HAL
open science

Non-adjacent Dependencies Processing in Human and Non-human Primates

Raphaëlle Malassis, Arnaud Rey, Joël R Fagot

To cite this version:

Raphaëlle Malassis, Arnaud Rey, Joël R Fagot. Non-adjacent Dependencies Processing in Human and Non-human Primates. Cognitive Science, 2018, 42 (5), pp.1677-1699. hal-01980246

HAL Id: hal-01980246

https://hal.science/hal-01980246v1

Submitted on 14 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Non-adjacent Dependencies Processing

in Human and Non-human Primates

Raphaëlle Malassis, Arnaud Rey and Joël Fagot

Laboratory of Cognitive Psychology
Institute for Language, Communication, and the Brain
Aix-Marseille University & CNRS
Marseille, France

Corresponding author: Joël Fagot, joel.fagot@univ-amu.fr

Laboratoire de Psychologie Cognitive – CNRS

Aix-Marseille Université

3, place Victor Hugo - Case D

13331 Marseille Cedex 03 – FRANCE

E-mail: joel.fagot@univ-provence.fr

Abstract

Human and non-human primates share the ability to extract adjacent dependencies and, under certain conditions, non-adjacent dependencies (i.e., predictive relationships between elements that are separated by one or several intervening elements in a sequence). In this study, we explore the online extraction dynamics of non-adjacent dependencies in humans and baboons using a serial reaction time task. Participants had to produce three-target sequences containing deterministic relationships between the first and last target locations. In Experiment 1, participants from the two species could extract these non-adjacent dependencies, but humans required less exposure than baboons. In Experiment 2, the data show for the first time in a non-human primate species the successful generalization of sequential non-adjacent dependencies over novel intervening items. These findings provide new evidence to further constrain current theories about the nature and the evolutionary origins of the learning mechanisms allowing the extraction of non-adjacent dependencies.

Keywords: language evolution, statistical learning, sequence learning, long-distance dependencies, animal cognition

Statistical learning can be defined as the implicit learning of regularities embedded in the environment, which has been proposed to play an important role in language acquisition (for a review, see Romberg & Saffran, 2010) and many other aspects of cognition (e.g., in visual perception, Fiser & Aslin, 2001). Several experimental paradigms have been used to study these fundamental learning mechanisms such as the Artificial Grammar Learning paradigm (AGL, Reber, 1967), the Artificial Language Learning paradigm (ALL, Saffran et al., 1996) or the Serial Response Time paradigm (SRT, Nissen & Bullemer, 1987). In these experiments, participants are typically exposed to sequences of nonsense stimuli (e.g., syllables, tones or visual shapes) that are organized with a specific grammar. Learning of the predictive relationships embedded in the input can then be assessed by presenting novel sequences that are either consistent or inconsistent with this grammar.

Comparative human/non-human studies can inform us about the nature and dynamic of these learning mechanisms, and their occurrence during the evolution, by distinguishing domain-general and evolutionary old processes from those that might have appeared more recently in the human lineage. Comparative experiments have demonstrated so far that animals, like humans, are highly proficient in learning predictive relationships between *adjacent* elements (i.e., elements that are presented one after the other in a sequence, without any delay or element in between; see for a review: Cate & Okanoya, 2012; Conway & Christiansen, 2001; Wilson, Marslen-Wilson, & Petkov, 2017). This ability has been reported in several primates (tamarins: Hauser, Newport, & Aslin, 2001; macaques: Wilson et al., 2013; Wilson, Smith, & Petkov, 2015; and baboons: Minier, Fagot, & Rey, 2016), and non-primate species (pigeons: Froehlich, Herbranson, Loper, Wood, & Shimp, 2004; rats: Toro &

Trobalón, 2005; and songbirds: Takahasi, Yamada, & Okanoya, 2010). Along with neuroimaging data (e.g., Wilson, Kikuchi, et al., 2015), these behavioral data have led to the hypothesis that the extraction of adjacent dependencies relies on evolutionarily conserved mechanisms (Friederici, 2004; Wilson et al., 2017).

In contrast, the evolutionary origins of the capacity to extract non-adjacent dependencies (i.e., predictive relationships between elements that are separated by one or several intervening elements in a sequence, hereafter "NADs") remain a matter of debate. A standard example of NADs in English is the relationship between auxiliaries and inflectional morphemes that are separated in the speech stream by the verbal root (e.g., *is* read*ing*). Crucially, extracting and recognizing these dependencies requires generalization over a variable verb. Studies conducted in humans indicate that NADs extraction is more challenging for humans than the extraction of regularities between adjacent elements (Cleeremans & McClelland, 1991; Gebhart, Newport, & Aslin, 2009; Newport & Aslin, 2004; Pacton, Sobaco, & Perruchet, 2015; Perruchet & Rey, 2005; Wilson, Smith, et al., 2015). This capacity develops later in human infancy (Gómez & Maye, 2005) than the sensitivity to adjacent dependencies (Saffran, Aslin, & Newport, 1996; but see Marchetto & Bonatti, 2013, for different findings). Moreover, the extraction of NADs looks especially challenging in human adults outside some specific facilitative contexts. For instance, it is facilitated when the non-adjacent elements have a high degree of perceptual (e.g., phonological) similarity (Creel, Newport, & Aslin, 2004; Gebhart et al., 2009; Onnis, Monaghan, Richmond, & Chater, 2005), when the intervening elements are highly variable (Gómez, 2002; Onnis, Monaghan, Christiansen, & Chater, 2005), or when the non-adjacent elements are located at

the edge of the sequences (Peña, Bonatti, Nespor, & Mehler, 2002). Studies also revealed similar performances and constraints on NADs learning in experiments using linguistic and nonlinguistic stimuli, such as tones (Creel et al., 2004; Endress, 2010; Gebhart et al., 2009), or actions (Endress & Wood, 2011), suggesting that the learning mechanisms involved in this ability are not language-specific. Several experiments aimed to test whether animals can also extract NADs. Wilson et al. (2015) exposed human participants and macaques to a complex grammar involving multiple adjacent and non-adjacent dependencies. All participants from both species detected violations of the adjacent dependencies. By contrast, half of the tested humans, and none of the monkeys were sensitive to the NADs. The authors concluded that when multiple regularities are present, monkeys rely preferentially on local ones, whereas humans exhibit more flexibility. Human learners can encounter some difficulties with the learning of several regularities simultaneously (e.g., Kovács & Mehler, 2009), but Wilson at al. (2015) suggests that this might be even more difficult for non-human learners. Animals' focus on various local cues rather than on NADs has also been reported when multiple embedded NADs must be processed, as in the so-called center-embedded grammars (e.g., Heijningen, Visser, Zuidema, & Cate, 2009). However, these results may point to a limitation in the monkeys' ability to track several NADs simultaneously, rather than an inability to extract NADs *per se.*

Complementary information on animals' ability to learn NADs has been obtained from experiments using simpler grammars of the form $AX^{(n)}B$. In these grammars, A and B are two paired elements, with a non-adjacent transitional probability of 1, and $X^{(n)}$ is one or several (n) variable intervening elements. Two main types of NAD have been investigated

thus far using these grammars: dependencies between elements that are perceptually more similar compared to the interspersed element(s) (i.e., that belong to the same perceptual category, hereafter "feature-based NAD"), and learning of NADs without such perceptual cues (often called "arbitrary associations").

Successful learning of feature-based NADs has been recently reported in two primate species, with visual (Sonnweber, Ravignani, & Fitch, 2015) and auditory stimuli (Ravignani, Sonnweber, Stobbe, & Fitch, 2013). Successful generalization across isomorphic visual and auditory sequences was also demonstrated recently (Ravignani & Sonnweber, 2017). In addition, positive results were obtained in rats, with NADs between elements that were phonologically similar (i.e., vowels vs. consonants, and *vice-versa*; de la Mora & Toro, 2013), whereas previous experiments conducted on NADs extraction in this species led to negative results in the absence of phonological similarity (Toro & Trobalón, 2005). Perceptual similarity therefore appears to facilitate the extraction of NADs in some non-human animal species, as it does in humans (e.g., Onnis et al., 2005). These findings support the hypothesis that NADs extraction is sustained by a general-purpose learning mechanism, interacting with general perceptual constraints (i.e., extra-linguistic, such as Gestalt principle of similarity, Creel et al., 2004), both being shared by phylogenetically distant species (Newport & Aslin, 2004; see however de la Mora & Toro, 2013, and Toro, Nespor, Mehler, & Bonatti, 2008, for a discussion of potential human-specific constraints on speech processing).

