 1: On the complexity of binary patterns

3.1 Methods & Participants

Participants We recruited from Twitter 49 participants, 10 of whom where filtered out due to them not being significantly better than chance (n=9), or being too slow (median response time 3 times higher than average; n=1). This left us with a sample of n=39 participants (15 females; mean age = 39.1; 3 missing demographics data).

Procedure In this experiment, we tested participant memory on a range of visual binary patterns using the same-different paradigm. As in a violation detection paradigm, participants had to detect in a query pattern deviations from a target pattern. The main difference is that this pattern was not learned through a habituation phase; instead, it was briefly displayed shortly before the query. In addition to this, participants faced a forced choice: rather than tagging violations, participants were asked to tell whether the target and the query were matching. Participants could answer directly upon query presentation. However query was only displayed a short amount of time; to prevent retinian persistence effect, a brief static mask was added after both presentations.

Stimuli We used the 10 length-16 binary sequences used in Planton et al. [Pla+21] as templates[1] (see Fig. 2.2). These templates presented various LoT-complexities (as computed by the binary version of Geo), but consistently had as many items (eight) of each category. Deviated versions were generated from templates by replacing a single item, from one category to another. This deviant item, or outlier was randomly chosen within the template. While it did unbalance the two categories (since one now had 9 items and the other 7), it kept the number of outliers minimal, reducing the effectiveness of strategies rote memorizing only parts of the sequence.

[1] Due to an error in the code, one sequence was actually modified, as its last item was of the wrong category. Complexity was adapted to the erroneous version in latter analyses.

Item sequences were displayed as static patterns at the center of the screen, aligned along the vertical axis. Items were displayed as black dots shifted either to the left (category A) or to the right (category B), and from top to bottom. The target pattern was presented 1200ms, after a 50ms blank its mask was flashed during 500ms; after another 1s blank, the query pattern was presented similarly to the target. The masks consisted in a static display of all possible item positions, that is of two vertical trails of dots shifted to the left and the right from the center. See Fig. 3.1 for an illustration.

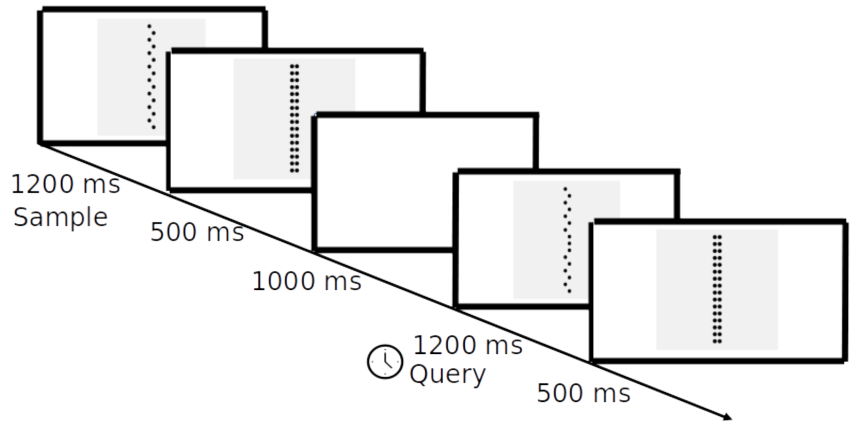

Fig. 3.1. Illustration of the time course of the experiment. 50ms blanks between sequences and masks were omitted. Here, this is a deviant trial given that one dot was moved in the query and the template was the "Alternate" sequence (in target). The clock indicates that we started measuring reaction time upon presentation of the target.

Across the whole experiment, test trials were presented in a random order. However, each of the 10 templates was used 8 times in total; this defined a balanced template condition. Within template condition, the trials were further balanced with three controlled parameters. First, half the 8 trials featured a deviated pattern; this was the deviancy condition. When no deviated pattern was introduced, both target and query patterns were visual displays of the template, and were thus matching; otherwise, only one of the two patterns was a deviated version, and thus target and query were different. Which of the two was deviated made the order condition. This condition was balanced as well, with 2 out of the 4 non-matching trial having a deviated target, and the other two a deviated query.

To artificially increase the variety of patterns displayed to participants, we also made it so that some of the trial used a negative display. Under such occurrences, the categories were mapped to the opposite side of the screen: an item of category A would be shifted to the right instead of left. This was of course consistent across target and query. This negativity condition was balanced as well, so that each combination of the previous conditions would feature as many trials under normal display as trials under negative display.

We thus tested participant with a 10 (template) x 2 (deviancy) x 2 (order) x 2 (negativity) balanced design, although order had no effect when the two sequences were matching. In addition to this, we made sure that the 4 outliers appearing for each template were all different from one another. This resulted in 80 test trials per participants, with 4 outliers out of the 16 possible ones for each template.

To familiarize participants to the task, we also presented them with 4 training trials. The displayed pattern was obtained from a quasi-monotone template (that is, all the items belonged to the same category, except for one or two; e.g. $AAABAABA$). In addition to training trials, control trials were integrated to the testing part, with one every 10 trials. In these controls, the query pattern was the negated version of the target. The target itself was a visual display of a template, and every template occurred once within these 10 control trials. This allowed us to test whether participants would distinguish left-right inversions.

3.2 Results

Every participant took 94 trials, 80 of which were testing them. Participants as a whole went through 3510 test trials in total, 68 of which were further filtered out due to abnormally long response times (over 3200ms, i.e. 2s after end of display). As a measure of performance, we will focus on error rate, as it was strongly correlated with reaction time (simple linear model predicting error rate by reaction time for all sequence: $R^2 = 0.93; \beta = 833ms; p < 0.001$).

Condition differences The error rate reached 26% on average across participants ($d' \sim 1.5$; see Fig. S1). However, these errors were heavily driven by deviancy cases: participants had a hard time spotting most sequence alterations (average miss rate: ~ 37%); conversely, participants rarely reported differences for two identical sequences (average false alarm rate: ~ 15%). A chi-squared test found this asymmetry in deviancy condition significant ($\chi^2 = 184.34; df = 1; p < 0.001$; see Fig. 2A.). We ran similar tests for negativity condition (related to the side on which the first item was displayed), which yielded no significance ($\chi^2 = 0.038; df = 1; p = 0.84$). Order condition, related to which of the template or the deviated pattern appeared first in deviancy cases, was also not significant ($\chi^2 = 0.21; df = 1; p = 0.65$). Template condition revealed that some sequences were however significantly easier to process than others, with error rates ranging from 9% to 36% ($\chi^2 = 109.07; df = 9; p < 0.001$; see Fig. 2A.).

Outlier position We then proceeded to give a closer look to deviancy cases, in which featured a variety of possibilities for the outlier position. Deviants occurring on later positions were mildly but significantly harder to spot (linear model over average error rate, for each position: $R^2 = 0.42; \beta = 1,6x10^{-2}; p = 0.006$).

Looking at the actual curves of error rate by outlier position (see Fig. 3.2 and Fig. S1), we observed the first three positions seemed to trigger considerably less errors: 23% less on average! We confirmed that this effect was significant using an ANOVA separating these trials from other deviant trials ($F = 22.27; df = 1; p < 0.001$).

When excluding those positions, outlier position appeared no longer correlated to error rate ($R^2 = 0.07; \beta = 0.6x10^{-2}; p = 0.38$). Nonetheless, a non-linear effect was still present as error rates still differed by position (chi-squared test over deviant trials with outlier beyond 3rd position: $\chi^2 = 26.697; df = 12; p = 0.008$).

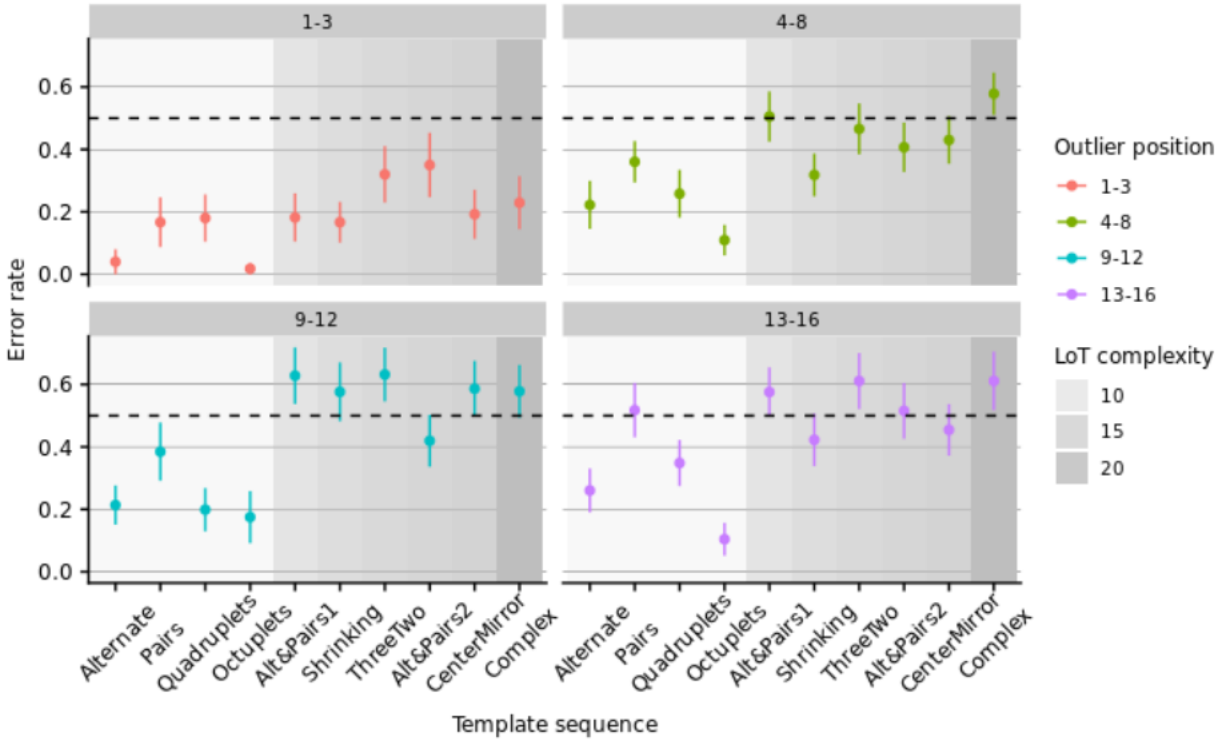

Fig. 3.2. Average error rate on deviant trials, depending on the template and the position of the outlier. The first group only features 3 positions because these have particularly low error rates. Bars stand for 1 standard error across participants. Dashed line represents chance level (50%)

LoT-complexity We then proceeded to investigated how difficulty varied within template condition. We tried to predict the average error rate across participants given the LoT-complexity of the template using a simple regression. As deviancy condition was shown to significantly affect performance, the analysis was run separately for matching and deviant trials. Under both conditions was LoT well correlated with errors (matching: $R^2 = 0.784; \beta = 0.51x10^{-2}; p < 0.001$; deviant: $R^2 = 0.623; \beta = 1.8x10^{-2}$; see Fig. 3.3 A & B.).

To rule out possible confounding factors, we joined LoT-complexity with other predictors in binomial Mixed-Effects Models. For matching trials, a mixed model was used to predict single-trial outcomes (whether answer was wrong), with LoT-complexity and the index of the trial (as a proxy for learning) as fixed effect, and a random intercept per participant as a random effect. To allow for effect size comparisons, all fixed effects were rescaled. All the effects came out significant (LoT: $\beta = 0.22 \pm 0.07; p = 0.001$; index: $\beta = -0.17 \pm 0.07; p = 0.02$). For deviant trials, we added outlier group (first three positions or other) and the Shannon surprise associated with the deviant as fixed effects on top of the others. Again, all fixed effects came out significant (LoT: $\beta = 0.36 \pm 0.06; p < 0.001$; surprise: $\beta = -0.40 \pm 0.08; p < 0.001$; outlier group: $-0.38 \pm 0.07; p < 0.001$; index: $\beta = -0.12 \pm 0.05; p = 0.038$).

