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

Raphaëlle Malassis, Arnaud Rey, Joël R Fagot

HAL Id: hal-01980246

https://hal.science/hal-01980246

Submitted on 14 Jan 2019

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

1   Human and non-human primates share the ability to extract adjacent dependencies

2   and, under certain conditions, non-adjacent dependencies (i.e., predictive relationships

3   between elements that are separated by one or several intervening elements in a sequence). In

4   this study, we explore the online extraction dynamics of non-adjacent dependencies in

5   humans and baboons using a serial reaction time task. Participants had to produce three-target

6   sequences containing deterministic relationships between the first and last target locations. In

7   Experiment 1, participants from the two species could extract these non-adjacent

8   dependencies, but humans required less exposure than baboons. In Experiment 2, the data

9   show for the first time in a non-human primate species the successful generalization of

10   sequential non-adjacent dependencies over novel intervening items. These findings provide

11   new evidence to further constrain current theories about the nature and the evolutionary

12   origins of the learning mechanisms allowing the extraction of non-adjacent dependencies.

Keywords: language evolution, statistical learning, sequence learning, long-distance

dependencies, animal cognition

13

14   Statistical learning can be defined as the implicit learning of regularities embedded in

15   the environment, which has been proposed to play an important role in language acquisition

16   (for a review, see Romberg & Saffran, 2010) and many other aspects of cognition (e.g., in

17   visual perception, Fiser & Aslin, 2001). Several experimental paradigms have been used to

18   study these fundamental learning mechanisms such as the Artificial Grammar Learning

19   paradigm (AGL, Reber, 1967), the Artificial Language Learning paradigm (ALL, Saffran et

20   al., 1996) or the Serial Response Time paradigm (SRT, Nissen & Bullemer, 1987). In these

21   experiments, participants are typically exposed to sequences of nonsense stimuli (e.g.,

22   syllables, tones or visual shapes) that are organized with a specific grammar. Learning of the

23   predictive relationships embedded in the input can then be assessed by presenting novel

24   sequences that are either consistent or inconsistent with this grammar.

25   Comparative human/non-human studies can inform us about the nature and dynamic

26   of these learning mechanisms, and their occurrence during the evolution, by distinguishing

27   domain-general and evolutionary old processes from those that might have appeared more

28   recently in the human lineage. Comparative experiments have demonstrated so far that

29   animals, like humans, are highly proficient in learning predictive relationships between

30   *adjacent* elements (i.e., elements that are presented one after the other in a sequence, without

31   any delay or element in between; see for a review: Cate & Okanoya, 2012; Conway &

32   Christiansen, 2001; Wilson, Marslen-Wilson, & Petkov, 2017). This ability has been reported

33   in several primates (tamarins: Hauser, Newport, & Aslin, 2001; macaques: Wilson et al.,

34   2013; Wilson, Smith, & Petkov, 2015; and baboons: Minier, Fagot, & Rey, 2016), and non-

35   primate species (pigeons: Froehlich, Herbranson, Loper, Wood, & Shimp, 2004; rats: Toro &

36   Trobalón, 2005; and songbirds: Takahasi, Yamada, & Okanoya, 2010). Along with

37   neuroimaging data (e.g., Wilson, Kikuchi, et al., 2015), these behavioral data have led to the

38   hypothesis that the extraction of adjacent dependencies relies on evolutionarily conserved

39   mechanisms (Friederici, 2004; Wilson et al., 2017).

40   In contrast, the evolutionary origins of the capacity to extract non-adjacent

41   dependencies (i.e., predictive relationships between elements that are separated by one or

42   several intervening elements in a sequence, hereafter "NADs") remain a matter of debate. A

43   standard example of NADs in English is the relationship between auxiliaries and inflectional

44   morphemes that are separated in the speech stream by the verbal root (e.g., *is* read*ing*).

45   Crucially, extracting and recognizing these dependencies requires generalization over a

46   variable verb. Studies conducted in humans indicate that NADs extraction is more challenging

47   for humans than the extraction of regularities between adjacent elements (Cleeremans &

48   McClelland, 1991; Gebhart, Newport, & Aslin, 2009; Newport & Aslin, 2004; Pacton,

49   Sobaco, & Perruchet, 2015; Perruchet & Rey, 2005; Wilson, Smith, et al., 2015). This

50   capacity develops later in human infancy (Gómez & Maye, 2005) than the sensitivity to

51   adjacent dependencies (Saffran, Aslin, & Newport, 1996; but see Marchetto & Bonatti, 2013,

52   for different findings). Moreover, the extraction of NADs looks especially challenging in

53   human adults outside some specific facilitative contexts. For instance, it is facilitated when

54   the non-adjacent elements have a high degree of perceptual (e.g., phonological) similarity

55   (Creel, Newport, & Aslin, 2004; Gebhart et al., 2009; Onnis, Monaghan, Richmond, &