Evidence for successful NADs learning in the absence of such perceptual cues can also be found in non-human animals. To date, three distinct experiments were conducted on that topic. Newport, Hauser, Spaepen, and Aslin (2004) assessed the ability of tamarins (*Saguinus*

oedipus) to extract dependencies between non-adjacent syllables in a speech stream, using a familiarization procedure. Six 3-syllable nonsense words of the form AXB (three instantiations of A-B pairs and one of two X syllables inserted in the middle) were presented in a continuous stream during several minutes via a loudspeaker. After this familiarization phase, the tamarins' reactions to words (e.g., A_1XB_1) *versus* part-words (B_1A_2X) were tested. Tamarins exhibited a higher rate of orientation responses towards the speaker after a part-word than after a word, suggesting that they learned the NADs. The authors concluded that the tamarins accurately segmented this continuous stream into triplets of syllables based on these NADs. However, one limitation of this study is that the tamarins' ability to generalize these dependencies over novel intervening X syllables was not tested. Therefore, this experiment leaves unexplored the question of how the NADs were processed and stored in their memory. Indeed, they could have extracted the three A-B non-adjacent relationships, or instead memorized the six AXB "words" as wholes. The same limitation also applies to Milne et al. (2016) who used a passive listening method in rhesus macaques (*Macaca mulatta*) and the recording of event-related potentials. Despite the results suggesting the learning of NADs, this study failed to provide generalization tests to confirm that the learned NADs could be maintained over novel intervening events.

To our knowledge, Sonnweber et al. (2015) is the only study proposing this kind of test in a non-human species. These data were obtained in one subgroup of chimpanzees (*Pan troglodytes)* referred to as the "arbitrary associative dependencies" group. These chimpanzees were presented on each trial with two "strings" of visual shapes depicted concurrently on a screen in a two-alternative forced choice task. The task required to choose the strings of the

form AX^nB (with five instantiations of A-B pairs and 60 instantiations of X items, with $n = 1$ or 2) over XX^nX strings. What the participants precisely learned was evaluated by a series of tests proposed after the training phase. The two main generalization tests (Test 2: extension of the number of X elements, and Test 3: introduction of novel X shapes) were successfully completed by one participant over the three tested chimpanzees. However, the fact that the items were presented concurrently rather than sequentially raises potential issues. The processing of associations between spatially distant elements, especially at rather small distance (up to 12 cm from each other during training, with a maximum of 30 cm in Test 2, according to the reported pixel measures), can arguably rely on strategies different from those available for the extraction of sequential NADs. For instance, once the chimpanzees have learned that only the left- and right-most shapes of each string matter in this task, they can make quick eye-movements from one to the other and process them as if they were adjacent, despite the spatial distance. Extraction and recognition of sequential NADs, in contrast, may require to hold an element in working memory and relate it to another element occurring later in the sequence (Wilson et al., 2017). Therefore, Sonnweber et al. (2015) offers promising findings about non-human primates' ability to generalize NADs, but these researches would need to be extended to the case of sequential NADs. Another limitation of the comparative literature is that it contains no comparative data on the temporal dynamics of NADs learning in human and non-human animal species. Earlier, we reported that baboons and humans learned adjacent dependencies at the same speed (Minier et al., 2016; Rey, Minier, Malassis, Bogaerts, & Fagot, submitted). A remaining issue is whether this is also true for NADs.

The aim of the present study was twofold in this context. It was firstly to provide human and baboon comparative data on the fine-grained dynamics of NADs extraction. It was secondly to test whether baboons can generalize sequential NADs over novel intervening elements. In Experiment 1, baboons and humans performed a serial response time task (SRT, Nissen & Bullemer, 1987), requiring to follow a stimulus appearing sequentially at different locations on a touchscreen. In a subset of the presented sequences, the first target location predicted the last target location, while the second target location varied. Response times provided an online behavioral metric of NADs learning. In Experiment 2, generalization of the NADs over novel intervening locations was tested. During this test, novel consistent sequences were contrasted with novel sequences violating the dependencies. This violation method was inspired from Gómez (2002). Only baboons took part in this second part of the study, as generalization has already been demonstrated in humans (Endress & Bonatti, 2007; Frost & Monaghan, 2016; Marchetto & Bonatti, 2015).

Experiment 1

Method

Baboons

Participants and apparatus

Participants are 16 Guinea baboons (*Papio papio*, 11 females, age range 4-21 years) from the CNRS primate facility (Rousset-sur-Arc, France). They live within a larger group of 26 individuals, within a 700 m2 outdoor enclosure and had a permanent access to ten Automated Learning Devices for Monkeys (ALDM, for a detailed description, see Fagot &

Bonté, 2010; Fagot & Paleressompoulle, 2009) equipped with a 19-inch touch screen and a food dispenser. The main feature of ALDM equipment is that a radio frequency identification reader (RFID) identifies each baboon via a microchip implanted in each arm. The baboons can therefore participate to the research at will, without being captured, as the test programs recognize them automatically. All baboons had previously participated to numerous computerized experiments using the ALDM test systems, including experiments on sequence processing (Minier et al., 2016), but they have never been exposed to tasks requiring the learning of NADs. The experiment was controlled by the EPrime software (Version 2.0, Psychology Software Tools, Pittsburgh).

Procedure

Participants had to perform a serial response time (SRT) task which required following and touching a dot moving in a pre-defined order on a touch screen. Each trial began by the display of a fixation cross (120x120 pixels) presented at the bottom of the screen (see Fig. 1a). Touching this stimulus triggered the display of a matrix of 9 locations (3x3) which contained 8 white crosses (60x60 pixels each) and a red circle target (80x80 pixels). Participants had to touch the target. The target then disappeared and moved to the next location on the screen. There was no delay between the offset of the target and its next display. A trial consisted in a sequence of three targets. An accurate completion of the 3-target sequence delivered grains of dry wheat inside the ALDM test unit. An incorrect response (i.e., selection of an incorrect location) stopped the trial and triggered a 5-sec timeout without food reward. Trials in which the participants failed to select a stimulus within 5-sec after the target's appearance were aborted and presented again in the next trial. The time elapsed between the appearance of the target (i.e., the red circle) and the baboon's touch on this target

was recorded as the response time. Response times were recorded for each response in the 3-target sequence.

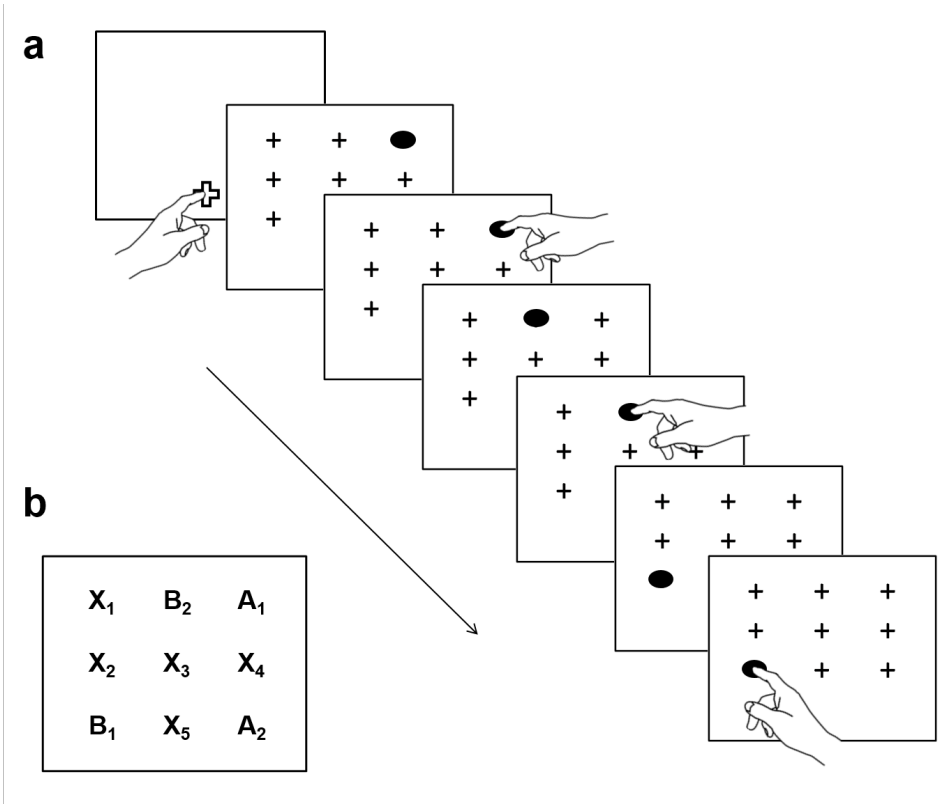

Figure 1: (a) General procedure of the serial response time task in Experiment 1. After touching the fixation cross, participants had to touch the red circle target that appeared successively at three different locations. (b) Locations on the touchscreen. Letters are given for illustrative purposes but were not displayed.