We also investigated whether LoT was still a relevant predictor for deviant trials where the outlier was on the first three positions. Using a simple LM, we still found a significant, albeit less decisive effect ($R^2 = 0.41; \beta = 1.1x10^{-2}; p = 0.046$).

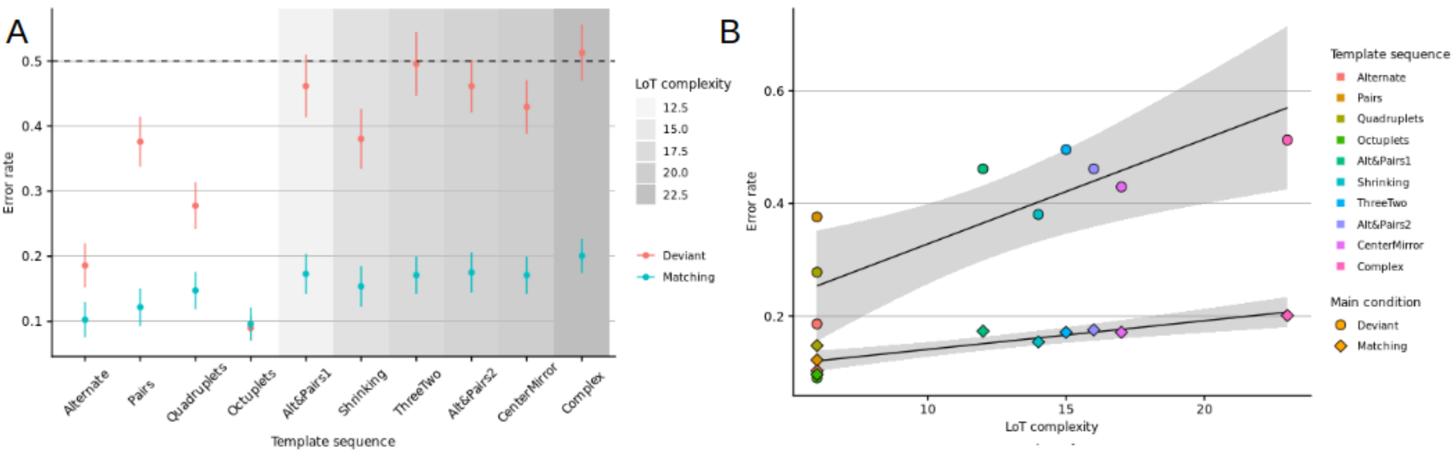

Fig. 3.3. A) Error rate depending on template, for both deviancy conditions. LoT-complexity is highlighted in the background. Error bars represent 1 standard error and the dashed line chance level (50%); B) Correlation of error rate with LoT-complexity. Each point represent performance over one template in either deviancy condition.

Predictors comparison We further tested how the LoT-complexity predictor compared with other complexity measures, already used as competitors by Planton et al. [Pla+21]. We adapted the models from above, replacing the LoT-complexity measure by one of its counterparts as a predictor. Models were compared on the basis of the Akaike Information Criterion (AIC) [Aka98], a measure of how good the model's fit is, while penalizing its complexity (degrees of freedom) for a given set of data. The best models have the lower AIC. To assess the relevance of the AIC difference between our models, we computed their relative Akaike weight (wAIC) [WF04]. Given a set of models, Akaike weights can be interpreted as the probability that the associated model is the best of the set. This measure is thus only valid as a comparison tool between competing models.

In matching condition (Fig. 3.4A), in which trials featured the exact same sequence as sample and as query, we fitted 7 different models predicting false alarms from one of the complexity measures. These models were Mixed-Effects Models similar to the ones previously used for the condition: they included learning as a covariate fixed effect and participants as a random effect. The best predictor for errors was LoT-complexity, with a wAIC of 0.30. The LoT-chunk complexity variant fitted almost as well, with a wAIC of 0.24. Other predictors' models had weights between 0.17 and 0.01, except chunk-complexity which was by far the worst predictor (wAIC = 0.003).

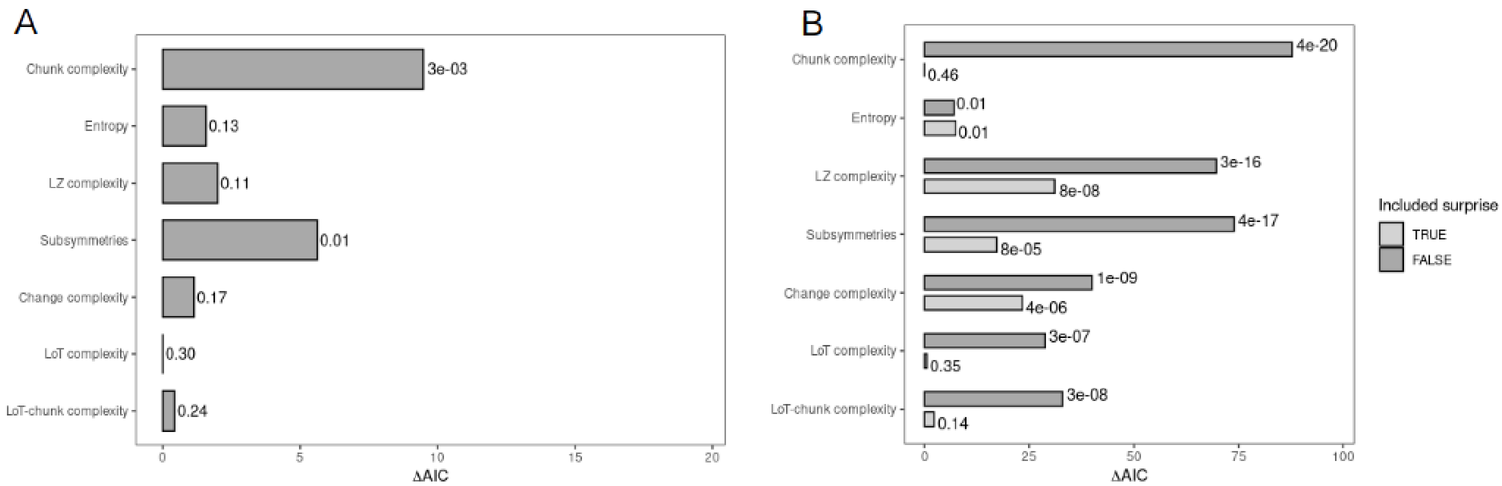

Fig. 3.4. Comparison of the different predictors for performance in Exp. 1. Models are compared to the best. Numbers at the tip of the bar represent Akaike weights of the model within the set. Light gray indicate that surprise was included as a predictor A) Matching condition; B) Deviant condition.

In deviancy condition (Fig. 3.4), we fitted 7 different pairs of models, again similar to those used for this condition. 7 models predicted misses from a fixed complexity effect (with each of our 7 measures) with a random participant effect and learning and outlier position as covariate fixed effects. The remaining 7 models also included surprise as a covariate fixed effect, in line with Planton et al. [Pla+21]. Similarly to them, we found that surprise improved the AIC for all models but entropy. Unlike this previous study however, the model with the best fit was the chunk-complexity (with surprise) model, with a wAIC of 0.46. The second best was LoT-complexity (wAIC = 0.35; with surprise) and its LoT-chunk-complexity variant (wAIC = 0.014; with surprise). Entropy followed with a wAIC of 0.01, with or without surprise. All other models were weighted below 0.01.

Did the good score of chunk-complexity in deviant condition advocate against LoT? To test this, we added chunk-complexity as a fixed effect to the original LoT-complexity binomial Mixed-Effects Model. All fixed effects, including LoT- and chunk-complexity, were significant (LoT: $\beta = 0.29 \pm 0.07$; $p < 0.001$; chunk: $\beta = 0.31 \pm 0.07$; $p < 0.001$; surprise: $\beta = -0.47 \pm 0.08$; $p < 0.01$; group: $\beta = -0.36 \pm 0.07$; $p < 0.001$; index: $\beta - 0.012 \pm 0.05$; $p = 0.03$). The results were almost identical when replacing LoT predictor by its chunk variant (LoT-chunk: $\beta = 0.29 \pm 0.06$; $p < 0.001$; chunk: $\beta = 0.32 \pm 0.07$; $p < 0.001$; surprise: $\beta = -0.47 \pm 0.08$; $p < 0.01$;

group: $\beta = -0.35 \pm 0.07; p < 0.001$; index: $\beta - 0.012 \pm 0.05; p = 0.03$).

Replication of Planton et al. We then proceeded to compare our results to Planton et al.'s [Pla+21]. To match their study closer, we restrained the data to the trials from deviant condition where the outlier was the 9th, 11th, 13th, or 15th item. We also used LISAS instead of error as the performance measure, and removed fixed effects other than complexity and surprise. We then ran the same comparisons as above (Fig. S2B). LoT-chunk-complexity model (with surprise) had the highest weight (wAIC = 0.23), immediately followed by the original LoT variant (with surprise; wAIC = 0.20). The 5 other predictors, when combined with surprise, had lower but close weights, ranging from 0.08 to 0.02. Despite a smaller scale, the profile of the comparison roughly resembled Planton et al.'s [Pla+21] (Fig. S2A & B).

The results above highlight a degree of qualitative similarity (along with some differences) with Planton et al. [Pla+21]. This similarity was also quantitative, as the LISAS measures for each1 length-16 template were significantly correlated between the two studies (linear model over average LISAS for all templates: $R^2 = 0.635; \beta = 0.65; p = 0.01$; see Fig. 3.5).

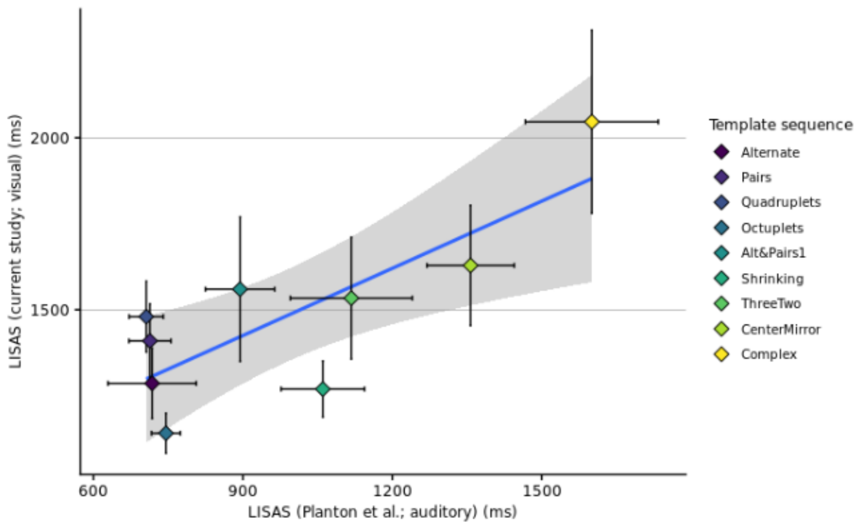

Fig. 3.5. Replication of Planton et al. with Exp. 1

3.3 Discussion

Our result show that participants do not process all binary patterns equally: some lead to significantly less errors than others. This difference is well-captured by the hypothesis of Language of Thought (LoT) processing patterns as combinations of primitives, with the complexity of said combination negatively affecting performance. Even though several competing measures were introduced, LoT-complexity remained one of the bests. It was only outmatched in deviancy cases by chunk-complexity, which measures how fragmented a pattern is.

Yet even when accounting for chunk complexity, LoT-complexity remained significant. It also remained significant under further comfounding factors, such as the surprise associated by the transition induced by the oulier, and the position of said outlier (first three positions or not). These results overall strongly argue for a LoT account of binary patterns processing in the visual modality, in line with the findings of Planton et al. [Pla+21] in the visual and auditory modalities using sequences. This convergence was highlighted by the correlation between our results and theirs.