56   Chater, 2005), when the intervening elements are highly variable (Gómez, 2002; Onnis,

57   Monaghan, Christiansen, & Chater, 2005), or when the non-adjacent elements are located at

4

58   the edge of the sequences (Peña, Bonatti, Nespor, & Mehler, 2002). Studies also revealed

59   similar performances and constraints on NADs learning in experiments using linguistic and

60   nonlinguistic stimuli, such as tones (Creel et al., 2004; Endress, 2010; Gebhart et al., 2009), or

61   actions (Endress & Wood, 2011), suggesting that the learning mechanisms involved in this

62   ability are not language-specific. Several experiments aimed to test whether animals can also

63   extract NADs. Wilson et al. (2015) exposed human participants and macaques to a complex

64   grammar involving multiple adjacent and non-adjacent dependencies. All participants from

65   both species detected violations of the adjacent dependencies. By contrast, half of the tested

66   humans, and none of the monkeys were sensitive to the NADs. The authors concluded that

67   when multiple regularities are present, monkeys rely preferentially on local ones, whereas

68   humans exhibit more flexibility. Human learners can encounter some difficulties with the

69   learning of several regularities simultaneously (e.g., Kovács & Mehler, 2009), but Wilson at

70   al. (2015) suggests that this might be even more difficult for non-human learners. Animals'

71   focus on various local cues rather than on NADs has also been reported when multiple

72   embedded NADs must be processed, as in the so-called center-embedded grammars (e.g.,

73   Heijningen, Visser, Zuidema, & Cate, 2009). However, these results may point to a limitation

74   in the monkeys' ability to track several NADs simultaneously, rather than an inability to

75   extract NADs *per se.*

76   Complementary information on animals' ability to learn NADs has been obtained

77   from experiments using simpler grammars of the form $AX^{(n)}B$. In these grammars, A and B

78   are two paired elements, with a non-adjacent transitional probability of 1, and $X^{(n)}$ is one or

79   several (n) variable intervening elements. Two main types of NAD have been investigated

80   thus far using these grammars: dependencies between elements that are perceptually more

81   similar compared to the interspersed element(s) (i.e., that belong to the same perceptual

82   category, hereafter "feature-based NAD"), and learning of NADs without such perceptual

83   cues (often called "arbitrary associations").

84   Successful learning of feature-based NADs has been recently reported in two primate

85   species, with visual (Sonnweber, Ravignani, & Fitch, 2015) and auditory stimuli (Ravignani,

86   Sonnweber, Stobbe, & Fitch, 2013). Successful generalization across isomorphic visual and

87   auditory sequences was also demonstrated recently (Ravignani & Sonnweber, 2017). In

88   addition, positive results were obtained in rats, with NADs between elements that were

89   phonologically similar (i.e., vowels vs. consonants, and *vice-versa*; de la Mora & Toro, 2013),

90   whereas previous experiments conducted on NADs extraction in this species led to negative

91   results in the absence of phonological similarity (Toro & Trobalón, 2005). Perceptual

92   similarity therefore appears to facilitate the extraction of NADs in some non-human animal

93   species, as it does in humans (e.g., Onnis et al., 2005). These findings support the hypothesis

94   that NADs extraction is sustained by a general-purpose learning mechanism, interacting with

95   general perceptual constraints (i.e., extra-linguistic, such as Gestalt principle of similarity,

96   Creel et al., 2004), both being shared by phylogenetically distant species (Newport & Aslin,

97   2004; see however de la Mora & Toro, 2013, and Toro, Nespor, Mehler, & Bonatti, 2008, for

98   a discussion of potential human-specific constraints on speech processing).

99   Evidence for successful NADs learning in the absence of such perceptual cues can also

100   be found in non-human animals. To date, three distinct experiments were conducted on that

101   topic. Newport, Hauser, Spaepen, and Aslin (2004) assessed the ability of tamarins (*Saguinus*

102   *oedipus*) to extract dependencies between non-adjacent syllables in a speech stream, using a

103   familiarization procedure. Six 3-syllable nonsense words of the form AXB (three

104   instantiations of A-B pairs and one of two X syllables inserted in the middle) were presented

105   in a continuous stream during several minutes via a loudspeaker. After this familiarization

106   phase, the tamarins' reactions to words (e.g., A_1XB_1) *versus* part-words (B_1A_2X) were tested.

107   Tamarins exhibited a higher rate of orientation responses towards the speaker after a part-

108   word than after a word, suggesting that they learned the NADs. The authors concluded that

109   the tamarins accurately segmented this continuous stream into triplets of syllables based on

110   these NADs. However, one limitation of this study is that the tamarins' ability to generalize

111   these dependencies over novel intervening X syllables was not tested. Therefore, this

112   experiment leaves unexplored the question of how the NADs were processed and stored in

113   their memory. Indeed, they could have extracted the three A-B non-adjacent relationships, or

114   instead memorized the six AXB "words" as wholes. The same limitation also applies to Milne

115   et al. (2016) who used a passive listening method in rhesus macaques (*Macaca mulatta*) and

116   the recording of event-related potentials. Despite the results suggesting the learning of NADs,

117   this study failed to provide generalization tests to confirm that the learned NADs could be

118   maintained over novel intervening events.