Two types of 3-target sequences were presented. In the first type ("NAD" condition), the first location of the target systematically predicted its last location, leading to a non-adjacent Transitional Probability (TP) of 1. Two pairs of A-B locations were used (A_1-B_1 and A_2-B_2), and a variable X location was inserted between A and B locations. The sequences were therefore of the form A_1XB_1 and A_2XB_2. The full set of 9 possible locations consisted in 4 locations for A_1, B_1, A_2 and B_2, and 5 locations for X elements (i.e., X_1, X_2, X_3, X_4, X_5; see

Fig. 1b). These locations were chosen pseudo-randomly among the nine possible spatial locations of the matrix, with the sole constraint that neither A_1 and B_1, nor A_2 and B_2, could be located next to each other. As the five X locations were presented at equal frequencies, the second location was uninformative about the third target location. The TP between the first and the second target locations was equal to 0.2, and the TP between the second and the third was equal to 0.5. A total of ten (two A-B pairs * five X locations) NAD sequences were used (see Table 1).

The second type of sequences had an XXB_1 and XXB_2 structure. These sequences were used in the control condition. To equate for the number of sequences per condition, 10 control sequences were constructed, as illustrated in Table 1. TPs between the first and third locations, as well as between the second and third locations, were equal to 0.5 in these sequences. It was therefore impossible to use adjacent or non-adjacent dependencies to predict the third target location. NAD and control sequences were presented at equal frequencies across the experiment. This equates motor practice in the two conditions. Experiment 1 involved a total of twenty 100-trial blocks. Each block contained five randomly intermixed presentations of the 10 NAD and 10 control sequences described above. All baboons were presented with the same sequences. Custom-written Python code (www.python.org) was used to generate the sequences.

Table 1 The 20 sequences presented in Experiment 1.

NAD	Control
A_1 X_1 B_1	X_3 X_1 B_1
A_1 X_2 B_1	X_5 X_2 B_1
A_1 X_3 B_1	X_2 X_3 B_1
A_1 X_4 B_1	X_1 X_4 B_1
A_1 X_5 B_1	X_4 X_5 B_1
A_2 X_1 B_2	X_3 X_1 B_2
A_2 X_2 B_2	X_5 X_2 B_2
A_2 X_3 B_2	X_2 X_3 B_2
A_2 X_4 B_2	X_1 X_4 B_2
A_2 X_5 B_2	X_4 X_5 B_2

Training

The baboons were familiar with the general principle of the task at the beginning of the experiment, as they previously performed a SRT task involving 9-target sequences (Minier et al., 2016). Before the experimental phase described above, the baboons received an initial training phase of twenty100-trial blocks in which the first, second and third target locations were selected randomly in the matrix. These random sequences were constructed with the constraint to systematically avoid the repetition of any given target location within the 3-target sequence. The aim of the training phase was twofold: (1) to familiarize the baboons with the 3-target sequences (instead of 9 as in Minier et al., 2016), and (2) to limit any long-term effect of the regularities previously learned.

Data analyses

Our set of statistical analyses compared response times on the third target (RTs) in the NAD and control conditions. Remember that the NAD and control sequences were presented at equal frequencies across the experiment. This procedure allowed us distinguishing response time accelerations due to a practice effect, as participants performed the control and NAD sequences an equal number of times, and those attributable to the learning of the predictive relationships present in the NAD sequences. Importantly, the final transitions between the second and the third target locations were the same across conditions (i.e., $X_{1\text{-}5}$-B_1 and $X_{1\text{-}5}$-B_2). Motor constraints were therefore controlled for that last transition. Shorter RTs in the NAD compared to the control condition would therefore indicate successful learning of the NADs.

For statistical analyses, the data of the exposure phase were grouped in five blocks of 400 trials. For each participant, mean response times on the third target (RTs) was computed for each condition and block. These data were then analyzed using repeated-measures ANOVA involving the Condition (NAD and Control) and Block (1-5) as within-participant factors. Post-hoc comparisons were performed using Tukey Honestly tests ($p < .05$)

On average, the sixteen baboons required five days of testing (range 3-7 days) to complete the experiment. Incorrect trials were removed from the data set (1.6%), as were removed all the trials with response times greater than two standard deviations from the mean (5.0% of the remaining trials, computed for each participant and each block).

Humans

Participants and apparatus

Ten human participants participated in this experiment (five females, age range 19-24). All participants were students of Aix-Marseille University and were paid for their participation. All participants were right-handed, had normal or corrected-to-normal vision, and were naïve as to the goal of the study. Participants performed the task while sited in front of a 19-inch touch screen. The experiment lasted approximately three hours per participant, and was divided in two sessions. Testing sessions were separated by two days at most.

Procedure and data analyses

The task and stimuli were similar in their general principles for humans and for baboons, with only slight differences between species. First, the number of blocks was reduced by half in humans (n = 10), in comparison to baboons (n = 20). This reduction was motivated by the long duration of the experiment (i.e., three hours with the current design). Second, humans did not receive a food reward at the end of each correct trial, contrary to baboons. The duration of the time-out and inter-trial interval were otherwise similar in the two species. Third, humans were instructed at the beginning of the experiment to follow the red target as quickly and accurately as possible. They were also told that a green screen meant that they did not correctly touch the target. Participants were interviewed at the end of the experiment to assess their declarative knowledge of the regularities embedded in the sequences. They were asked: “Did you notice any regularity, or rule, in some of the sequences?”.

Regarding the training phase, human participants were first trained on the SRT task using the same procedure as for baboons, but their number of training trials was reduced to

ten 100-trial blocks. Statistical analyses were the same as for baboons, except that blocks of the exposure phase were grouped in five blocks of 200 trials (instead of 400).

Incorrect trials were removed from the data set (7.3%), as well as trials with response times greater than two standard deviations from the mean (3.3% of the remaining trials, computed for each participant and each block).

Results

Baboons

Mean RTs obtained in baboons are reported for each condition and block in Fig. 2a. The repeated measures ANOVA revealed that the effect of Condition did not reach the standard significance level (i.e., .05), $F(1, 15) = 4.21$, $p = .06$, $\eta_p^2 = .22$. No main effect of Block was found[1], $F(1, 15) = 1.21$, $p = .29$, but there was a significant interaction between these two factors, $F(1, 15) = 13.73$, $p < .01$, $\eta_p^2 = .48$. Post-hoc analyses revealed that RTs were significantly shorter in NAD compared to the control trials for the last block (Block 5, Cohen's $d = .29$), but not for the previous blocks (1-4). Altogether, the findings suggest that baboons successfully learned the NADs in our task, but that this learning took a substantial number of trials.

[1] As noted by some reviewers, an increase in RTs for both conditions is observed between Blocks 3 and 4 (see Fig. 2a). We have no definitive explanation to this increase, which can be explained by a series of factors not directly related to the task, such as an interfering social event in the group of baboons, or a change in weather conditions.