The good scores of other predictors independently of LoT complexity also raise questions. Sequential predictors, such as surprise and outlier position, hint at a step-by-step processing of the sequence, and not of its whole structure. Since LoT-complexity did not account for how the template sequence was modified, we cannot conclude any further on how these findings relate to LoT processing. They could be signatures of an item-by-item comparison to a memorized template as a LoT-representation (with errors being related to memory load, i.e. complexity); but they could also belong to an entirely independent process happening in parallel.

Previous research in fact argues for the existence if processes independent from LoT to an extent: Sablé-Meyer et al. [Sab+21] found that baboons were insensitive to LoT-complexity differences, but could nonetheless achieve a related task. Remarkably, their strategy also predicted human behavior to a lesser extent[2]. Although the uniqueness of human LoT remains an open question, it is thus not far-fetched to assume that we still run ancestral processes for binary sequences in parallel to LoT. In fact, the tracking of transition probabilities and simple patterns have been observed in both humans and non-human primates [Deh+15].

Chunk complexity, precisely, may capture such alternative process. This predictor outmatched LoT-complexity in deviancy cases, and is tightly linked to interesting properties of the sequence. It could e.g. measure fragmentation (it is maximal for many small chunks and minimal for one single chunk). Very interestingly, chunk complexity only fared well in deviancy cases, and not when the two compared sequences were identical. It was also a poor predictor in Planton et al.'s [Pla+21] study. Given its limited advantage ($\Delta AIC < 10$) we cannot tell whether this result is due to noise or really highlights some concurrent strategy specific to our experimental design. In the latter case, this strategy could be specific to the visual modality or the use of patterns; but it could also be built on the constant length (16) of our sequences, while Planton et al.'s were of length 6 to 16. Either case, prediction would be greatly improved ($\Delta AIC = 12.44$; $wAIC > 0,99$; see Fig. S3) by considering the gap variant of chunk-complexity. This gap variant was computed as the absolute difference in complexity between target and query. Since the difference is null when target and query are matching, this improvement is consistent with the poor performance of chunk-complexity in matching cases. Given the focus of our experiment, we didn't push further the analysis of this predictor, although it could be a trail for future work.

[2] A key difference with our results, however, was that this processing wasn't observable anymore within adults educated in mathematics. This may be explained by the fact that a) their task involved geometrical quadrilatera and was thus more mathematics-driven than ours; and that b) the alternative strategy was much less efficient than in our case, which incentized resorting to LoT-based judgment.

Among possible alternative strategies, some degree of rote memorization is also a likely candidate. One could e.g. expect working memory to store the first items and their positions, with a limitation to the first 3 or 4 items due to memory constraints [MF12]. In agreement with this, we found that participants fared much better when the deviant was on the first 3 positions (error rate of 19% vs 41% otherwise on average). However, even when restricting ourselves to these outliers, the effect of LoT was still significant, albeit less decisive (R^2 of 0,41 instead of 0.62). This may suggest that LoT operates in parallel to working memory, and may marginally help improving its output. This weak effect of LoT despite a "basic", independent strategy is also observable in Planton et al. [Pla+21]. In their experiment, LoT-complexity predicted performance in detecting super-deviants (sounds absent from the original sequence).

Overall, our results in this experiment suggest that binary patterns processing are driven by a mechanism that can be analyzed as a "language of thought" (LoT) [Fod75]. This mechanism appears to be similar to the one observed by Planton et al. [Pla+21]. Notably, and unlike Planton et al.'s visual experiment, the similarity held across a different domain from temporal sequences: static patterns. Although the full extent will be more generally discussed later in chapter 5, this raises a major question: do we observe two "accidentally" similar processes, or is there rather a common, cross-domain mechanism underlying both? The possible cross-modal and cross-domain nature of LoT will be investigated in the following Experiment 2.

CHAPTER 4

Experiment 2: Testing priming from auditory sequences to visual patterns

4.1 Methods & Participants

Participants We recruited 10 participants (8 females; mean age: 38.7) from the RISC platform. Participants took the test on site, under our supervision, using a laptop and its keyboard. They were paid 10€ for 45 minutes on average. Due to our small sample size, we did not filter out any participant.

Procedure Participants went through a succession of pairs of blocks, alternating between an auditory and a visual task. Auditory trials were based on Planton et al.'s study [Pla+21]: first, we asked participants to listen and memorize sequences of auditory tones 8 times. If participants were confident that they had perfectly memorized the sequence, they could press the space bar to indicate it and move to the next block. Then, participants took 16 auditory test trials in which they had to spot outliers in the previously learned sequence. We asked participants to report an outlier by pressing the J key as soon as they detected one. If they did not spot any outlier, participants had to press F at the end of the sequence to move on to the next trial. This procedure was adapted from Exp 1, for consistency with visual trials. In fact, we also reused the very same procedure as Exp. 1 for visual trials. After each auditory/visual pair, participants were allowed to take a small break, and resumt at will.

Stimuli For this cross-modal experiment, auditory and visual trials were combined in a 5x5 design, using five length-16 template sequences from Planton et al. [Pla+21]. Each auditory blocks used a single template, hypothesized to induce priming for the following visual block, and visual trials used all five sequences to test this priming.

In auditory trials, we used two different, monotonous, 250ms tones: the first was at 425Hz and the second at 925Hz. Binary templates items were mapped onto these tones depending on their category, so that the mapping was consistent across the experiment. The mapping allowed conversion of templates into binary sequences, with an inter-item interval of 250ms.

To distinguish between them, sequences were separated with 1s empty periods. Outliers could only occur within the second half of each sequence, equiprobably at positions 10, 12, 14 or 16. This is similar to Planton et al., save for the set of positions which was offset by one item compared to their design. We thus used a close version of their experiment while making sure that the effect occurred on all positions, as predicted by our Language of Thought Hypothesis.

For visual trials, templates were converted into patterns by mapping each item category to a left-shifted or right-shifted position with respect to the central axis of the screen, as in Exp. 1. We nonetheless slightly deviated from stimuli from Exp. 1: matching trials could either display twice the exact template pattern (as in Exp. 1), or twice a deviated version of it. Every target and query had individually 50% chance to be deviated. We thus had a 2x2 balanced design, where outliers occurred 50% of the time, and every trial had at most one outlier. Every template was used 8 times in the visual part of each block. Those 8 occurences were distributed in 8 different groups of trials within the block, so that each group included every template once and only once. We then concatenated those group to form the visual part of the block, as we aimed to avoid immediate repetitions of a template across two consecutive trials. Since repetitions could still occur at the boundary of two blocks, we shuffled them when their first template was the same as the last one of the previous group.

Training was very important to match performance across blocks: instructions were followed by short practice blocks, and the experiment started with a fake test block which was ignored in our analyses. Training sequences and patterns were both simple and different from test ones : in auditory trials, we used length-8 patterns that were almost monotone (used only one of the two tones except for one or two items); in visual trials we kept the regular length for display consistency, but again used quasi-monotone templates as well as one difficult, handcrafted one to force participants to attend carefully.

4.2 Results

Participants took 3500 trials in total (auditory habituations were counted as 1 trial), among which 2850 test trials. We filtered out 11 trials on the basis of two criteria: for auditory trials, we removed trials where a lack of violation was reported before the end of sequence (n = 3); for visual trials, we removed answers over 3200ms reaction time, as in Exp. 1.

As a a common measure of performance for both types of trials, we focused on error rate. As Planton et al. [Pla+21], a deviant report was considered a correct only if it occurred between 200 and 2500ms after the deviant onset. Participants failed on average 23% of trials in the auditory modality and 28% in the visual modality. The difference between modalities wasn't significant (t-test pairing participant error rates: $t = -1.47; df = 9; p = 0.18$).

Auditory trials Auditory trials started with a habituation part, where a given sequence was repeated up to 8 times. Participants could trigger an early stop: all participants used this possibility at least at least once, and 6 out of 10 used it for every block. As a result, sequences were repeated 5.7 times on average. To measure this accurately, we computed encoding time (HT) across sequences as the duration between the start of the first repetition of the sequence and the pressing of space bar by participants. If participants did not press the space bar, we used the total duration of the trial (8 x 5,000ms = 40,000ms) as HT. HT was intended as a similar measure to Al Roumi et al.'s "encoding time" [Al +21].

We then analyzed HT as a function of sequence using an ANOVA across all participants. This revealed a significant difference between templates ($F = 6.21; df = 4,36; ges = .223; p < 0.001$; Fig. 4.1A). This effect was well-captured using LoT-complexity measure (linear model over average HT per template: $R^2 = 0.86; \beta = 632.5ms; p = 0.02$; Fig. 4.1B). Conversely, average HT did not correlate well with average order of the block ($R^2 = 0.24; \beta = -9,607ms; p = 0.40$).

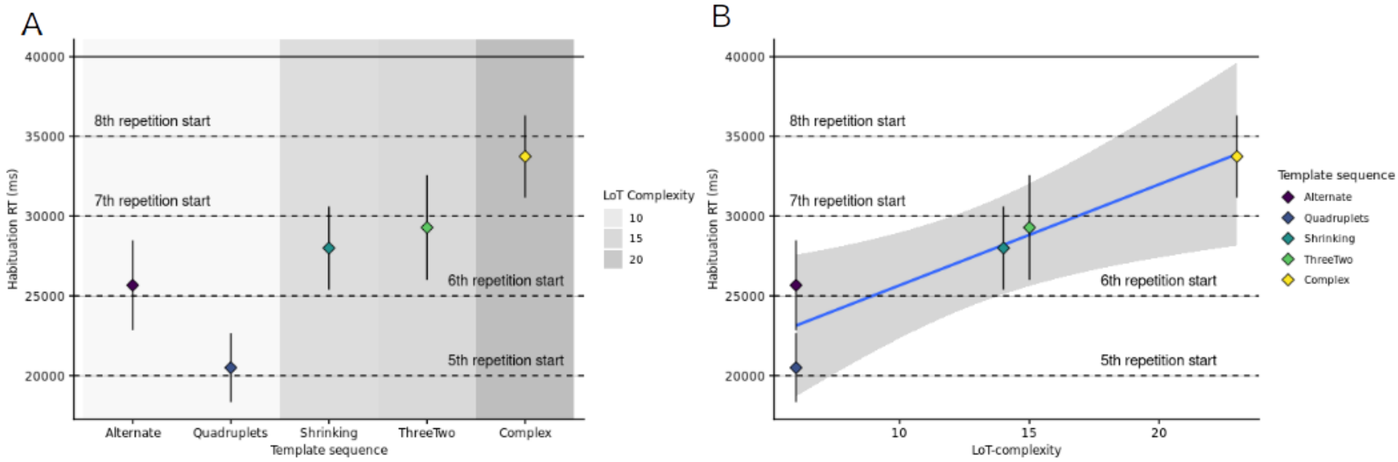

Fig. 4.1. A) Habituation time for each sequence. Error bars represent 1 standard error. Dashed lines represent repetitions starts, and plain line represent the forced end of habituation. B) Correlation of habituation time with LoT-complexity of the sequence.