119   To our knowledge, Sonnweber et al. (2015) is the only study proposing this kind of

120   test in a non-human species. These data were obtained in one subgroup of chimpanzees (*Pan*

121   *troglodytes)* referred to as the "arbitrary associative dependencies" group. These chimpanzees

122   were presented on each trial with two "strings" of visual shapes depicted concurrently on a

123   screen in a two-alternative forced choice task. The task required to choose the strings of the
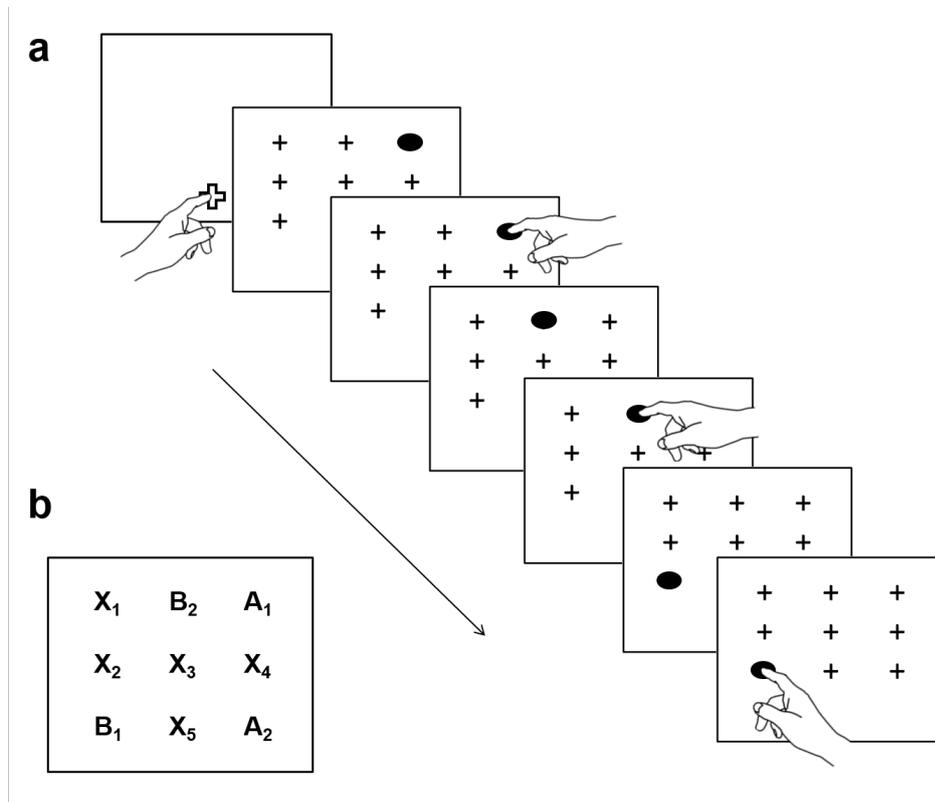
124   form AX^nB (with five instantiations of A-B pairs and 60 instantiations of X items, with $n = 1$

125   or 2) over XX^nX strings. What the participants precisely learned was evaluated by a series of

126   tests proposed after the training phase. The two main generalization tests (Test 2: extension of

127   the number of X elements, and Test 3: introduction of novel X shapes) were successfully

128   completed by one participant over the three tested chimpanzees. However, the fact that the

129   items were presented concurrently rather than sequentially raises potential issues. The

130   processing of associations between spatially distant elements, especially at rather small

131   distance (up to 12 cm from each other during training, with a maximum of 30 cm in Test 2,

132   according to the reported pixel measures), can arguably rely on strategies different from those

133   available for the extraction of sequential NADs. For instance, once the chimpanzees have

134   learned that only the left- and right-most shapes of each string matter in this task, they can

135   make quick eye-movements from one to the other and process them as if they were adjacent,

136   despite the spatial distance. Extraction and recognition of sequential NADs, in contrast, may

137   require to hold an element in working memory and relate it to another element occurring later

138   in the sequence (Wilson et al., 2017). Therefore, Sonnweber et al. (2015) offers promising

139   findings about non-human primates' ability to generalize NADs, but these researches would

140   need to be extended to the case of sequential NADs. Another limitation of the comparative

141   literature is that it contains no comparative data on the temporal dynamics of NADs learning

142   in human and non-human animal species. Earlier, we reported that baboons and humans

143   learned adjacent dependencies at the same speed (Minier et al., 2016; Rey, Minier, Malassis,

144   Bogaerts, & Fagot, submitted). A remaining issue is whether this is also true for NADs.

145   The aim of the present study was twofold in this context. It was firstly to provide

146   human and baboon comparative data on the fine-grained dynamics of NADs extraction. It was

147   secondly to test whether baboons can generalize sequential NADs over novel intervening

148   elements. In Experiment 1, baboons and humans performed a serial response time task (SRT,