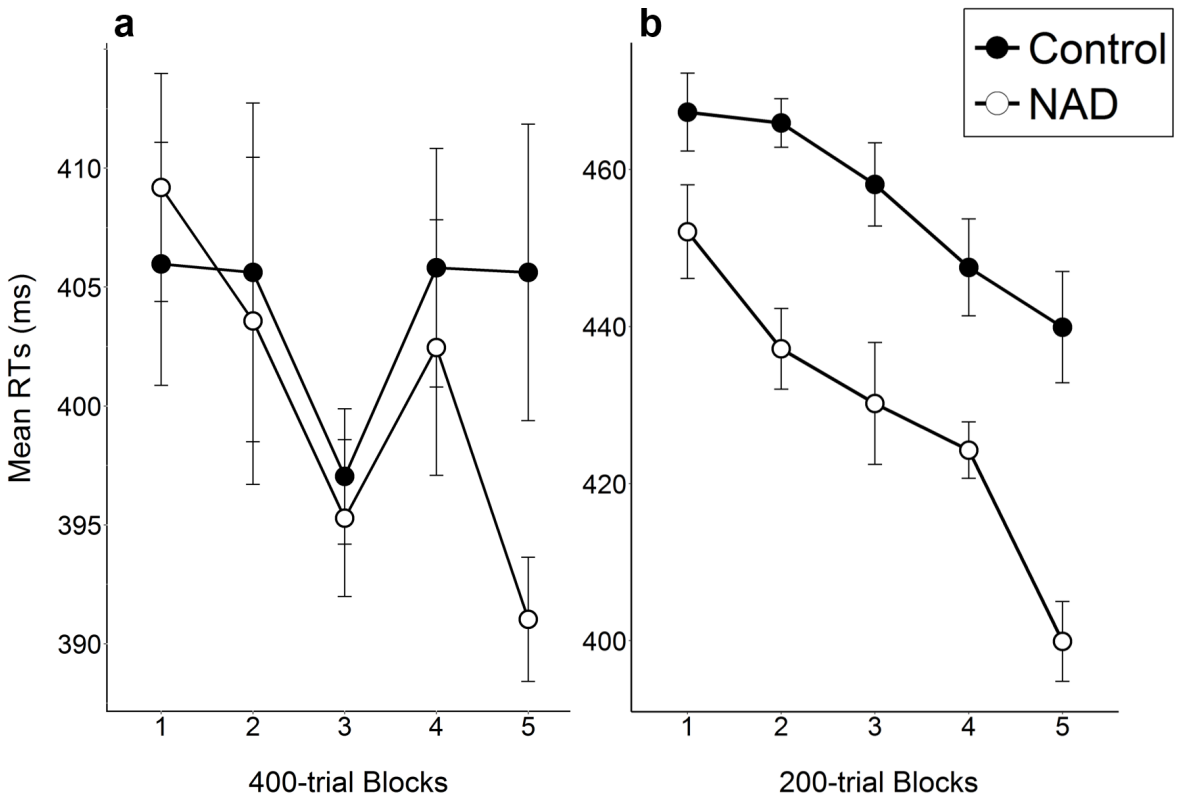

Figure 2. Response times for the third target (RTs) depending on the block, for the control and the non-adjacent dependencies (NAD) conditions in Experiment 1, for baboons (a) and for humans (b). The bars represent standard errors from normalized data.

Humans

Mean RTs are reported for each condition and block in Fig. 2b. Repeated measure ANOVAs revealed a main effect of Condition, F (1, 9) = 38.60, $p < .001$, $\eta_p^2 = .81$, indicating faster responses in the NAD (429 ± 65 ms) compared to the control condition (456 ± 58). A reliable main effect of Block was found as well, F (1, 9) = 30.99, $p < .001$, $\eta_p^2 = .77$, as was found a significant Block*Condition interaction, F (1, 9) = 11.87, $p < .01$, $\eta_p^2 = .57$. Post hoc

analyses on this interaction indicated a RTs advantage in the NAD condition compared to the control condition that was significant from the first block (*d*s = Block 1: .25, Block 2: .42, Block 3: .39, Block 4: .39, Block 5: .70). These results suggest that humans began to learn the NADs earlier than baboons, during the first block, and that this learning was amplified throughout the course of the experiment.

Post-experiment interviews revealed that none of the participants reported the existence of a predictive relationship between the first and the last locations (i.e., neither A_1-B_1 nor A_2-B_2 relationships were reported). However, 4 out of 10 participants correctly reported that two locations only were used at the last step of the sequence. In addition, four participants could explicitly verbalize some of the triplets (from one to four triplets, depending on the participant). Six of these triplets of locations were NAD sequences, and four were control sequences. To assess whether the results we obtained were triggered by the participants' declarative knowledge of this subset of sequences, we replicated the analysis described above after removing from the dataset the corresponding trials, and these exclusions did not change the pattern of statistical results. Overall, these data suggest that human participants could extract the NADs. They further suggest that while human participants exhibited learning of the regularities, as inferred from their response times, this learning was implicit.

Discussion

In Experiment 1, baboons and humans could extract the NADs, as both species responded faster to the third target when its location was predicted by the first target location

(NAD condition) than when it was not (control condition). Fig. 1 suggests that this difference progressively increased in both species, with substantial differences between them. A significant difference in RTs between the two conditions was found from the first block of exposure in humans. This effect only emerged after five blocks in baboons, revealing a substantial difference in the amount of exposure necessary for each species to exhibit successful learning of the NADs. Considering that each of the two NADs was presented 100 times within a 400-trial block, we estimate that baboons demonstrated learning after 400 (100 * 4 blocks) presentations of each NAD. In contrast, this advantage in response times reached significance since the first 200-trial block in humans, meaning that learning took place within the first 50 presentations of these NADs. Moreover, effect size of the difference between the two conditions on the last block was small to medium in baboons (d = .29), while it was medium to large in humans (d = .70).

However, a discrepancy in task practice between the two species might partly explain these differences between humans and baboons. Indeed, all our baboons had already participated to a study involving a SRT task (Minier et al., 2016), and performed 20 blocks of random sequences before completing the current experiment. In contrast, human participants only performed 10 random blocks and never had been trained to the SRT task before. This discrepancy in task practice between the two species might explain why humans' response times decreased progressively throughout the experiment in both the control and the NAD conditions, whereas baboons only showed this acceleration in the NAD condition. Importantly, the baboons responded faster on average than humans on the first block (Mean ± SD = 406 ± 43 and 460 ± 62 ms, respectively, t (24) = 2.58, p = .02, d = .99), probably as a consequence of their extensive practice with the task, while the two species show comparable

RTs on the last block (400 ± 54 and 420 ± 56 ms, respectively, t (24) = 0.89, p = .38). These data suggest that a ceiling effect might have limited baboon's acceleration in the NAD condition.

One limitation of this first experiment is that it does not rule out the possibility that the participants progressively learned the NAD sequences by rote, or a subset of them, and responded faster at the third location because of this rote learning, but without extracting the NADs embedded in these sequences. Experiment 2 addresses this hypothesis in baboons by testing their ability to generalize the NADs over novel intervening locations.

Experiment 2

Experiment 2 tested whether baboons merely learned the ten A_1XB_1 and A_2XB_2 sequences as wholes, or extracted instead the A_1_B_1 and A_2_B_2 non-adjacent relationships. Baboons were exposed successively to two different "languages"[2] in a within-participant design, and two experimental phases were proposed for each language. The first phase corresponded to the exposure phase of Experiment 1, except that the set of possible locations was increased from 9 to 16 locations. Six of these locations remained unused during that phase. During the second phase, which is referred to as the test phase, the baboons were presented with two types of novel sequences. In half of these sequences, the NADs learnt during the exposure phase were presented over the six locations not used during the exposure

[2] We borrow here the term "language" to the literature on artificial language learning, the set of sequences being considered here as a set of "words".

phase. The dependencies were inconsistent in the remaining half of the sequences (i.e., A_1 was presented with B_2 and A_2 with B_1). Importantly, the consistent sequences in Language 1 were the inconsistent sequences in Language 2, and *vice-versa*. This two-language design therefore guaranties that any difference in response times between the conditions will reflect the learning and generalization of the NADs, rather than idiosyncratic sequence features related to differences in motor constraints.

Methods

Participants and apparatus

Experiment 2 involved 14 (10 females, age range 4-21 years) out of 16 baboons of Experiment 1, because we were unable to test two of the previous participants for reasons unrelated to our research. The apparatus was the same as in the previous experiment. The general testing procedure was also the same as in Experiment 1, except that the display contained 16 (4x4) possible locations rather than only 9 (3x3).