Following habituation, participants took 16 test trials in which they had to spot outliers in the sequence they just learned. Detecting outliers was harder (33% errors on average by participants) than checking for their absence (14%, chi-squared test over all trials: $\chi^2 = 42.765; df = 1; p < 0.001$). We investigated whether sequences elicited different difficulties under either condition (deviancy or matching; see Fig. 4.2A), using ANOVAs across subjects for each template sequence. There was a significant difference for deviancy cases ($F = 7.78; df = 4,36; ges = 0.322; p < 0.001$), but not matching cases ($F = 1.12; df = 4,36; ges = 0.097; p = 0.36$). We completed this analysis with a linear regression of the average error rate across participant for every template sequence as a function of LoT-complexity. This model predicted very well error rates (see Fig. 4.2B) in both deviancy cases ($R^2 = 0.90; \beta = 2.3x10^{-2}; p = 0.01$) and matching cases ($R^2 = 0.89; B = 0.7x10^{-2}; p = 0.02$). To capture between-subject variability, we modeled single trial outcomes (if the answer was wrong) using a binomial Mixed-Effects Model. It included a random intercept per participants as a random effect, and LoT-complexity, deviancy condition and whether the habituation was stopped as fixed effects, with the later two being binary factors. As above, the

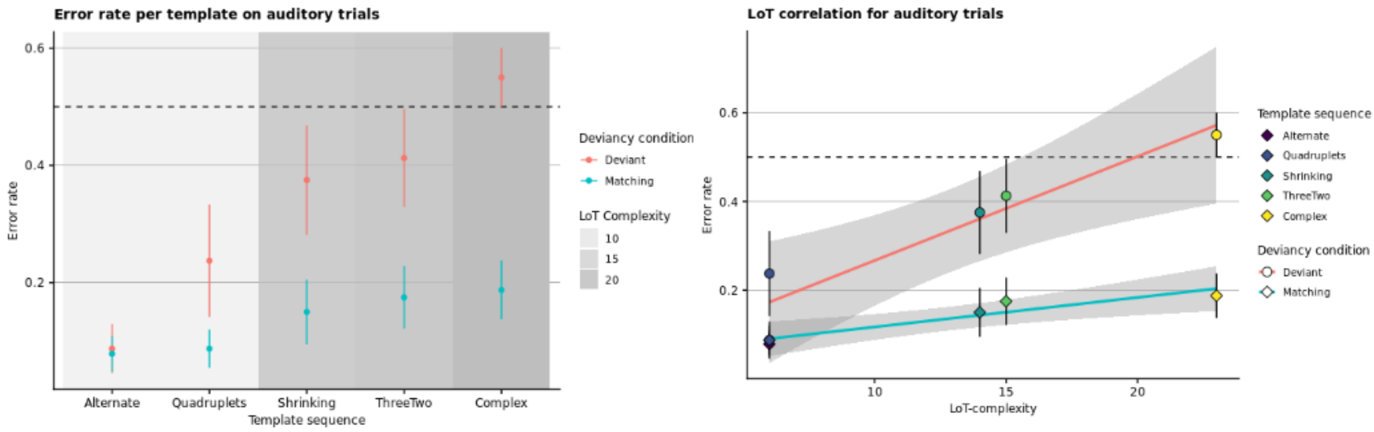

Fig. 4.2. A) Error rate on auditory trials depending on sequence and deviancy condition. Error bars represent 1 standard error. B) Correlation between performance and LoT-complexity for either condition on auditory trials.

LoT-complexity and deviancy condition came out significant (LoT: $\beta = 10.1x10^{-2}; p < 0.001$; deviancy: $\beta = 65x10^{-2}; p < 0.001$), but habituation stopping wasn't ($\beta = 40x10^{-2}; p = 0.32$).

To see if we replicated Planton et al.'s [Pla+21], we tried to see if the average LISAS measure for each sequence in our experiment could be predicted by the LISAS measure in the original study. Using a simple linear regression, we found a significant correlation between the two experiments ($R^2 = 0.88; \beta = 1.60ms; p = 0.02$). The intercept wasn't significatively different from zero ($\beta = 216ms; p = 0.60$).

Visual trials Visual trials presented one target pattern, followed by one query pattern, in which participants had to check for outlier. Either target, or query, or both could have one (which they shared if both had one). This resulted in a 2x2 local design, or pair condition, balanced over the block. We analyzed if this condition drove differences in response using chi-squared tests over single trials. The overall effect between pair conditions was highly significant ($\chi^2 = 86.922; df = 3; p < 0.001$). However, it was heavily driven by whether or not the two patterns were matching (deviancy condition; $X^2 = 85.74; df = 1; p < 0.001$; see Fig. 4.3A). Indeed, when the patterns were matching, there was no significant effect of whether they deviated from the template or not ($X^2 = 0.25186; df = 1; p = 0.62$). Similarly, there was no significant effect of which of target or query was deviated when the two differed (order condition; $X^2 = 0.8x10-30; df = 1; p \sim 1$).

Given that pair condition seems to reduce to deviancy condition, we ran the very same set of analyses on templates as for auditory trials. The ANOVAs revealed a significant difference in error rates between template patterns for deviant trials ($F = 19.17; df = 4,36; ges = 0.497; p < 0.001$), but not for matching ones ($F = 1.52; df = 4,36; p < 0.218$). In the former case, LoT-complexity was only correlated to error rate as a trend ($R^2 = 0.71; \beta = 2.0x10^{-2}; p = 0.07$), and wasn't significantly correlated at all in the latter case ($R^2 = 0.023; \beta = 0.07x10^{-2}; p = 0.80$; see Fig. 4.3B). When using a more sensitive binomial Mixed-effect Model over single trials, with participants as a random effect, the correlation became very significant for deviant cases ($\beta = 0.091; p < 0.001$). It still wasn't for matching cases ($\beta = 0.005; p = 0.71$).

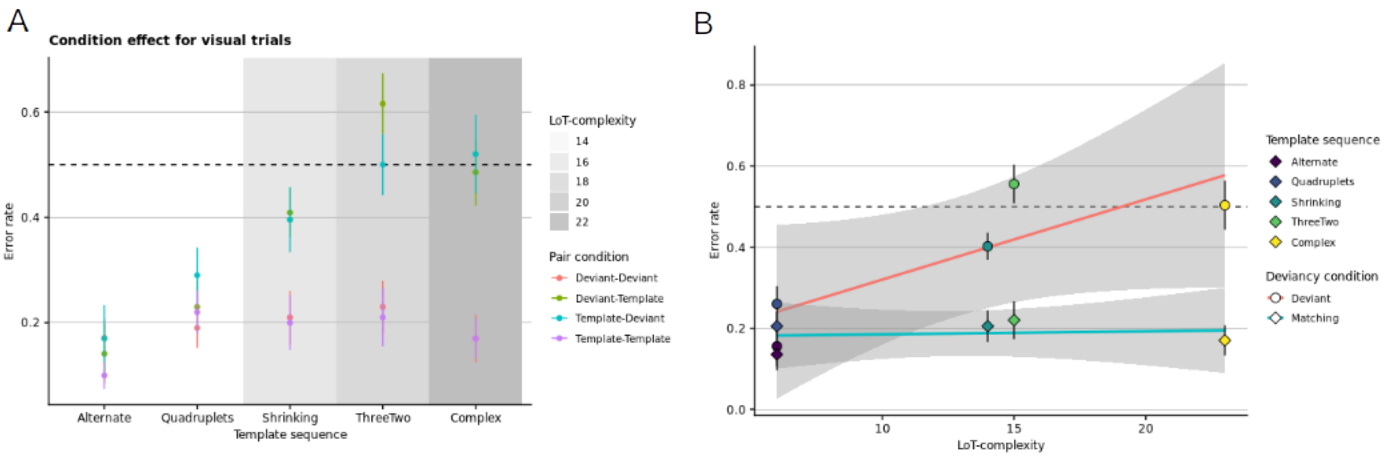

Fig. 4.3. A) Error rate depending on template and condition for visual trials in Experiment 2. Error bars stand for 1 standard error. Note that the profiles only differ along deviancy condition. B) Error rate depending on LoT-complexity in either deviancy condition.

As for auditory trials with Planton et al. [Pla+21], we checked whether our results replicated Exp. 1. Although cases where both target and sequences were deviated didn't occur in Exp. 1, we disregarded this difference given its lack of apparent effect (see above). With a simple linear regression, we checked whether the average error rate for each template pattern in this Exp. 2 could be predicted from those of Exp. 1. We found a strong correlation between the two experiments ($R^2 = 0.87; \beta = 1.31; p = 0.02$).

Priming effect Our aim was to investigate whether being habituated to a sequence in the auditory trials affected performance for the subsequent visual trials of the block. In particular, we wanted to test whether performance improved on the visual patterns sharing their templates with the habituated sequence. To that end, we computed the average error rate for each participant and each habituated-tested pair of sequences and patterns. We then computed the average error rate for participants when the templates underlying the pair were the same, and when they weren't. We analyzed these two paired distributions using a t-test. The difference wasn't found to be significant ($t = 1.37; df = 49; p = 0.18$; Fig. 4.4; also see Fig. S3 for graphs for each template). Without correction for multiple tests, one template ("Shrinking") had significantly better results in the primed block (paired t-test restricted to this template: $t = 2.445; df = 9; p = 0.037$).

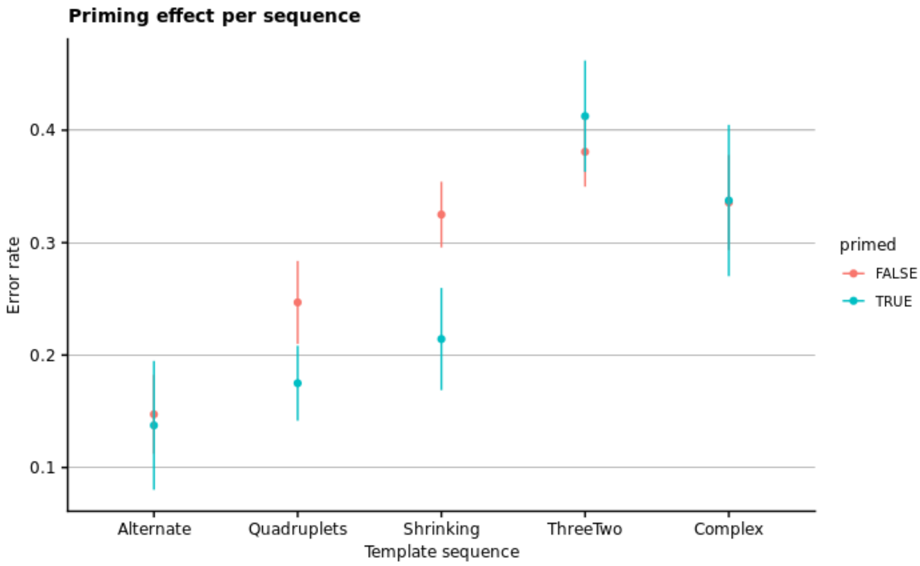

Fig. 4.4. Error rate on primed vs non-primed blocks per template. Error bars stand for 1 standard error.

Focusing back on habituation-tested pairs, we computed the average error rate across participants for each one of them (Fig. 4.5A). We aimed at predicting error rates using a combination of three different models. The first model was based on task complexity, and assumed error rate to simply be proportional to the LoT-complexity of the tested template, regardless of the habituation (Fig. 4.5B). The second, opposite model, was based on memory load, and assumed error rate to be proportional to the LoT-complexity of the habituation template (Fig. 4.5C). The last model was a model for priming, assuming error rate to be maximal for non-primed templates, and minimal for primed ones, irrespective of anything else (Fig. 4.5D). These models were used as predictors in a LM on error rate. Only task complexity was found a significant predictor ($R^2 = 0.48; taskcomplexity : \beta = 0.17; p < 0.001$; memory load: $\beta = 0.023; p = 0.57$; priming: $\beta = 0.032; p = 0.41$).