149   Nissen & Bullemer, 1987), requiring to follow a stimulus appearing sequentially at different

150   locations on a touchscreen. In a subset of the presented sequences, the first target location

151   predicted the last target location, while the second target location varied. Response times

152   provided an online behavioral metric of NADs learning. In Experiment 2, generalization of

153   the NADs over novel intervening locations was tested. During this test, novel consistent

154   sequences were contrasted with novel sequences violating the dependencies. This violation

155   method was inspired from Gómez (2002). Only baboons took part in this second part of the

156   study, as generalization has already been demonstrated in humans (Endress & Bonatti, 2007;

157   Frost & Monaghan, 2016; Marchetto & Bonatti, 2015).

158

159   **Experiment 1**

160   **Method**

161   **Baboons**

162   *Participants and apparatus*

163   Participants are 16 Guinea baboons (*Papio papio*, 11 females, age range 4-21 years)

164   from the CNRS primate facility (Rousset-sur-Arc, France). They live within a larger group of

165   26 individuals, within a 700 m2 outdoor enclosure and had a permanent access to ten

166   Automated Learning Devices for Monkeys (ALDM, for a detailed description, see Fagot &

167   Bonté, 2010; Fagot & Paleressompoulle, 2009) equipped with a 19-inch touch screen and a

168   food dispenser. The main feature of ALDM equipment is that a radio frequency identification

169   reader (RFID) identifies each baboon via a microchip implanted in each arm. The baboons

170   can therefore participate to the research at will, without being captured, as the test programs

171   recognize them automatically. All baboons had previously participated to numerous

172   computerized experiments using the ALDM test systems, including experiments on sequence

173   processing (Minier et al., 2016), but they have never been exposed to tasks requiring the

174   learning of NADs. The experiment was controlled by the EPrime software (Version 2.0,

175   Psychology Software Tools, Pittsburgh).

176   ***Procedure***

177   Participants had to perform a serial response time (SRT) task which required following

178   and touching a dot moving in a pre-defined order on a touch screen. Each trial began by the

179   display of a fixation cross (120x120 pixels) presented at the bottom of the screen (see Fig.

180   1a). Touching this stimulus triggered the display of a matrix of 9 locations (3x3) which

181   contained 8 white crosses (60x60 pixels each) and a red circle target (80x80 pixels).

182   Participants had to touch the target. The target then disappeared and moved to the next

183   location on the screen. There was no delay between the offset of the target and its next

184   display. A trial consisted in a sequence of three targets. An accurate completion of the 3-target

185   sequence delivered grains of dry wheat inside the ALDM test unit. An incorrect response (i.e.,

186   selection of an incorrect location) stopped the trial and triggered a 5-sec timeout without food

187   reward. Trials in which the participants failed to select a stimulus within 5-sec after the

188   target's appearance were aborted and presented again in the next trial. The time elapsed

189   between the appearance of the target (i.e., the red circle) and the baboon's touch on this target

10

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

191   target sequence.

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

197

198   Two types of 3-target sequences were presented. In the first type ("NAD" condition),

199   the first location of the target systematically predicted its last location, leading to a non-

200   adjacent Transitional Probability (TP) of 1. Two pairs of A-B locations were used (A_1-B_1 and

201   A_2-B_2), and a variable X location was inserted between A and B locations. The sequences

202   were therefore of the form A_1XB_1 and A_2XB_2. The full set of 9 possible locations consisted in

203   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

204   Fig. 1b). These locations were chosen pseudo-randomly among the nine possible spatial

205   locations of the matrix, with the sole constraint that neither A_1 and B_1, nor A_2 and B_2, could

206   be located next to each other. As the five X locations were presented at equal frequencies, the

207   second location was uninformative about the third target location. The TP between the first

208   and the second target locations was equal to 0.2, and the TP between the second and the third

209   was equal to 0.5. A total of ten (two A-B pairs * five X locations) NAD sequences were used

210   (see Table 1).

211   The second type of sequences had an XXB_1 and XXB_2 structure. These sequences

212   were used in the control condition. To equate for the number of sequences per condition, 10

213   control sequences were constructed, as illustrated in Table 1. TPs between the first and third

214   locations, as well as between the second and third locations, were equal to 0.5 in these

215   sequences. It was therefore impossible to use adjacent or non-adjacent dependencies to predict

216   the third target location. NAD and control sequences were presented at equal frequencies

217   across the experiment. This equates motor practice in the two conditions. Experiment 1

218   involved a total of twenty 100-trial blocks. Each block contained five randomly intermixed

219   presentations of the 10 NAD and 10 control sequences described above. All baboons were

220   presented with the same sequences. Custom-written Python code (www.python.org) was used

221   to generate the sequences.

222

223

Table 1 The 20 sequences presented in Experiment 1.