Procedure

Table 2 provides the full lists of sequences used in Languages 1 and 2. The first exposure phase was identical to Experiment 1 in both languages. In Language 1, sequences took the form A_1XB_1 and A_2XB_2 for the NAD trials (50% of the trials), and XXB_1 and XXB_2 for the control trials (50% of the trials; see Table 2), with six different instantiations of the X locations. In Language 2, the sequences took the form A_1XB_2 and A_2XB_1 for the NAD trials, and control sequences were the same as in Language 1. The remaining six locations (noted $Y_{1\text{-}6}$) were never used during the exposure phase. Locations have been chosen pseudo-

randomly, with the sole constraint that A_1 and B_1, as well as A_2 and B_2, could not be located next to each other in the matrix. Fig. 3 shows the sixteen locations on the touchscreen that were used in Experiment 2. Note that the locations were the same in both languages, and that only the NADs differed between them.

Baboons then received three types of sequences during the test phase. The first type (50% of the trials) corresponded to the NAD sequences of the exposure phase. For the two other types of trials, sequences were either consistent with the NADs of the exposure phase ("Consistent" condition, 25% of the test trials), or inconsistent with those dependencies ("Inconsistent" condition, 25%). In these two latter conditions, the second location of the sequence was drawn from the set of Y locations never used during the exposure phase. Therefore, sequences were novel in these two conditions. In Language 1, consistent sequences took the form A_1YB_1 and A_2YB_2, while inconsistent sequences took the form A_1YB_2 and A_2YB_1. In Language 2, the consistent sequences were the inconsistent sequences of Language 1 (i.e., A_1YB_2 and A_2YB_1), while the inconsistent sequences were the consistent ones of Language 1 (i.e., A_1YB_1 and A_2YB_2). Twelve sequences were constructed for each condition. NAD, consistent and inconsistent trials were randomly intermixed within each block of the test phase.

Each baboon participated successively to the two language sessions. For each language, the baboons received 40 blocks of 96 trials during the exposure phase, and 20 blocks of 96 trials in the test phase. Note that the amount of exposure changed from 20 blocks in Experiment 1 to 40 in Experiment 2, for both languages, because we hypothesized that learning was still in progress during the last blocks of Experiment 1. To limit potential

influences of Language 1 acquisition on the learning of Language 2, six months elapsed between the two language sessions, during which the baboons were exposed to different unrelated tasks. Furthermore, a random phase consisting of 35 blocks of 96 trial blocks was given to the baboons at the beginning of the Language 2 session.

X_1	X_2	A_1	Y_1
B_1	Y_2	X_3	Y_3
Y_4	X_4	Y_5	Y_6
A_2	X_5	B_2	X_6

Figure 3. Locations on the touchscreen in Experiment 2. Letters are given for illustrative purposes but were not displayed. Note that the locations were the same for Languages 1 and 2.

429

Table 2. The 48 sequences presented during Experiment 2, for Language 1 (L1) and Language 2 (L2).

NAD (L1)	NAD (L2)	Control (L1 & L2)	Consistent (L1) / Inconsistent (L2)	Inconsistent (L1) / Consistent (L2)
A_1 X_1 B_1	A_2 X_1 B_1	X_2 X_1 B_1	A_1 Y_1 B_1	A_2 Y_1 B_1
A_1 X_2 B_1	A_2 X_2 B_1	X_3 X_2 B_1	A_1 Y_2 B_1	A_2 Y_2 B_1
A_1 X_3 B_1	A_2 X_3 B_1	X_6 X_3 B_1	A_1 Y_3 B_1	A_2 Y_3 B_1
A_1 X_4 B_1	A_2 X_4 B_1	X_1 X_4 B_1	A_1 Y_4 B_1	A_2 Y_4 B_1
A_1 X_5 B_1	A_2 X_5 B_1	X_4 X_5 B_1	A_1 Y_5 B_1	A_2 Y_5 B_1
A_1 X_6 B_1	A_2 X_6 B_1	X_5 X_6 B_1	A_1 Y_6 B_1	A_2 Y_6 B_1
A_2 X_1 B_2	A_1 X_1 B_2	X_2 X_1 B_2	A_2 Y_1 B_2	A_1 Y_1 B_2
A_2 X_2 B_2	A_1 X_2 B_2	X_3 X_2 B_2	A_2 Y_2 B_2	A_1 Y_2 B_2
A_2 X_3 B_2	A_1 X_3 B_2	X_6 X_3 B_2	A_2 Y_3 B_2	A_1 Y_3 B_2
A_2 X_4 B_2	A_1 X_4 B_2	X_1 X_4 B_2	A_2 Y_4 B_2	A_1 Y_4 B_2
A_2 X_5 B_2	A_1 X_5 B_2	X_4 X_5 B_2	A_2 Y_5 B_2	A_1 Y_5 B_2
A_2 X_6 B_2	A_1 X_6 B_2	X_5 X_6 B_2	A_2 Y_6 B_2	A_1 Y_6 B_2

430 ***Data analyses***

431 Two separated sets of analyses were conducted on the exposure and test phases.
432 Blocks were grouped for these analyses in 10 (Exposure) and 5 (Test) blocks of 384 trials
433 each. Learning of NADs during the exposure phase was evaluated with repeated-measures
434 ANOVA performed on RTs and involving the Condition (NAD, Control), Language (1, 2)
435 and Block (1-10) as within-participant factors. Post-hoc comparisons were performed using
436 Tukey Honestly tests ($p < .05$).

437 Generalization was evaluated in the test phase with repeated-measures ANOVA on
438 RTs involving the Condition (NAD, Consistent, Inconsistent), Language (1, 2) and Block (1-

5) as within-participant factors. RTs in the consistent and inconsistent conditions were further compared for the first test block using repeated-measures ANOVA involving the Condition (Consistent, Inconsistent) and Language (1, 2) as within-participant factors. RTs in the consistent and inconsistent conditions were further compared for the first test block using a repeated-measures ANOVA involving the Condition (Consistent, Inconsistent) and Language (Language 1, Language 2).

We also investigated potential effects of spatial factors on the generalization of NADs. This analysis was conducted with an ANOVA using the Condition (Consistent, Inconsistent) and Pair (A_1-B_1, A_2-B_2, A_1-B_2, A_2-B_1) as within-participant factors, and the RTs obtained in the first test block as dependent variable. An additional set of analyses tested whether generalization occurred for every novel Y location. One-tailed paired t-tests ($p < .05$) were used to assess if RTs were reliably longer in the inconsistent compared to the consistent condition for each Y location.

To complete Language 1 session, the fourteen baboons required 8 days (range 5-12) on average. The same baboons required on average 10 days (range 5-17) of testing to complete Language 2 session. Incorrect trials were removed from the data set (Language 1: 1.7%, Language 2: 2.1%), as well as trials with RTs greater than two standard deviations from the mean (4.9 % and 5.1%, respectively, computed for each participant and each block).

Main Results

Mean RTs are reported in Fig. 4 for each language, phase and condition. For the exposure phase, the main effect of Condition was significant, $F(1, 13) = 16.28$, $p = .001$, $\eta_p^2 =$

.56, corresponding to faster response times in the NAD (Mean ± SD = 422 ± 44 ms) than in the control condition (430 ± 43 ms). Also significant was the main effect of Block, F (1, 13) = 16.88, p = .001, η_p^2 = .56, showing that RTs decreased with practice. Finally, the main effect of Language was also significant, F (1, 13) = 6.60, p < .01, η_p^2 = .34, indicating faster response times on average for the second (414 ± 32 ms) compared to the first language (438 ± 50 ms). The Condition*Block interaction was close to significance level, F (1, 13) = 3.21, p = .10, η_p^2 = .20. There were no other interactions (all *ps* > .42). The main effect of condition, along with the absence of Condition*Language interaction, indicate that the baboons were able to learn the NADs in both languages.