As a finer grained-analysis, we broke this prediction at the level of the participant (Fig. S4, aiming to predict their error rate on template pairs with our three models. We were mostly interested in a potential priming effect, and focused on the associated coefficient yielded by the regression for each participant. Using a one-sample t-test, we found that

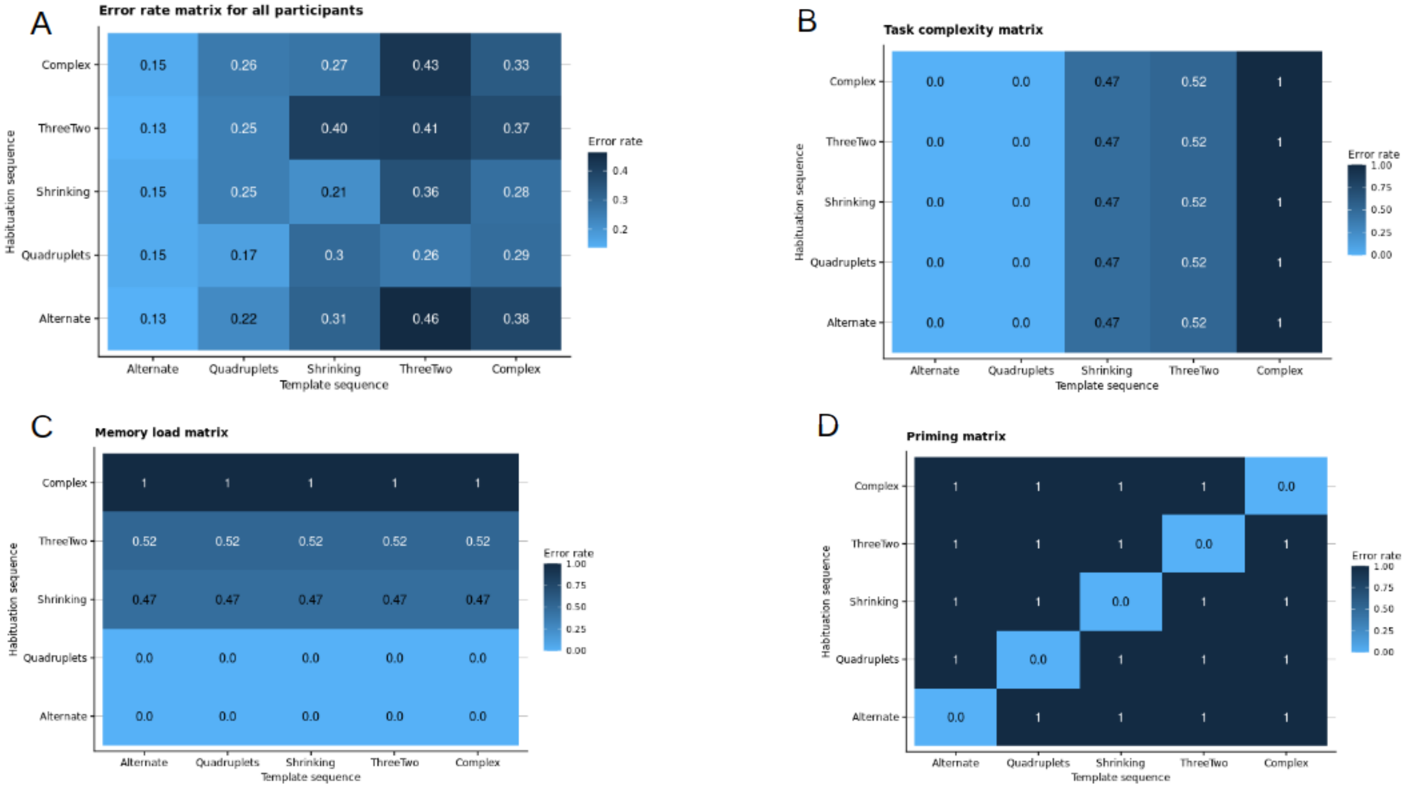

Fig. 4.5. A) Average error rates matrix for each habituation-visual template pair. Brighter colors indicate lower error rate. B) Expected error rates matrix for ideal participants, under the assumption that only task complexity affects error rate. C) As before for memory load. D) As before for priming.

the distribution of β-estimates was not significantly different from a zero-centered one ($t = 1.1068; M = 0.032; df = 9; p = 0.30$).

Training effect Aside from priming effect, which was defined as an effect from a habituated sequence on the patterns sharing the same template only within the same block, we considered a training effect spanning over the rest of the experiment following its habituation. We thus ran similar analyses as above, but this time grouping visual trials by position to their template habituation. We thus differentiated them based on whether they occurred in the same/a following block, or one before said habituation (see Fig. 4.6, and also Fig. S5 for full split along relative position to the block). As above, our first analysis was a paired t-test using difference in error rates for each participants for trials before/after prime. Given that the first (resp. last) primes did not have any trials before (resp. after) them, they were *de facto* excluded from the pair analysis. The t-test found a significant effect of occurring after the prime ($t = -3.4549; \Delta M = -0.11; df = 29; p = 0.002$).

We then regressed error rates again with our previous models, this time replacing priming with training. In training, error rates were supposed minimal for blocks including and after the habituation to the template, and maximal for those before. Since training model made different predictions for each participant (depending on the order of the block), we directly analyzed the distribution over participants of the estimates for the training model. We

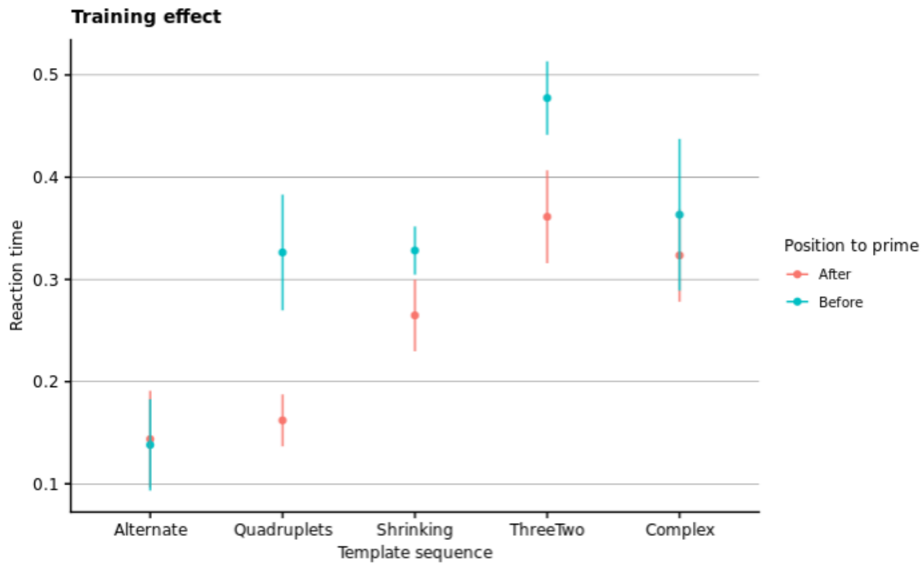

Fig. 4.6. Error rate before vs. after being habituated for each template. Error bars stand for 1 standard error across participants.

found no significant difference from a zero-centered one (one-sampled t-test: $t = 1.34; M = 0.039; df = 9; p = 0.21$).

To rule out a generic effect of task learning, we ran two binomial Mixed-Effects Models on single trials (to predict wrong answers). Both models had an intercept per participant as a random effect, and LoT-complexity, the binary deviancy condition, and the binary relative position to prime (before/after) as fixed effects. One of the models also included trial index as a further predictor and confounder to position to prime. In both cases, LoT-complexity (without trial index: $\beta = 0.06; p < 0.001$; with trial index: $\beta = 0.06; p < 0.001$) and deviancy condition (without trial index: $\beta = 0.51; p < 0.001$; with trial index: $\beta = 0.51; p < 0.001$) were highly significant. When trial index wasn't included, occurring after prime had a trending effect ($\beta = -0.10; p = 0.0578$). This effect was no longer significant at all ($\beta = 0.03; p = 0.67$) once trial index was included, and the effect was very significant for trial index instead ($\beta = -0.0027; p < 0.001$).

4.3 Discussion

Despite a low number of subjects (n=10), this experiment provides a good replication of both Planton et al.'s [Pla+21] experiment, and Exp. 1 of the present thesis. For both auditory sequences and visual patterns, we were able to accurately predict participant performance using the Language of Thought (LoT) Geo [Ama+17]. This result was not affected by the minimal changes we introduced. Participants could stop habituation to sequences early in the auditory part, and were forced to answer in the subsequent trials; outliers also occurred on different position than by Planton et al. In the visual part, the ordering of the trials was better controlled by avoiding immediate repetitions of a template. As such insensitivity was predicted by our LoT hypothesis (LoTH), the latter is left strengthened by these observations. Beyond solely confirming previous findings, it also demonstrated the viability of a common paradigm for the two domains investigated. This step is essential in the investigation of the interactions between them.

The purpose of this experiment was precisely to investigate a potential priming effect across these domains. It relied on the structural similarity of the templates used to generate their respective stimuli, adapting to a much simpler situation prior research on maths-language syntactic priming [SS14]. Unfortunately, we were not able to find a significant effect of the auditorily habituated sequence on visual patterns sharing the same template during the rest of the block. Of course, this may indicate that the priming effect we were hypothesizing simply isn't a thing. However, this lack of observed effect, could also derive from the experiment proper: first, we only had a sample of 10 participants. Although their data was of decent individual quality due to onsite testing (as highlighted by the low number of filtered trials), the priming effect size could be small, especially given our design.

In fact, we deviated from the typical priming experiment: primes did not occur right before trials, but instead were repeated before a group of trials (although some studies highlighted effect over several trials [BG00]). On top of that the groups were fairly long (40 trials), as we wanted a balanced design with enough data points for each of the 5 templates. The result is that we assumed the priming to last over two minutes (counting three seconds per trial)! This assumption was arguably bold, although it was intended to hide the priming purpose of the experiment, which the alternation of auditory and visual trials might have revealed. This part at least was well achieved, given that only one participant spontaneously reported that sequences and patterns were linked. When it was revealed, 8 of the other 9 participants reported not noticing it, while the last one was "unsure".

Just as we may have overestimated the duration of the priming, we could have underestimated it: priming may have carried on to the following blocks. Our design acted against such lasting phenomenon but only loosely: participants took small pauses between blocks, and each new auditory habituation was assumed to overwrite the previous one. We couldn't ensure that this was sufficient to prevent overlap between hypothesized primings. We decided to fully bite the bullet, and investigate the extreme version of this possibility: that priming never worn off across the experiment. In other words, we assumed auditory habituation on a template to train participant on this very template for the visual part. We then compared participant performance before and after this hypothesized training. We found that, although performance increased after a habituation part, this wasn't specific to the primed template; instead, trials involving each template were better solved. We thus associated this improvement with a better generic mastery of the task rather than a better knowledge of a given stimulus. No effect was observed after accounting for this, perhaps due to the lack of statistical power from our limited sample size.

As a result, none of our analyses hinted any kind of priming effect whatsoever. Nonetheless, many shortcomings of the experimental setup, and especially the 10-participant sample size, do not push for a blunt rejection of the hypothesis of a shared syntax across domains. Most notably, one template showed significant improvement upon priming, although this significance did not survive multiple-tests correction (see Fig. 4.4 and Fig. S3, template

"Shrinking").

Furthermore, the sequences where the priming effect was the least probable statistically were the most complex ones ("ThreeTwo", and "Complex"). As such, some sequences might be too complex for priming to occur properly, although this is only speculation at this time. This possibility is also disputable, since "ThreeTwo" has a barely higher complexity than "Shrinking" (15 vs 14; although this becomes 18 vs. 15 under Planton et al.'s [Pla+21] LoT-chunk variant), which witnessed the highest improvement. However, a sharp cut-off is not impossible either. The question of how LoT-complexity would affect priming thus remains unanswered.

CHAPTER 5

General discussion

During this internship, we aimed to explore the possibility of a "Language of Thought" (LoT) [Fod75] in the human brain. Prior studies proposed that such a mechanism would apply across a wide range of domains [VH16], including sequences of items [Pla+21]. This domain had already been shown to allow for cross-modal transfer learning [YJ15]; across more complex domains, such as mathematics and language, shared syntactic structures were found to trigger priming effects. We thus wanted to test the potential domain-generality of LoT. To better control for confounding effects, we tested syntactic interactions across the simple domains of binary sequences and binary patterns. To further disentangle the two, the domains were associated with different modalities.

Since binary sequences of sounds were already known to display LoT-complexity effects [Pla+21], Experiment 1 was designed to check whether these occurred in binary visual patterns as well. Using a same-different paradigm, we found support for a LoT account of their processing. Alongside, we also found evidence of simultaneously occurring processes, such as rote memorization. Overall, this experiment attested that our two domains displayed the effects a similar mechanism.

Experiment 2 was intended to test whether this similarity could be credited to a shared, syntactic mechanism for the two domains. Using a non-conventional priming design, we tried to see if habituating participants to an auditory sequence would improve their performance on patterns sharing the same template. Perhaps due to a low sample size (10 participants), this experiment wasn't successful, as performance did not significantly improve for primed patterns. This may be due to a lack of statistical power given the relatively small effect size that was expected [BG00]. Nonetheless, we observed an almost significant effect on one template, although it did not survive multiple-test correction.