NAD	Control
$A_1\,X_1\,B_1$	$X_3\,X_1\,B_1$
$A_1\,X_2\,B_1$	$X_5\,X_2\,B_1$
$A_1\,X_3\,B_1$	$X_2\,X_3\,B_1$
$A_1\,X_4\,B_1$	$X_1\,X_4\,B_1$
$A_1\,X_5\,B_1$	$X_4\,X_5\,B_1$
$A_2\,X_1\,B_2$	$X_3\,X_1\,B_2$
$A_2\,X_2\,B_2$	$X_5\,X_2\,B_2$
$A_2\,X_3\,B_2$	$X_2\,X_3\,B_2$
$A_2\,X_4\,B_2$	$X_1\,X_4\,B_2$
$A_2\,X_5\,B_2$	$X_4\,X_5\,B_2$

224

225   ***Training***

226   The baboons were familiar with the general principle of the task at the beginning of

227   the experiment, as they previously performed a SRT task involving 9-target sequences

228   (Minier et al., 2016). Before the experimental phase described above, the baboons received an

229   initial training phase of twenty100-trial blocks in which the first, second and third target

230   locations were selected randomly in the matrix. These random sequences were constructed

231   with the constraint to systematically avoid the repetition of any given target location within

232   the 3-target sequence. The aim of the training phase was twofold: (1) to familiarize the

233   baboons with the 3-target sequences (instead of 9 as in Minier et al., 2016), and (2) to limit

234   any long-term effect of the regularities previously learned.

235   ***Data analyses***

236   Our set of statistical analyses compared response times on the third target (RTs) in the

237   NAD and control conditions. Remember that the NAD and control sequences were presented

238   at equal frequencies across the experiment. This procedure allowed us distinguishing response

239   time accelerations due to a practice effect, as participants performed the control and NAD

240   sequences an equal number of times, and those attributable to the learning of the predictive

241   relationships present in the NAD sequences. Importantly, the final transitions between the

242   second and the third target locations were the same across conditions (i.e., X_{1-5}-B_1 and X_{1-5}-

243   B_2). Motor constraints were therefore controlled for that last transition. Shorter RTs in the

244   NAD compared to the control condition would therefore indicate successful learning of the

245   NADs.

246   For statistical analyses, the data of the exposure phase were grouped in five blocks of

247   400 trials. For each participant, mean response times on the third target (RTs) was computed

248   for each condition and block. These data were then analyzed using repeated-measures

249   ANOVA involving the Condition (NAD and Control) and Block (1-5) as within-participant

250   factors. Post-hoc comparisons were performed using Tukey Honestly tests ($p < .05$)

251   On average, the sixteen baboons required five days of testing (range 3-7 days) to

252   complete the experiment. Incorrect trials were removed from the data set (1.6%), as were

253   removed all the trials with response times greater than two standard deviations from the mean

254   (5.0% of the remaining trials, computed for each participant and each block).

255

256   **Humans**

257   *Participants and apparatus*

258   Ten human participants participated in this experiment (five females, age range 19-

259   24). All participants were students of Aix-Marseille University and were paid for their

260   participation. All participants were right-handed, had normal or corrected-to-normal vision,

261   and were naïve as to the goal of the study. Participants performed the task while sited in front

262   of a 19-inch touch screen. The experiment lasted approximately three hours per participant,

263   and was divided in two sessions. Testing sessions were separated by two days at most.

264   *Procedure and data analyses*

265   The task and stimuli were similar in their general principles for humans and for

266   baboons, with only slight differences between species. First, the number of blocks was

267   reduced by half in humans (n = 10), in comparison to baboons (n = 20). This reduction was

268   motivated by the long duration of the experiment (i.e., three hours with the current design).

269   Second, humans did not receive a food reward at the end of each correct trial, contrary to

270   baboons. The duration of the time-out and inter-trial interval were otherwise similar in the two

271   species. Third, humans were instructed at the beginning of the experiment to follow the red

272   target as quickly and accurately as possible. They were also told that a green screen meant

273   that they did not correctly touch the target. Participants were interviewed at the end of the

274   experiment to assess their declarative knowledge of the regularities embedded in the

275   sequences. They were asked: "Did you notice any regularity, or rule, in some of the

276   sequences?".

277   Regarding the training phase, human participants were first trained on the SRT task

278   using the same procedure as for baboons, but their number of training trials was reduced to

279   ten 100-trial blocks. Statistical analyses were the same as for baboons, except that blocks of

280   the exposure phase were grouped in five blocks of 200 trials (instead of 400).