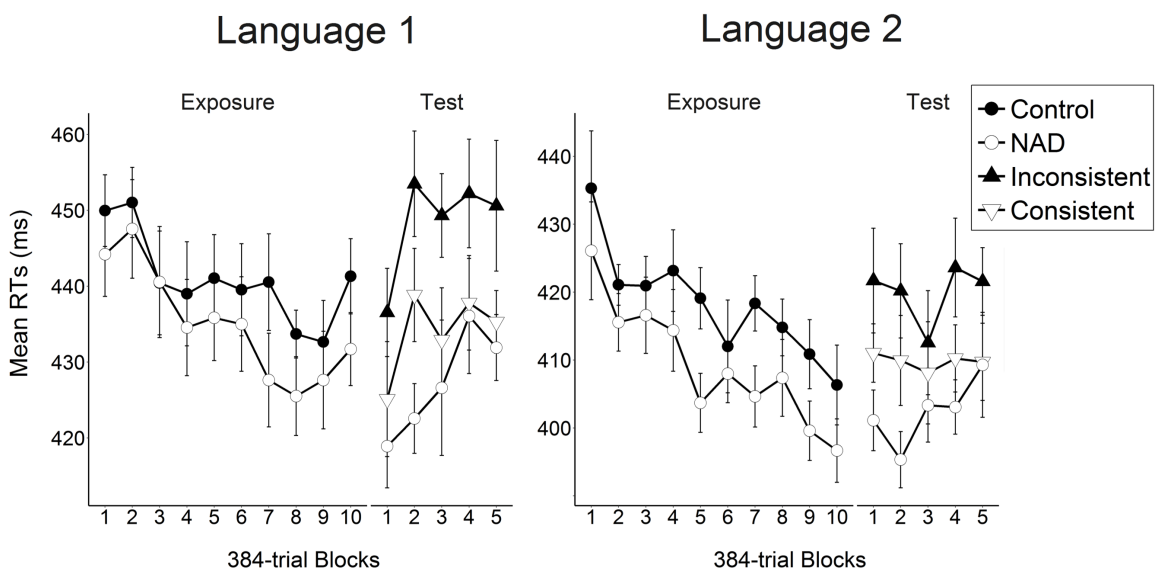

Figure 4. Mean response times (RTs) on the third target obtained in each condition of the exposure and test phases of Experiment 2 and for each language. Bars represent standard errors from normalized data.

Analyses conducted on the test phase revealed a significant effect of Condition, F (2, 26) = 13.39, $p < .001$, $\eta_p^2 = .51$. Post hoc analyses revealed that the three conditions differed significantly on average from each other. Baboons were faster to respond to the third target in the NAD condition (415 ± 46 ms) than in the two other conditions (Inconsistent condition: 434 ± 47 ms, $d = .41$; Consistent condition: 441 ± 46 ms, $d = .15$). Critically, they were also significantly slower in the inconsistent compared to the consistent condition ($d = .26$). A main effect of Language was found, F (1, 13) = 8.75, $p = .011$, $\eta_p^2 = .40$, RTs being faster in the second compared to the first language session. No reliable effect of Block was found, F (1, 13) = 1.52, $p = .24$, but the Block*Condition interaction was significant, F (2, 26) = 3.82, $p < .05$, $\eta_p^2 = .23$, RTs in the NAD condition only progressively increasing throughout the test phase. No other reliable interactions among these factors were found (all *p*s > .54). An ANOVA conducted on the first test block confirmed that a significant difference between the

consistent and inconsistent conditions was already present at the onset of that phase, as revealed by a main effect of Condition, $F(1, 13) = 9.53$, $p < .01$, $\eta_p^2 = .42$. No other main effect or interaction were found (*p*s > 12).

Spatial factors

An ANOVA revealed a significant interaction between Condition and Pair, $F(3, 39) = 4.56$, $p < .01$, $\eta_p^2 = .26$. Post-hoc tests revealed significantly longer RTs in the inconsistent compared to the consistent condition for two pairs: A_2-B_2 (416 ± 68 and 389 ± 44 ms, respectively**,** $d = .44$) and A_2-B_1 (467 ± 61 and 438 ± 53 ms, $d = .50$), but not for the A_1-B_1 (458 ± 38 and 463 ± 58 ms) and A_1-B_2 **(**379 ± 45 and 384 ± 43 ms) pairs. These results suggest that baboons have been able to generalize two out of the four NADs.

Table 3 reports the results of the one-tailed paired t-tests evaluating whether the generalization obtained with the two NADs mentioned above occurred on every novel Y location. RTs were longer in the inconsistent compared to the consistent condition for all Y locations, and this difference was significant for four out of the six Y locations (i.e., Y_2, Y_3, Y_4, Y_6), and approached significance for a fifth one (i.e., Y_5). Therefore, these results suggest that the baboons generalized the NADs over several novel Y locations, with only one exception for Y_1.

Table 3 Response times (Mean ± SD, ms) and results of the one-tailed paired t-tests assessing generalization over each novel Y locations, for the A_2-B_1 and A_2-B_2 NADs during the first test block. *df*s were all equal to 13.

	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
Inconsistent	426 ± 33	445 ± 69	434 ± 37	472 ± 76	435 ± 54	436 ± 49
Consistent	419 ± 39	391 ± 40	413 ± 35	423 ± 49	419 ± 65	415 ± 38
t	0.91	3.51	3.36	2.72	1.32	1.80
Cohen's *d*	.15	.96	.60	.77	.27	.48
p	.19	< .01	< .01	< .01	.10	< .05

Discussion

Two results were obtained in Experiment 2. First, baboons successfully learned the NADs during the exposure phase, replicating the findings already obtained in Experiment 1. Second, for two NADs response times were slower for inconsistent than for consistent test trials. This finding suggests generalization of these NADs over the novel intervening locations. Importantly, these results suggesting generalization cannot be accounted for by learning of the consistent sequences, since each beginning of a sequence (i.e., A_1-$Y_{1\text{-}6}$ and A_2-$Y_{1\text{-}6}$) was followed by a B_1 or B_2 location at equal frequency within each language session. Analyses moreover confirmed that the speed advantage for consistent sequences was already present at the onset of the test phase.

Interestingly, reliable generalization was only observed for the two NADs beginning by A_2 which was located in the lower-left corner of the matrix (see Fig. 3). By contrast, the baboons failed to properly generalize the two NADs beginning at the A_1 location, which was not located in a corner. This result suggests that the saliency of the predictive target had an impact on baboons' ability to generalize NADs in this task. We also note that generalization

of the above two NADs was not observed for the only Y element located in a corner of the matrix (Y_1), possibly because of its greater saliency, or eccentricity in the matrix, compared to the other Y locations which were more central. Taken together, these findings suggest that the baboons can generalize NADs over novel locations, but that this ability may be hindered by spatial factors.

General discussion

In Experiment 1, humans and baboons responded faster to the last target in the NAD compared to the control sequences. Importantly, participants did not merely learn that the sequences always finished by one of two locations (with a probability of .5 for each), but processed the conditional probabilities of the third location, given the first location. This result was replicated in baboons in Experiment 2. These data overall provide evidence of baboons' ability to learn NADs. However, while humans presented a significant difference in response times between the two conditions within the first exposure block, this difference appeared after approximately 400 presentations of each NAD in baboons. This difference in learning speed differs from what has been found for adjacent dependencies (Minier et al., 2016; Rey et al., in press). In these earlier studies, baboons and humans learned at similar speed (in less than 200 trials) the predictive relationships between two locations presented at the end of three different triplets. This suggests that learning of NADs is more difficult for baboons than humans, while these two species seem equally fast for learning adjacent dependencies. Importantly, the amount of exposure required by the baboons to extract NADs in the current experiment is in the same range as previously found with habituation-

dishabituation paradigms in other primate species (e.g., 672 presentations in Newport et al., 2004). This amount may serve as a reference for designing future experiments investigating related topics.

For two NADs we found that when presented with novel intervening locations the baboons were slower to respond when the sequences were inconsistent with the learned NADs than when the dependencies were preserved. To our knowledge, this result is the first demonstration in a non-human species of a successful generalization of sequentially non-adjacent dependencies over novel intervening elements. The current results therefore extend to the case of sequential NADs Sonnweber et al.'s findings (2015), in which generalization of NADs was tested between visual shapes presented concurrently. Along with previous findings in humans (e.g., Frost & Monaghan, 2016), our study indicates that this ability is shared across different primate species. However, successful generalization of the NADs in the current SRT study appeared to be modulated by spatial features of the stimuli. Indeed, the baboons only generalized the two NADs whose predictive target was located in a corner of the matrix and generalization was not observed for one of the novel intervening location which was also located in a corner. These data present some similarities with previous findings in the auditory domain. For instance, Endress, Scholl, and Mehler (2005, see also Peña et al., 2002) reported that human participants could properly extract some specific regularities from auditory sequences only when the relevant stimuli were presented at an edge of the sequences. Moreover, several studies manipulating the phonological parameters of the stimuli suggested that the relative saliency of the non-adjacent and intervening elements impacts the extraction of NADs (Creel et al., 2004; Onnis et al., 2005; Toro & Trobalón, 2005). Taken together, these findings suggest (1) that perceptual factors constrain the

extraction of NADs, and (2) that this effect occurs in both the auditory and visuo-spatial domains.