Despite the lack of significance, we deem our observations to be encouraging overall. A larger-extent study could thus be envisioned beyond what was intended as a simple pilot. Still, improvements of the design might be considered to improve our chances of detecting an effect, and especially its reliance on a 2-minute long priming effect! To lessen the need for such a long priming, due to priming being tested over a single block, we could split this block into several others. As an example, our 40-visual-trial block, featuring 8 times each template, could be split into 8 mini-blocks featuring each template once. By keeping the exact same trials as the original block, but dispatching them over the mini-blocks, we would not change the whole set of visual trials that participants go through. However, shuffling all the mini-blocks, that would all start with the same auditory part as of now, we would only require priming to last for a handful of trials. This would also provide us with potentially more data on the auditory modality, which we could relate to the intensity of a putative priming effect.

Beyond our two tested domains, our findings could potentially decisively address the possibility of a domain-general LoT. The LoT Hypothesis has in fact been suffering from contradicting results between behavioral experiments, highlighting syntactic priming across mathematics and language [Sch+11; SS14], and brain imaging, which revealed dissociated networks for the two domains [AD19]. The existence of a shared mechanism beyond domains, i.e. semantics, relying on syntax, might reconcile the two: mathematics and language may (partly) share their syntax, but not their semantics. However, it must be noted that no brain region was identified as a good candidate for this shared mechanism in prior fMRI studies [AD19; Wan+19]. Assuming that we identify a syntactic interaction across domains, future work may thus focus on brain imagery to investigate which brain area could be shared across these domains. In particular, an experiment from Pallier et al. [PDD11] may offer an interesting lead. The authors investigated "jabberwocky" pseudo-sentences featuring constituents of various length as well as pseudo-function-words. This experiment could probably be adapted to include a mathematical part and provide a simple framework for maths and language syntactic comparison.

Although we were not did not succeed in supporting the hypothesis of a central, cross-domain-and-modality brain network dedicated to syntactic structures, we paved the way for it to be tested. Highlighting a common mechanism dedicated to syntax, or otherwise multiple, domain-specific ones, would improve greatly our understanding of how human language and mathematics evolved. It may also lead us to better account for our remarkable capacity for abstraction [Sab+21].

APPENDIX A

Statistics

A.0.1 Tools and softwares

Experiment were coded in `HTML 5` and `Javascript`, using `jsPsych` 6.3.1. Data analysis was performed with `R` 3.6.3, with the software `RStudio 1.1.463`. For ANOVAs and Mixed-Effects Model, `afex` package was used.

A.0.2 Statistical tests significance conventions

For all our statistical tests (ANOVAs, chi-squared, t-tests, correlation tests...), we will be refuting the null hypothesis when $p < 0.05$.

APPENDIX B

Supplementary figures

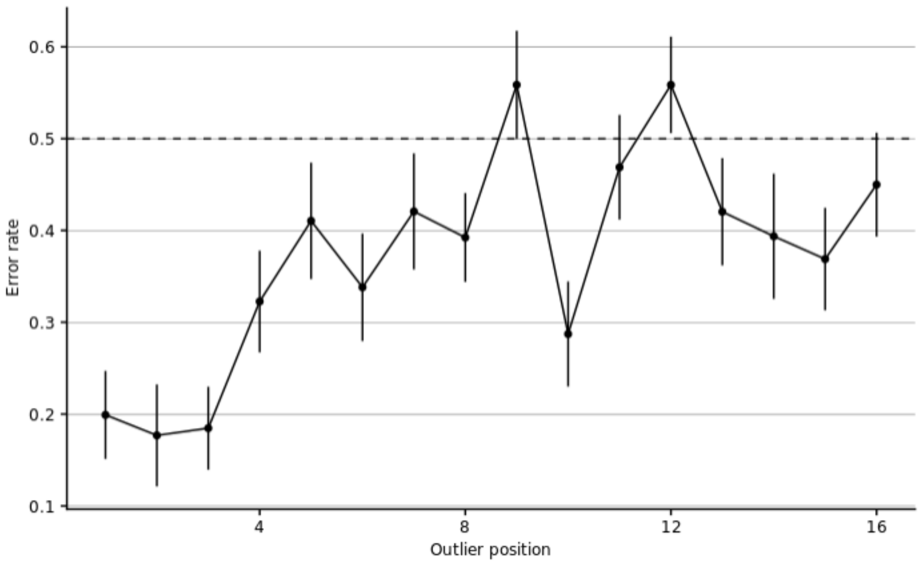

Fig. S1. Error rate by position of the outlier, irrespective of the sequence. Notice the plateau on the first 3 positions. Dashed line represent chance level (50%)

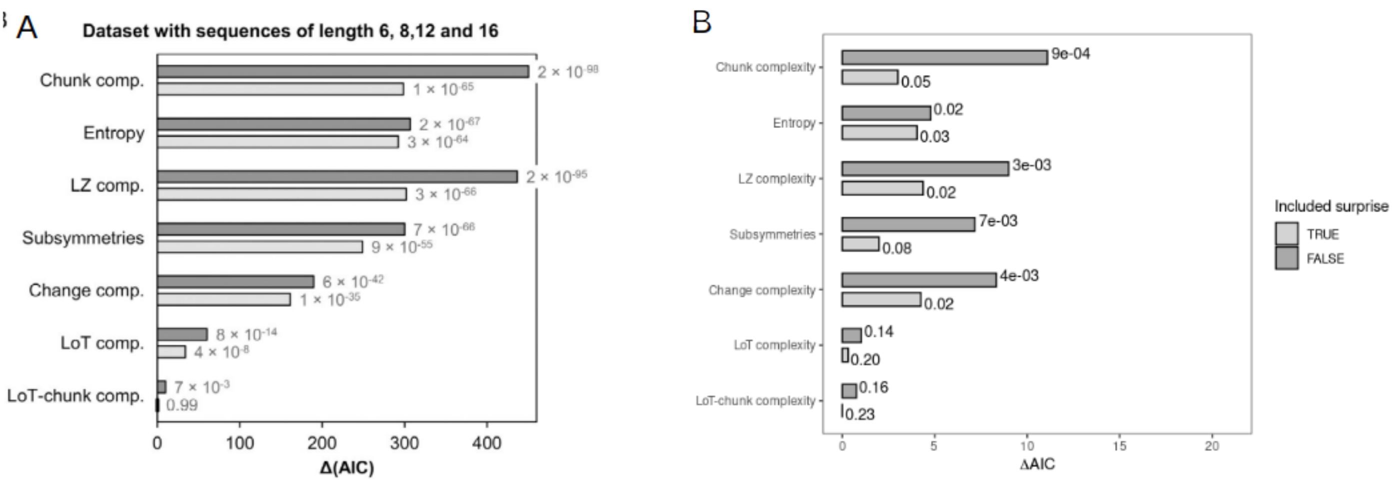

Fig. S2. Comparison on predictors across length-16 sequences. Models are compared to the best one, on the basis on fit to LISAS performance for every trials with an outlier on position 9, 11, 13, 15. Numbers at the tip of the bars represent respective Akaike weights. A) Results from Planton et al's [[Pla+21] experiment (auditory sequences). B) Results from present Exp. 1. (visual patterns).

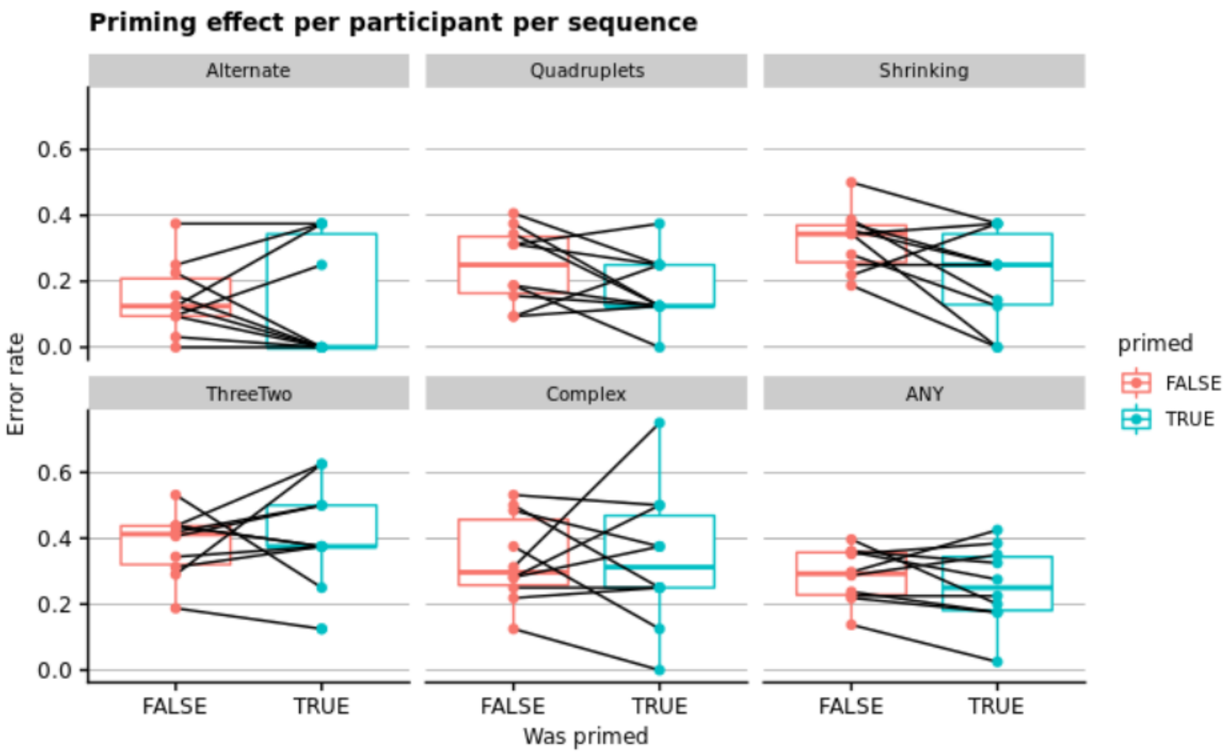

Fig. S3. Error rate of primed vs non-primed block for each template. Each point represents a participants. The boxplots represent median and quartiles. The black lines highlights whether error rates decreases or not during priming. The overall effect is displayed under "ANY".

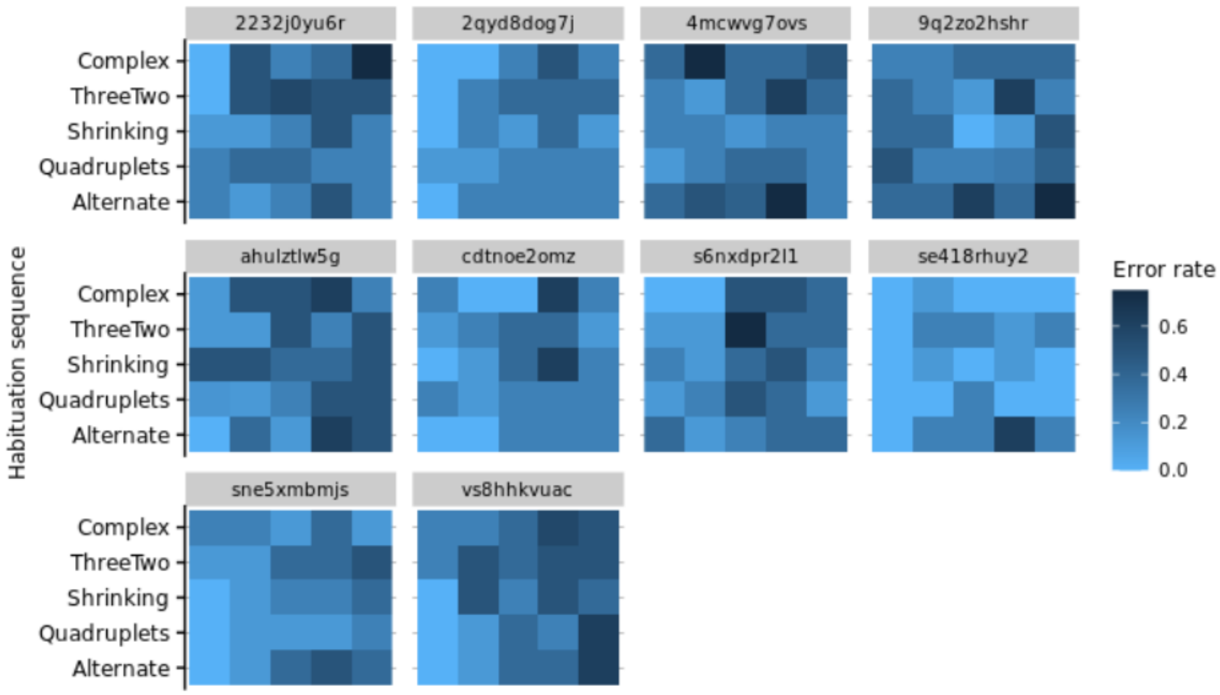

Fig. S4. Error rates matrices for each participants. Columns represent the template to which participants were habituated at the start of the block, and rows the template of the current trial. Lighter colors indicate lower error rate.