281   Incorrect trials were removed from the data set (7.3%), as well as trials with response

282   times greater than two standard deviations from the mean (3.3% of the remaining trials,

283   computed for each participant and each block).
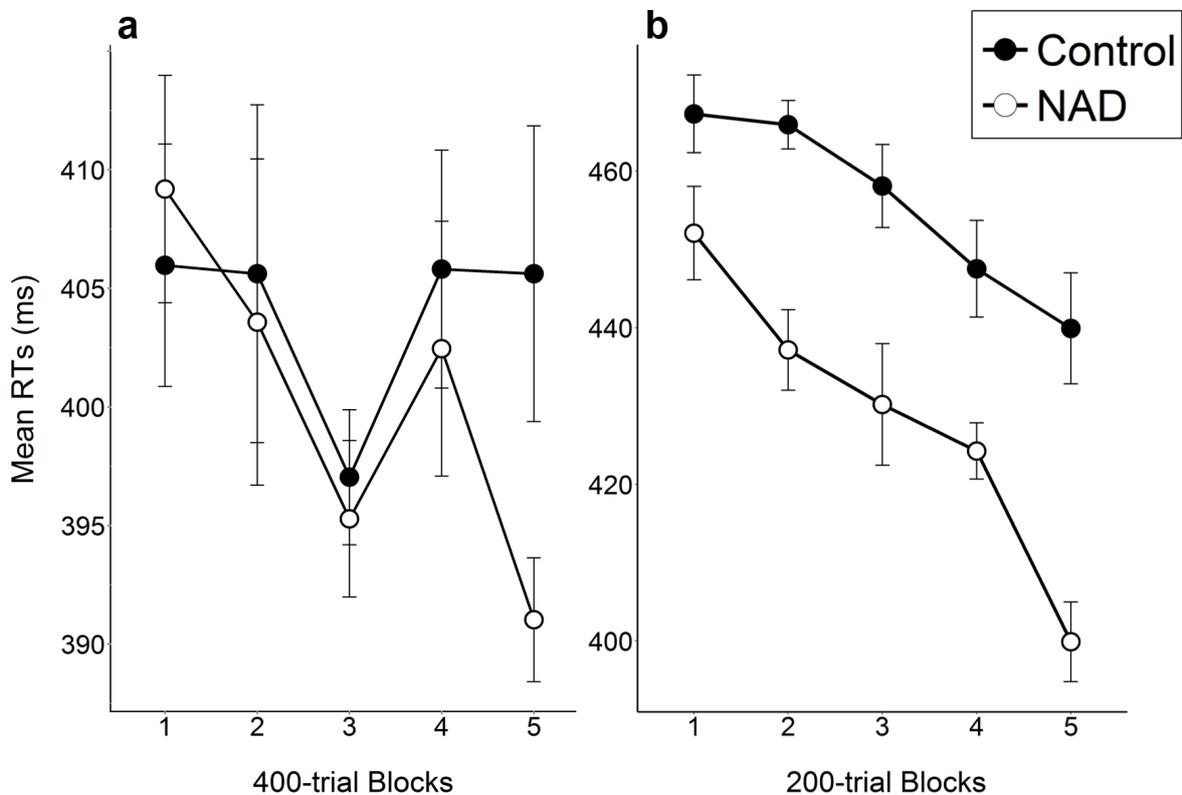
284

285   **Results**

286   **Baboons**

287   Mean RTs obtained in baboons are reported for each condition and block in Fig. 2a.

288   The repeated measures ANOVA revealed that the effect of Condition did not reach the

289   standard significance level (i.e., .05), $F (1, 15) = 4.21$, $p = .06$, $\eta_p^2 = .22$. No main effect of

290   Block was found[1], $F (1, 15) = 1.21$, $p = .29$, but there was a significant interaction between

291   these two factors, $F (1, 15) = 13.73$, $p < .01$, $\eta_p^2 = .48$. Post-hoc analyses revealed that RTs

292   were significantly shorter in NAD compared to the control trials for the last block (Block 5,

293   Cohen's $d = .29$), but not for the previous blocks (1-4). Altogether, the findings suggest that

294   baboons successfully learned the NADs in our task, but that this learning took a substantial

295   number of trials.

296

---

[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.

297

298

299 <u>Figure 2</u>. Response times for the third target (RTs) depending on the block, for the control and
300 the non-adjacent dependencies (NAD) conditions in Experiment 1, for baboons (a) and for
301 humans (b). The bars represent standard errors from normalized data.
302

303

304 **Humans**

305 　　　　Mean RTs are reported for each condition and block in Fig. 2b. Repeated measure

306 ANOVAs revealed a main effect of Condition, $F_{(1, 9)} = 38.60$, $p < .001$, $\eta_p^2 = .81$, indicating

307 faster responses in the NAD (429 ± 65 ms) compared to the control condition (456 ± 58). A

308 reliable main effect of Block was found as well, $F_{(1, 9)} = 30.99$, $p < .001$, $\eta_p^2 = .77$, as was

309 found a significant Block*Condition interaction, $F_{(1, 9)} = 11.87$, $p < .01$, $\eta_p^2 = .57$. Post hoc

310   analyses on this interaction indicated a RTs advantage in the NAD condition compared to the

311   control condition that was significant from the first block (ds = Block 1: .25, Block 2: .42,

312   Block 3: .39, Block 4: .39, Block 5: .70). These results suggest that humans began to learn the

313   NADs earlier than baboons, during the first block, and that this learning was amplified

314   throughout the course of the experiment.