The exact nature of the processes involved in NADs extraction and recognition is still debated in the literature, and at least three theories are proposed. According to a first theory, the mechanisms that support NADs extraction are the same associative mechanisms as those allowing the learning of adjacent dependencies (Perruchet, Tyler, Galland, & Peereman, 2004). This single-process account proposes that the joint attentional processing of two elements is a necessary and sufficient condition to associate them, even when these are distant in space or time (Pacton & Perruchet, 2008). A second and complementary theory proposes that the extraction of adjacent and non-adjacent dependencies involves similar statistical learning mechanisms, but emphasize the role played by working memory for NADs extraction (Conway, Deocampo and Smith, 2016; Wilson et al., 2017). According to this dual-system theory, the detection of adjacent relationships would be sustained by modality-specific processes involving perceptual-motor neural networks (Conway et al., 2016), or ventral regions of the frontal cortex (Wilson, Kikuchi, et al., 2015), while the extraction of non-adjacent relationships would additionally recruit more dorsal areas (Wilson et al., 2017). According to a last hypothesis (Endress & Bonatti, 2007; Endress, Nespor, & Mehler, 2009; Peña et al., 2002), humans and possibly other animals (Endress, Carden, Versace, & Hauser, 2010) may be endowed with some rule-based mechanisms dedicated to fast learning of "classes" of elements occurring at the edges of sequences (e.g., here, learning that A_1 and A_2 always occur in the first position), as well as statistical learning mechanisms for the computation of transitional probabilities (e.g. A_1_B_1). The current experiment did not aim at

assessing the edge-based positional learning hypothesis but examined the extraction time course of those transitional probabilities.

In humans, learning occurred early in the experiment, and was strengthened throughout the entire exposure phase. In baboons, learning of the NADs occurred later in the experiment, but a similar progressive emergence was observed. In previous SRT experiments, we found that the learning of adjacent dependencies in triplets (Minier et al., 2016), or pairs of shapes (Fagot, Malassis, & Medam, 2017) gave rise to similar progressive (albeit earlier) decrease in response times throughout the exposure to these regularities. These data, along with previous findings (Frost & Monaghan, 2016; Romberg & Saffran, 2013; Vuong et al., 2016), suggest that adjacent and non-adjacent dependencies might be extracted by similar statistical learning mechanisms. However, they do not allow disentangling between the single-process and the dual-system proposals described above. Indeed, the difference in speed learning that is observed between baboons and humans might reflect an involvement of working memory in NADs extraction, as working memory capacity differ between these species (e.g., Fagot & De Lillo, 2011). However, addressing this question would require additional behavioral and brain imaging studies. Regarding behavioral studies, correlational designs in humans and non-human primates would allow examining individual differences in statistical learning ability across different types of dependencies (i.e., adjacent and non-adjacent, Siegelman, Bogaerts, Christiansen, & Frost, 2017; Siegelman & Frost, 2015), as well as exploring the relationships between working memory capacity and NADs extraction. Comparative neuroimaging studies have already suggested that adjacent dependencies are processed by evolutionary conserved brain structures, including the frontal operculum and anterior insula (Wilson, Kikuchi, et al., 2015). However, further researches are also needed in

this domain to discover whether NADs processing is sustained by similar or different neural substrates in human and non-human primates (Milne et al., 2016; Wilson et al., 2017).

Another important question is the domain generality of the mechanisms involved during the learning of NADs. For instance, are the extraction of NADs in the current SRT task and the extraction of NADs in natural languages (such as in *is* read*ing*) supported by the same learning mechanisms? Evidences accumulated so far reveal a contrasted pattern (Frost, Armstrong, Siegelman, & Christiansen, 2015). On the one hand, individuals with specific language impairment have been found to exhibit poor performance in auditory statistical learning tasks (e.g., Evans, Saffran, & Robe-Torres, 2009) as well as in SRT tasks (Lum, Conti-Ramsden, Morgan, & Ullman, 2014), compared with typically-developing individuals. They also express greater difficulties in processing NADs in an AGL-SRT task (Hsu, Tomblin, & Christiansen, 2014). In another study, Misyak, Christiansen, and Tomblin (2010) found a positive correlation between typically-developing individuals' ability to process long-distance dependencies from a natural language and NADs in an AGL-SRT task. On the other hand, statistical learning appears to be subject to modality and stimulus specificity (e.g., Conway & Christiansen, 2005, 2006). Siegelman and Frost (2015) reported for instance that the performance of their participants did not correlate between any of four statistical learning tasks they have used (auditory-verbal, auditory-nonverbal, visual, and SRT tasks). These results suggest the existence of modality and stimulus specific constraints on statistical learning mechanisms, which prevent drawing strong conclusions regarding the similarity between NADs extraction in natural languages and in our visuo-spatial task.

From an evolutionary standpoint, the ability to keep track of predictive relationships between non-adjacent events might present various functional advantages, for instance for

planning complex behavioral or motor sequences, or for monitoring social interactions and their outcomes (Sonnweber et al., 2015). In baboons, this process may allow for instance to detect that a threat behavior from an individual A is followed by a scream from an individual B, even if other individuals emit unrelated vocalizations in between. One general advantage of the learning of NADs might also be a reduction in memory demands. For instance in song birds, admitting an optional or variable song element between two others (thereby being non-adjacent) could limit the number of pairwise transitions needed to be memorized (Petkov & Wilson, 2012). In our experiments, the extraction of two NADs may be cognitively less demanding than the learning and storage of ten sequences.

Acknowledgments

This work, carried out within the Labex BLRI (ANR-11-LABX-0036) and the Institut Convergence ILCB (ANR-16-CONV-0002), has benefited from support from the French government, managed by the French National Agency for Research (ANR) and the Excellence Initiative of Aix-Marseille University (A*MIDEX). This research was supported by the Premilang2 ANR-project (#ANR-13-BSH2-0002) and the Chunked ANR-project (#ANR-17-CE28-0013-02).

References

Cate, C. ten, & Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1598), 1984–1994. https://doi.org/10.1098/rstb.2012.0055

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology. General*, *120*(3), 235–253.

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, *5*(12), 539–546. https://doi.org/10.1016/S1364-6613(00)01800-3

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(1), 24–39. https://doi.org/10.1037/0278-7393.31.1.24

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: pitting abstract against stimulus-specific representations. *Psychological Science*, *17*(10), 905–912. https://doi.org/10.1111/j.1467-9280.2006.01801.x

Conway, C. M., Deocampo, J., A., & Smith, G..N.L. (2016). Multiple Routes to Implicit Statistical Learning? A Dual-System Perspective. Abstracts of the Psychonomic Society, 57th annual meeting, Boston.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(5), 1119–1130. https://doi.org/10.1037/0278-7393.30.5.1119

de la Mora, D. M., & Toro, J. M. (2013). Rule learning over consonants and vowels in a non-human animal. *Cognition*, *126*(2), 307–312. https://doi.org/10.1016/j.cognition.2012.09.015

Endress, A. D. (2010). Learning melodies from non-adjacent tones. *Acta Psychologica*, *135*(2), 182–190. https://doi.org/10.1016/j.actpsy.2010.06.005

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*(2), 247–299. https://doi.org/10.1016/j.cognition.2006.09.010

Endress, A. D., Carden, S., Versace, E., & Hauser, M. D. (2010). The apes' edge: positional learning in chimpanzees and humans. *Animal Cognition*, *13*(3), 483–495. https://doi.org/10.1007/s10071-009-0299-8

Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences*, *13*(8), 348–353. https://doi.org/10.1016/j.tics.2009.05.005

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology. General*, *134*(3), 406–419. https://doi.org/10.1037/0096-3445.134.3.406

Endress, A. D., & Wood, J. N. (2011). From movements to actions: two mechanisms for learning action sequences. *Cognitive Psychology*, *63*(3), 141–171. https://doi.org/10.1016/j.cogpsych.2011.07.001

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical Learning in Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *52*(2), 321–335. https://doi.org/10.1044/1092-4388(2009/07-0189)

Fagot, J., & Bonté, E. (2010). Automated testing of cognitive performance in monkeys: use of a battery of computerized test systems by a troop of semi-free-ranging baboons (Papio papio). *Behavior Research Methods*, *42*(2), 507–516. https://doi.org/10.3758/BRM.42.2.507

Fagot, J., & De Lillo, C. (2011). A comparative study of working memory: Immediate serial spatial recall in baboons (Papio papio) and humans. *Neuropsychologia*, *49*(14), 3870–3880. https://doi.org/10.1016/j.neuropsychologia.2011.10.003

Fagot, J., Malassis, R., & Medam, T. (2017). The processing of positional information in a two-item sequence limits the emergence of symmetry in baboons (Papio papio), but not in humans (Homo sapiens). *Learning & Behavior*. https://doi.org/10.3758/s13420-017-0290-1

Fagot, J., & Paleressompoulle, D. (2009). Automatic testing of cognitive performance in baboons maintained in social groups. *Behavior Research Methods*, *41*(2), 396–404. https://doi.org/10.3758/BRM.41.2.396

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504.