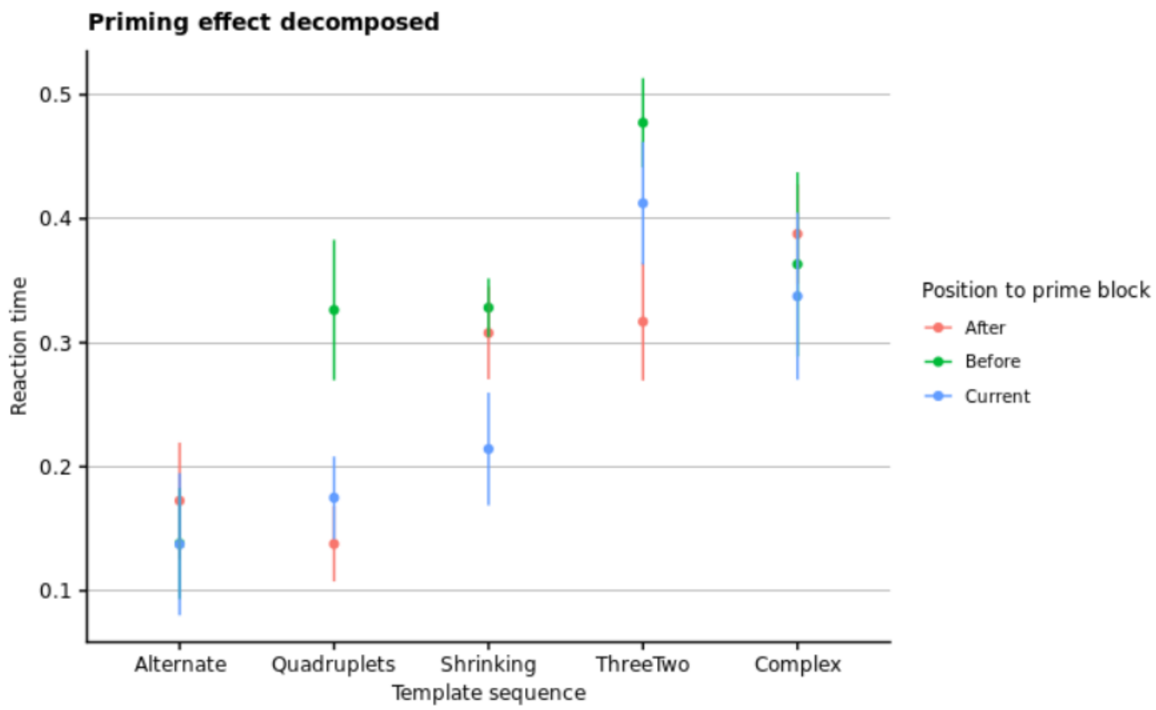

Fig. S5. Error rates for each template depending on position relative to primed block for this template.

APPENDIX C

Additional works

C.1 Codes for MEG-javascript interaction

In the context of a fellow intern's work on risk estimation, I developed a mini-python/javascript library to communicate with a MEG device. The idea was to adapt previous code for a MEG experiment in which the participant would have to make bets by pressing buttons. The challenge was that our response devices only communicated through parallel ports, while the experiment was coded in javascript: for safety reasons, javascript cannot read data from parallel ports.

We successfully adapted the experiment using a python pipeline, which retrieved the data using the expyriment library. The data was then sent to a server which was regularly queried by javascript. All the code can be found at `https://github.com/MaximeCaute/MEG_interactions`.

Given that neither my colleague nor I had a significant previous experience with MEG, this side project was both challenging and very enriching. I hope the acquired knowledge for good use for an eventual MEG experiment.

C.2 Compositionality in chimpanzees

As a term project for the course of super semantics at the CogMaster, I wrote an essay on a paper from Oña et al. [OSL19]. The authors observed responses of chimpanzees to combinations of faces and gestures by their conspecifics. They found out that a face in particular had two opposite effects on affiliativeness (how "friendly" an interaction is) depending on the associated gesture. This inconsistency led them to claim that face-gesture semantics in chimpanzees were non-compositional. Their conclusion may however be objected that their measure of semantics was indirect, and that compositional systems could similarly account for their data.

On a weekly basis, I thus worked with Emmanuel Chemla and Philippe Schlenker on a preparing a response to Oña et al. We also more generally started investigating the formal properties of animal semantic systems. This work is as of now heavily in progress, and will carry on over the next year.

List of Figures

Bibliography

[AD19] Marie Amalric and Stanislas Dehaene. "A distinct cortical network for mathematical knowledge in the human brain". en. In: *NeuroImage* 189 (Apr. 2019), pp. 19–31. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2019.01.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811919300011 (visited on 03/21/2022).

[Aka98] Hirotogu Akaike. "Information theory and an extension of the maximum likelihood principle". In: *Selected papers of hirotugu akaike*. Springer, 1998, pp. 199–213.

[Al +21] Fosca Al Roumi et al. "Mental compression of spatial sequences in human working memory using numerical and geometrical primitives". en. In: *Neuron* 109.16 (Aug. 2021), 2627–2639.e4. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2021.06.009. URL: https://www.sciencedirect.com/science/article/pii/S0896627321004244 (visited on 02/02/2022).

[Ama+17] Marie Amalric et al. "The language of geometry: Fast comprehension of geometrical primitives and rules in human adults and preschoolers". en. In: *PLOS Computational Biology* 13.1 (Jan. 2017). Publisher: Public Library of Science, e1005273. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1005273. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005273 (visited on 09/24/2021).

[Ber+11] Robert C. Berwick et al. "Songs to syntax: the linguistics of birdsong". en. In: *Trends in Cognitive Sciences* 15.3 (Mar. 2011), pp. 113–121. ISSN: 13646613. DOI: 10.1016/j.tics.2011.01.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S1364661311000039 (visited on 01/08/2022).

[BG00] Kathryn Bock and Zenzi M. Griffin. "The persistence of structural priming: Transient activation or implicit learning?" In: *Journal of Experimental Psychology: General* 129.2 (2000). Place: US Publisher: American Psychological Association, pp. 177–192. ISSN: 1939-2222. DOI: 10.1037/0096-3445.129.2.177.

[Boc86] J. Kathryn Bock. “Syntactic persistence in language production”. en. In: *Cognitive Psychology* 18.3 (July 1986), pp. 355–387. ISSN: 0010-0285. DOI: 10.1016/0010-0285(86)90004-6. URL: https://www.sciencedirect.com/science/article/pii/0010028586900046 (visited on 06/05/2022).

[Cäs+13] Cristiane Cäsar et al. “Titi monkey call sequences vary with predator location and type”. In: *Biology Letters* 9.5 (Oct. 2013). Publisher: Royal Society, p. 20130535. DOI: 10.1098/rsbl.2013.0535. URL: https://royalsocietypublishing.org/doi/10.1098/rsbl.2013.0535 (visited on 01/11/2022).

[Cho56] N. Chomsky. “Three models for the description of language”. In: *IRE Transactions on Information Theory* 2.3 (Sept. 1956). Conference Name: IRE Transactions on Information Theory, pp. 113–124. ISSN: 2168-2712. DOI: 10.1109/TIT.1956.1056813.

[Cho57] Noam Chomsky. *Syntactic structures.* English. OCLC: 308125. The Hague: Mouton, 1957. ISBN: 978-90-279-3385-0.

[Coy+15] Camille Coye et al. “Suffixation influences receivers’ behaviour in non-human primates”. In: *Proceedings of the Royal Society B: Biological Sciences* 282.1807 (May 2015). Publisher: Royal Society, p. 20150265. DOI: 10.1098/rspb.2015.0265. URL: https://royalsocietypublishing.org/doi/10.1098/rspb.2015.0265 (visited on 01/09/2022).

[Dav18] Paul Fisher Davies. “The Nested Spaces of Graphic Narrative”. en. In: *Between* 8.15 (July 2018). Number: 15. ISSN: 2039-6597. DOI: 10.13125/2039-6597/3248. URL: https://ojs.unica.it/index.php/between/article/view/3248 (visited on 06/05/2022).

[Deh+15] Stanislas Dehaene et al. “The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees”. en. In: *Neuron* 88.1 (Oct. 2015), pp. 2–19. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2015.09.019. URL: https://www.sciencedirect.com/science/article/pii/S089662731500776X (visited on 06/02/2022).

[Fer+20] Stephen Ferrigno et al. “Recursive sequence generation in monkeys, children, U.S. adults, and native Amazonians”. EN. In: *Science Advances* (June 2020). Publisher: American Association for the Advancement of Science. DOI: 10.1126/sciadv.aaz1002. URL: https://www.science.org/doi/abs/10.1126/sciadv.aaz1002 (visited on 01/11/2022).

[FF12] W. Tecumseh Fitch and Angela D. Friederici. “Artificial grammar learning meets formal language theory: an overview”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1598 (July 2012). Publisher: Royal Society, pp. 1933–1955. DOI: 10.1098/rstb.2012.0103. URL: https://royalsocietypublishing.org/doi/full/10.1098/rstb.2012.0103 (visited on 06/05/2022).

[Fia+07] Wim Fias et al. “Processing of Abstract Ordinal Knowledge in the Horizontal Segment of the Intraparietal Sulcus”. en. In: *Journal of Neuroscience* 27.33 (Aug. 2007). Publisher: Society for Neuroscience Section: Brief Communications, pp. 8952–8956. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.2076-07.2007.

URL: https://www.jneurosci.org/content/27/33/8952 (visited on 06/05/2022).

[Fit14] W. Tecumseh Fitch. "Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition". en. In: *Physics of Life Reviews* 11.3 (Sept. 2014), pp. 329–364. ISSN: 1571-0645. DOI: 10.1016/j.plrev.2014.04.005. URL: https://www.sciencedirect.com/science/article/pii/S157106451400058X (visited on 06/05/2022).

[Fod75] Jerry A. Fodor. *The Language of Thought*. en. Google-Books-ID: XZwGLBYLbg4C. Harvard University Press, 1975. ISBN: 978-0-674-51030-2.

[FP14] Anna Fiveash and Kristen Pammer. "Music and language: Do they draw on similar syntactic working memory resources?" en. In: *Psychology of Music* 42.2 (Mar. 2014). Publisher: SAGE Publications Ltd, pp. 190–209. ISSN: 0305-7356. DOI: 10.1177/0305735612463949. URL: https://doi.org/10.1177/0305735612463949 (visited on 10/19/2021).

[HCF02] Marc D. Hauser, Noam Chomsky, and W. Tecumseh Fitch. "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" EN. In: *Science* (Nov. 2002). Publisher: American Association for the Advancement of Science. DOI: 10.1126/science.298.5598.1569. URL: https://www.science.org/doi/abs/10.1126/science.298.5598.1569 (visited on 01/08/2022).