315   Post-experiment interviews revealed that none of the participants reported the

316   existence of a predictive relationship between the first and the last locations (i.e., neither A_1-

317   B_1 nor A_2-B_2 relationships were reported). However, 4 out of 10 participants correctly

318   reported that two locations only were used at the last step of the sequence. In addition, four

319   participants could explicitly verbalize some of the triplets (from one to four triplets,

320   depending on the participant). Six of these triplets of locations were NAD sequences, and four

321   were control sequences. To assess whether the results we obtained were triggered by the

322   participants' declarative knowledge of this subset of sequences, we replicated the analysis

323   described above after removing from the dataset the corresponding trials, and these exclusions

324   did not change the pattern of statistical results. Overall, these data suggest that human

325   participants could extract the NADs. They further suggest that while human participants

326   exhibited learning of the regularities, as inferred from their response times, this learning was

327   implicit.

328

329   **Discussion**

330   In Experiment 1, baboons and humans could extract the NADs, as both species

331   responded faster to the third target when its location was predicted by the first target location

18

332   (NAD condition) than when it was not (control condition). Fig. 1 suggests that this difference

333   progressively increased in both species, with substantial differences between them. A

334   significant difference in RTs between the two conditions was found from the first block of

335   exposure in humans. This effect only emerged after five blocks in baboons, revealing a

336   substantial difference in the amount of exposure necessary for each species to exhibit

337   successful learning of the NADs. Considering that each of the two NADs was presented 100

338   times within a 400-trial block, we estimate that baboons demonstrated learning after 400 (100

339   * 4 blocks) presentations of each NAD. In contrast, this advantage in response times reached

340   significance since the first 200-trial block in humans, meaning that learning took place within

341   the first 50 presentations of these NADs. Moreover, effect size of the difference between the

342   two conditions on the last block was small to medium in baboons ($d = .29$), while it was

343   medium to large in humans ($d = .70$).

344   However, a discrepancy in task practice between the two species might partly explain

345   these differences between humans and baboons. Indeed, all our baboons had already

346   participated to a study involving a SRT task (Minier et al., 2016), and performed 20 blocks of

347   random sequences before completing the current experiment. In contrast, human participants

348   only performed 10 random blocks and never had been trained to the SRT task before. This

349   discrepancy in task practice between the two species might explain why humans' response

350   times decreased progressively throughout the experiment in both the control and the NAD

351   conditions, whereas baboons only showed this acceleration in the NAD condition.

352   Importantly, the baboons responded faster on average than humans on the first block (Mean ±

353   SD = 406 ± 43 and 460 ± 62 ms, respectively, t (24) = 2.58, $p = .02$, $d = .99$), probably as a

354   consequence of their extensive practice with the task, while the two species show comparable

355   RTs on the last block (400 ± 54 and 420 ± 56 ms, respectively, t (24) = 0.89, p = .38). These

356   data suggest that a ceiling effect might have limited baboon's acceleration in the NAD

357   condition.

358   One limitation of this first experiment is that it does not rule out the possibility that the

359   participants progressively learned the NAD sequences by rote, or a subset of them, and

360   responded faster at the third location because of this rote learning, but without extracting the

361   NADs embedded in these sequences. Experiment 2 addresses this hypothesis in baboons by

362   testing their ability to generalize the NADs over novel intervening locations.

363

364   **Experiment 2**

365   Experiment 2 tested whether baboons merely learned the ten A_1XB_1 and A_2XB_2

366   sequences as wholes, or extracted instead the $A_1\_B_1$ and $A_2\_B_2$ non-adjacent relationships.

367   Baboons were exposed successively to two different "languages"[2] in a within-participant

368   design, and two experimental phases were proposed for each language. The first phase

369   corresponded to the exposure phase of Experiment 1, except that the set of possible locations

370   was increased from 9 to 16 locations. Six of these locations remained unused during that

371   phase. During the second phase, which is referred to as the test phase, the baboons were

372   presented with two types of novel sequences. In half of these sequences, the NADs learnt

373   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".

374 phase. The dependencies were inconsistent in the remaining half of the sequences (i.e., A_1

375 was presented with B_2 and A_2 with B_1). Importantly, the consistent sequences in Language 1

376 were the inconsistent sequences in Language 2, and *vice-versa*. This two-language design

377 therefore guaranties that any difference in response times between the conditions will reflect

378 the learning and generalization of the NADs, rather than idiosyncratic sequence features

379 related to differences in motor constraints.

380

381 **Methods**

382   ***Participants and apparatus***

383   Experiment 2 involved 14 (10 females, age range 4-21 years) out of 16 baboons of

384 Experiment 1, because we were unable to test two of the previous participants for reasons

385 unrelated to our research. The apparatus was the same as in the previous experiment. The

386 general testing procedure was also the same as in Experiment 1, except that the display

387 contained 16 (4x4) possible locations rather than only 9 (3x3).