Friederici, A. D. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends in Cognitive Sciences*, *8*(6), 245–247. https://doi.org/10.1016/j.tics.2004.04.013

Froehlich, A. L., Herbranson, W. T., Loper, J. D., Wood, D. M., & Shimp, C. P. (2004). Anticipating by Pigeons Depends on Local Statistical Information in a Serial Response Time Task. *Journal of Experimental Psychology: General*, *133*(1), 31–45. https://doi.org/10.1037/0096-3445.133.1.31

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010

Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, *147*, 70–74. https://doi.org/10.1016/j.cognition.2015.11.010

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review*, *16*(3), 486–490. https://doi.org/10.3758/PBR.16.3.486

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436.

Gómez, R., & Maye, J. (2005). The Developmental Trajectory of Nonadjacent Dependency Learning. *Infancy*, *7*(2), 183–206. https://doi.org/10.1207/s15327078in0702_4

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53–B64. https://doi.org/10.1016/S0010-0277(00)00132-3

Heijningen, C. A. A. van, Visser, J. de, Zuidema, W., & Cate, C. ten. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0908113106

Hsu, H. J., Tomblin, J. B., & Christiansen, M. H. (2014). Impaired statistical learning of non-adjacent dependencies in adolescents with specific language impairment. *Frontiers in Psychology*, *5*, 175. https://doi.org/10.3389/fpsyg.2014.00175

Kovács, Á. M., & Mehler, J. (2009). Flexible Learning of Multiple Speech Structures in Bilingual Infants. *Science*, *325*(5940), 611–612. https://doi.org/10.1126/science.1173947

Lum, J. A. G., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. T. (2014). Procedural learning deficits in specific language impairment (SLI): A meta-analysis of serial reaction time task performance. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *51*(100), 1–10. https://doi.org/10.1016/j.cortex.2013.10.011

Marchetto, E., & Bonatti, L. L. (2013). Words and possible words in early language acquisition. *Cognitive Psychology*, *67*(3), 130–150. https://doi.org/10.1016/j.cogpsych.2013.08.001

Marchetto, E., & Bonatti, L. L. (2015). Finding words and word structure in artificial speech: the development of infants' sensitivity to morphosyntactic regularities. *Journal of Child Language*, *42*(4), 873–902. https://doi.org/10.1017/S0305000914000452

Milne, A. E., Mueller, J. L., Männel, C., Attaheri, A., Friederici, A. D., & Petkov, C. I. (2016). Evolutionary origins of non-adjacent sequence processing in primate brain potentials. *Scientific Reports*, *6*, 36259. https://doi.org/10.1038/srep36259

Minier, L., Fagot, J., & Rey, A. (2016). The Temporal Dynamics of Regularity Extraction in Non-Human Primates. *Cognitive Science*, *40*(4), 1019–1030. https://doi.org/10.1111/cogs.12279

Misyak, J. B., Christiansen, M. H., & Bruce Tomblin, J. (2010). Sequential Expectations: The Role of Prediction-Based Learning in Language. *Topics in Cognitive Science*, *2*(1), 138–153. https://doi.org/10.1111/j.1756-8765.2009.01072.x

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162.

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, *49*(2), 85–117. https://doi.org/10.1016/j.cogpsych.2003.12.002

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8

Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2005). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies [Proceedings paper].

Onnis, Luca, Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, *53*(2), 225–237. https://doi.org/10.1016/j.jml.2005.02.011

Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(1), 80–96. https://doi.org/10.1037/0278-7393.34.1.80

Pacton, S., Sobaco, A., & Perruchet, P. (2015). Is an attention-based associative account of adjacent and nonadjacent dependency learning valid? *Acta Psychologica*, *157*, 195–199. https://doi.org/10.1016/j.actpsy.2015.03.002

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science (New York, N.Y.)*, *298*(5593), 604–607. https://doi.org/10.1126/science.1072901

Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: no need for algebraic-like computations. *Journal of Experimental Psychology. General*, *133*(4), 573–583. https://doi.org/10.1037/0096-3445.133.4.573

Petkov, C. I., & Wilson, B. (2012). On the pursuit of the brain network for proto-syntactic learning in non-human primates: conceptual issues and neurobiological hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 2077–2088. https://doi.org/10.1098/rstb.2012.0073

Ravignani, A., & Sonnweber, R. (2017). Chimpanzees process structural isomorphisms across sensory modalities. *Cognition*, *161*, 74–79. https://doi.org/10.1016/j.cognition.2017.01.005

Ravignani, A., Sonnweber, R.-S., Stobbe, N., & Fitch, W. T. (2013). Action at a distance: dependency sensitivity in a New World primate. *Biology Letters*, *9*(6), 20130852. https://doi.org/10.1098/rsbl.2013.0852

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

Rey, A., Minier, L., Malassis, R., Bogaerts, L., & Fagot, J. (in press). Regularity extraction across species: associative learning mechanisms shared by human and non-human primates. *Topics in Cognitive Science*.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews. Cognitive Science*, *1*(6), 906–914. https://doi.org/10.1002/wcs.78

Romberg, A. R., & Saffran, J. R. (2013). All Together Now: Concurrent Learning of Multiple Structures in an Artificial Language. *Cognitive Science*, *37*(7), 1290–1320. https://doi.org/10.1111/cogs.12050

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–1928.

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1711). https://doi.org/10.1098/rstb.2016.0059

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105–120. https://doi.org/10.1016/j.jml.2015.02.001

Sonnweber, R., Ravignani, A., & Fitch, W. T. (2015). Non-adjacent visual dependency learning in chimpanzees. *Animal Cognition*, *18*(3), 733–745. https://doi.org/10.1007/s10071-015-0840-x

Takahasi, M., Yamada, H., & Okanoya, K. (2010). Statistical and Prosodic Cues for Song Segmentation Learning by Bengalese Finches (Lonchura striata var. domestica). *Ethology*, *116*(6), 481–489. https://doi.org/10.1111/j.1439-0310.2010.01772.x

Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. L. (2008). Finding words and rules in a speech stream: functional differences between vowels and consonants. *Psychological Science*, *19*(2), 137–144. https://doi.org/10.1111/j.1467-9280.2008.02059.x

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, *67*(5), 867–875. https://doi.org/10.3758/BF03193539

Vuong, L. C., Meyer, A. S., & Christiansen, M. H. (2016). Concurrent Statistical Learning of Adjacent and Nonadjacent Dependencies. *Language Learning*, *66*(1), 8–30. https://doi.org/10.1111/lang.12137

Wilson, B., Kikuchi, Y., Sun, L., Hunter, D., Dick, F., Smith, K., … Petkov, C. I. (2015). Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. *Nature Communications*, *6*, 8901. https://doi.org/10.1038/ncomms9901

Wilson, B., Marslen-Wilson, W. D., & Petkov, C. I. (2017). Conserved Sequence Processing in Primate Frontal Cortex. *Trends in Neurosciences*, *0*(0). https://doi.org/10.1016/j.tins.2016.11.004

Wilson, B., Slater, H., Kikuchi, Y., Milne, A. E., Marslen-Wilson, W. D., Smith, K., & Petkov, C. I. (2013). Auditory Artificial Grammar Learning in Macaque and Marmoset Monkeys. *The Journal of Neuroscience*, *33*(48), 18825–18835. https://doi.org/10.1523/JNEUROSCI.2414-13.2013

Wilson, B., Smith, K., & Petkov, C. I. (2015). Mixed-complexity artificial grammar learning in humans and macaque monkeys: evaluating learning strategies. *The European Journal of Neuroscience*, *41*(5), 568–578. https://doi.org/10.1111/ejn.12834