[HCN18] Ailsa Humphries, Zhe Chen, and Ewald Neumann. "Comparing Repetition Priming Effects in Words and Arithmetic Equations: Robust Priming Regardless of Color or Response Hand Change". In: *Frontiers in Psychology* 8 (2018), p. 2326. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2017.02326. URL: https://www.frontiersin.org/article/10.3389/fpsyg.2017.02326 (visited on 10/12/2021).

[HCW20] Ailsa Humphries, Zhe Chen, and Jonathan Wiltshire. "Repetition priming with no antipriming in picture identification". en. In: *Vision Research* 168 (Mar. 2020), pp. 9–17. ISSN: 0042-6989. DOI: 10.1016/j.visres.2019.09.011. URL: https://www.sciencedirect.com/science/article/pii/S0042698920300043 (visited on 10/08/2021).

[Jia+18] Xinjian Jiang et al. "Production of Supra-regular Spatial Sequences by Macaque Monkeys". en. In: *Current Biology* 28.12 (June 2018), 1851–1859.e4. ISSN: 0960-9822. DOI: 10.1016/j.cub.2018.04.047. URL: https://www.sciencedirect.com/science/article/pii/S0960982218305190 (visited on 06/05/2022).

[KM85] Walter Kintsch and Ernest F. Mross. "Context effects in word identification". In: *Journal of Memory and Language* 24.3 (1985). Place: Netherlands Publisher: Elsevier Science, pp. 336–349. ISSN: 1096-0821. DOI: 10.1016/0749-596X(85)90032-4.

[Koe+13] Stefan Koelsch et al. "Processing of hierarchical syntactic structure in music". In: *Proceedings of the National Academy of Sciences* 110.38 (Sept. 2013). Publisher: Proceedings of the National Academy of Sciences, pp. 15443–15448. DOI: 10.1073/pnas.1300272110. URL: https://www.pnas.org/doi/full/10.1073/pnas.1300272110 (visited on 06/05/2022).

[Las+51] Karl Spencer Lashley et al. *The problem of serial order in behavior*. Vol. 21. Bobbs-Merrill Oxford, United Kingdom, 1951.

[MDF20] Raphaëlle Malassis, Stanislas Dehaene, and Joël Fagot. "Baboons (Papio papio) Process a Context-Free but Not a Context-Sensitive Grammar". en. In: *Scientific Reports* 10.1 (Dec. 2020), p. 7381. ISSN: 2045-2322. DOI: 10.1038/s41598-020-64244-5. URL: http://www.nature.com/articles/s41598-020-64244-5 (visited on 01/11/2022).

[MF12] Fabien Mathy and Jacob Feldman. "What's magic about magic numbers? Chunking and data compression in short-term memory". en. In: *Cognition* 122.3 (Mar. 2012), pp. 346–362. ISSN: 0010-0277. DOI: 10.1016/j.cognition.2011.11.003. URL: https://www.sciencedirect.com/science/article/pii/S0010027711002733 (visited on 02/23/2022).

[MMD22] Maxime Maheu, Florent Meyniel, and Stanislas Dehaene. "Rational arbitration between statistics and rules in human sequence processing". en. In: *Nature Human Behaviour* (May 2022). Publisher: Nature Publishing Group, pp. 1–17. ISSN: 2397-3374. DOI: 10.1038/s41562-021-01259-6. URL: https://www.nature.com/articles/s41562-021-01259-6 (visited on 05/05/2022).

[NO18] Tomoya Nakai and Kazuo Okanoya. "Neural Evidence of Cross-domain Structural Interaction between Language and Arithmetic". en. In: *Scientific Reports* 8.1 (Dec. 2018), p. 12873. ISSN: 2045-2322. DOI: 10.1038/s41598-018-31279-8. URL: http://www.nature.com/articles/s41598-018-31279-8 (visited on 09/30/2021).

[OSL19] Linda Oña, Wendy Sandler, and Katja Liebal. "A stepping stone to compositionality in chimpanzee communication". In: *PeerJ* 7 (Sept. 2019), e7623. DOI: 10.7717/peerj.7623.

[PDD11] Christophe Pallier, Anne-Dominique Devauchelle, and Stanislas Dehaene. "Cortical representation of the constituent structure of sentences". en. In: *Proceedings of the National Academy of Sciences* 108.6 (Feb. 2011), pp. 2522–2527. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1018711108. URL: https://pnas.org/doi/full/10.1073/pnas.1018711108 (visited on 06/06/2022).

[PHS18] Céline Pozniak, Barbara Hemforth, and Christoph Scheepers. "Cross-Domain Priming From Mathematics to Relative-Clause Attachment: A Visual-World Study in French". In: *Frontiers in Psychology* 9 (2018), p. 2056. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2018.02056. URL: https://www.frontiersin.org/article/10.3389/fpsyg.2018.02056 (visited on 10/08/2021).

[Pia+04] Manuela Piazza et al. "Tuning Curves for Approximate Numerosity in the Human Intraparietal Sulcus". en. In: *Neuron* 44.3 (Oct. 2004), pp. 547–555. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2004.10.014. URL: https://www.sciencedirect.com/science/article/pii/S0896627304006786 (visited on 06/05/2022).

[Pia21] Steven T. Piantadosi. “The Computational Origin of Representation”. en. In: *Minds and Machines* 31.1 (Mar. 2021), pp. 1–58. ISSN: 0924-6495, 1572-8641. DOI: 10.1007/s11023-020-09540-9. URL: http://link.springer.com/10.1007/s11023-020-09540-9 (visited on 09/24/2021).

[Pla+21] Samuel Planton et al. “A theory of memory for binary sequences: Evidence for a mental compression algorithm in humans”. en. In: *PLOS Computational Biology* 17.1 (Jan. 2021). Publisher: Public Library of Science, e1008598. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1008598. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008598 (visited on 09/24/2021).

[PTG12] Steven T. Piantadosi, Joshua B. Tenenbaum, and Noah D. Goodman. “Bootstrapping in a language of thought: A formal model of numerical concept learning”. en. In: *Cognition* 123.2 (May 2012), pp. 199–217. ISSN: 00100277. DOI: 10.1016/j.cognition.2011.11.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0010027711002769 (visited on 06/05/2022).

[Rom+18] Sergio Romano et al. “Bayesian validation of grammar productions for the language of thought”. en. In: *PLOS ONE* 13.7 (July 2018). Publisher: Public Library of Science, e0200420. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0200420. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0200420 (visited on 12/17/2021).

[RSF13] Sergio Romano, Mariano Sigman, and Santiago Figueira. “LT^2C^2: A language of thought with Turing-computable Kolmogorov complexity”. en. In: *Papers in Physics* 5 (Mar. 2013). arXiv:1303.0875 [cs, q-bio], p. 050001. ISSN: 1852-4249. DOI: 10.4279/PIP.050001. URL: http://arxiv.org/abs/1303.0875 (visited on 06/04/2022).

[Sab+21] Mathias Sablé-Meyer et al. “Sensitivity to geometric shape regularity in humans and baboons: A putative signature of human singularity”. en. In: *Proceedings of the National Academy of Sciences* 118.16 (Apr. 2021), e2023123118. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2023123118. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.2023123118 (visited on 09/24/2021).

[Sch+11] Christoph Scheepers et al. “Structural Priming Across Cognitive Domains: From Simple Arithmetic to Relative-Clause Attachment”. en. In: *Psychological Science* 22.10 (Oct. 2011), pp. 1319–1326. ISSN: 0956-7976, 1467-9280. DOI: 10.1177/0956797611416997. URL: http://journals.sagepub.com/doi/10.1177/0956797611416997 (visited on 12/17/2021).

[Sha48] Claude Elwood Shannon. “A mathematical theory of communication”. In: *The Bell system technical journal* 27.3 (1948), pp. 379–423.

[Shi88] Stuart M. Shieber. “Evidence Against the Context-Freeness of Natural Language”. en. In: *Philosophy, Language, and Artificial Intelligence: Resources for Processing Natural Language*. Ed. by Jack Kulas, James H. Fetzer, and Terry L. Rankin. Studies in Cognitive Systems. Dordrecht: Springer Netherlands, 1988, pp. 79–89. ISBN: 978-94-009-2727-8. DOI: 10.1007/978-94-009-2727-8_4. URL: https://doi.org/10.1007/978-94-009-2727-8_4 (visited on 06/05/2022).

[SS14] Christoph Scheepers and Patrick Sturt. “Bidirectional Syntactic Priming across Cognitive Domains: From Arithmetic to Language and Back”. en. In: *Quarterly Journal of Experimental Psychology* 67.8 (Aug. 2014), pp. 1643–1654. ISSN: 1747-0218, 1747-0226. DOI: 10.1080/17470218.2013.873815. URL: http://journals.sagepub.com/doi/10.1080/17470218.2013.873815 (visited on 09/30/2021).

[Tra+21] N.-Han Tran et al. “Entropy trade-offs in artistic design: A case study of Tamil kolam”. en. In: *Evolutionary Human Sciences* 3 (2021). Publisher: Cambridge University Press. ISSN: 2513-843X. DOI: 10.1017/ehs.2021.14. URL: https://www.cambridge.org/core/journals/evolutionary-human-sciences/article/entropy-tradeoffs-in-artistic-design-a-case-study-of-tamil-kolam/C1EBAE83D2E57EC19A22F53E010339A0 (visited on 01/07/2022).

[Van17] André Vandierendonck. “A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure”. en. In: *Behavior Research Methods* 49.2 (Apr. 2017), pp. 653–673. ISSN: 1554-3528. DOI: 10.3758/s13428-016-0721-5. URL: https://doi.org/10.3758/s13428-016-0721-5 (visited on 01/11/2022).

[VH16] Joris Van de Cavey and Robert J. Hartsuiker. “Is there a domain-general cognitive structuring system? Evidence from structural priming across music, math, action descriptions, and language”. en. In: *Cognition* 146 (Jan. 2016), pp. 172–184. ISSN: 00100277. DOI: 10.1016/j.cognition.2015.09.013. URL: https://linkinghub.elsevier.com/retrieve/pii/S001002771530069X (visited on 10/12/2021).

[Wan+19] Liping Wang et al. “Representation of spatial sequences using nested rules in human prefrontal cortex”. en. In: *NeuroImage* 186 (Feb. 2019), pp. 245–255. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2018.10.061. URL: https://www.sciencedirect.com/science/article/pii/S1053811918320330 (visited on 06/06/2022).

[WF04] Eric-Jan Wagenmakers and Simon Farrell. “AIC model selection using Akaike weights”. In: *Psychonomic bulletin & review* 11.1 (2004), pp. 192–196.

[YJ15] Ilker Yildirim and Robert A. Jacobs. “Learning multisensory representations for auditory-visual transfer of sequence category knowledge: a probabilistic language of thought approach”. en. In: *Psychonomic Bulletin & Review* 22.3 (June 2015), pp. 673–686. ISSN: 1069-9384, 1531-5320. DOI: 10.3758/s13423-014-0734-y. URL: http://link.springer.com/10.3758/s13423-014-0734-y (visited on 05/31/2022).

[ZML18] Tao Zeng, Wen Mao, and Rongfeng Liu. “Structural priming from arithmetic to language in Chinese: Evidence from adults and children”. en. In: *Quarterly Journal of Experimental Psychology* 71.7 (July 2018), pp. 1552–1560. ISSN: 1747-0218, 1747-0226. DOI: 10.1080/17470218.2017.1340968. URL: http://journals.sagepub.com/doi/10.1080/17470218.2017.1340968 (visited on 10/12/2021).

[ZMZ21] Tao Zeng, Yating Mu, and Taoyan Zhu. “Structural priming from simple arithmetic to Chinese ambiguous structures: evidence from eye movement”. en. In: *Cognitive Processing* 22.2 (May 2021), pp. 185–207. ISSN: 1612-4790. DOI: 10.1007/s10339-020-01003-4. URL: https://doi.org/10.1007/s10339-020-01003-4 (visited on 10/08/2021).