388   ***Procedure***

389   Table 2 provides the full lists of sequences used in Languages 1 and 2. The first

390 exposure phase was identical to Experiment 1 in both languages. In Language 1, sequences

391 took the form A_1XB_1 and A_2XB_2 for the NAD trials (50% of the trials), and XXB_1 and XXB_2

392 for the control trials (50% of the trials; see Table 2), with six different instantiations of the X

393 locations. In Language 2, the sequences took the form A_1XB_2 and A_2XB_1 for the NAD trials,

394 and control sequences were the same as in Language 1. The remaining six locations (noted

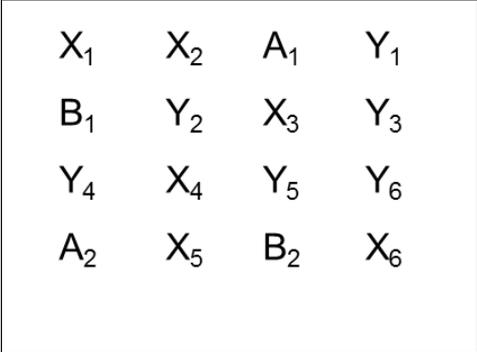395 Y_{1-6}) were never used during the exposure phase. Locations have been chosen pseudo-

396   randomly, with the sole constraint that A_1 and B_1, as well as A_2 and B_2, could not be located

397   next to each other in the matrix. Fig. 3 shows the sixteen locations on the touchscreen that

398   were used in Experiment 2. Note that the locations were the same in both languages, and that

399   only the NADs differed between them.

400   Baboons then received three types of sequences during the test phase. The first type

401   (50% of the trials) corresponded to the NAD sequences of the exposure phase. For the two

402   other types of trials, sequences were either consistent with the NADs of the exposure phase

403   ("Consistent" condition, 25% of the test trials), or inconsistent with those dependencies

404   ("Inconsistent" condition, 25%). In these two latter conditions, the second location of the

405   sequence was drawn from the set of Y locations never used during the exposure phase.

406   Therefore, sequences were novel in these two conditions. In Language 1, consistent sequences

407   took the form A_1YB_1 and A_2YB_2, while inconsistent sequences took the form A_1YB_2 and

408   A_2YB_1. In Language 2, the consistent sequences were the inconsistent sequences of Language

409   1 (i.e., A_1YB_2 and A_2YB_1), while the inconsistent sequences were the consistent ones of

410   Language 1 (i.e., A_1YB_1 and A_2YB_2). Twelve sequences were constructed for each condition.

411   NAD, consistent and inconsistent trials were randomly intermixed within each block of the

412   test phase.

413   Each baboon participated successively to the two language sessions. For each

414   language, the baboons received 40 blocks of 96 trials during the exposure phase, and 20

415   blocks of 96 trials in the test phase. Note that the amount of exposure changed from 20 blocks

416   in Experiment 1 to 40 in Experiment 2, for both languages, because we hypothesized that

417   learning was still in progress during the last blocks of Experiment 1. To limit potential

418   influences of Language 1 acquisition on the learning of Language 2, six months elapsed

419   between the two language sessions, during which the baboons were exposed to different

420   unrelated tasks. Furthermore, a random phase consisting of 35 blocks of 96 trial blocks was

421   given to the baboons at the beginning of the Language 2 session.

422

$$
\begin{array}{cccc}
X_1 & X_2 & A_1 & Y_1 \\
B_1 & Y_2 & X_3 & Y_3 \\
Y_4 & X_4 & Y_5 & Y_6 \\
A_2 & X_5 & B_2 & X_6
\end{array}
$$

423

424   Figure 3. Locations on the touchscreen in Experiment 2. Letters are given for illustrative
425   purposes but were not displayed. Note that the locations were the same for Languages 1 and
426   2.

427

428

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-

439   5) as within-participant factors. RTs in the consistent and inconsistent conditions were further

440   compared for the first test block using repeated-measures ANOVA involving the Condition

441   (Consistent, Inconsistent) and Language (1, 2) as within-participant factors. RTs in the

442   consistent and inconsistent conditions were further compared for the first test block using a

443   repeated-measures ANOVA involving the Condition (Consistent, Inconsistent) and Language

444   (Language 1, Language 2).

445   We also investigated potential effects of spatial factors on the generalization of NADs.

446   This analysis was conducted with an ANOVA using the Condition (Consistent, Inconsistent)

447   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

448   the first test block as dependent variable. An additional set of analyses tested whether

449   generalization occurred for every novel Y location. One-tailed paired t-tests ($p < .05$) were

450   used to assess if RTs were reliably longer in the inconsistent compared to the consistent

451   condition for each Y location.

452   To complete Language 1 session, the fourteen baboons required 8 days (range 5-12)

453   on average. The same baboons required on average 10 days (range 5-17) of testing to

454   complete Language 2 session. Incorrect trials were removed from the data set (Language 1:

455   1.7%, Language 2: 2.1%), as well as trials with RTs greater than two standard deviations from

456   the mean (4.9 % and 5.1%, respectively, computed for each participant and each block).
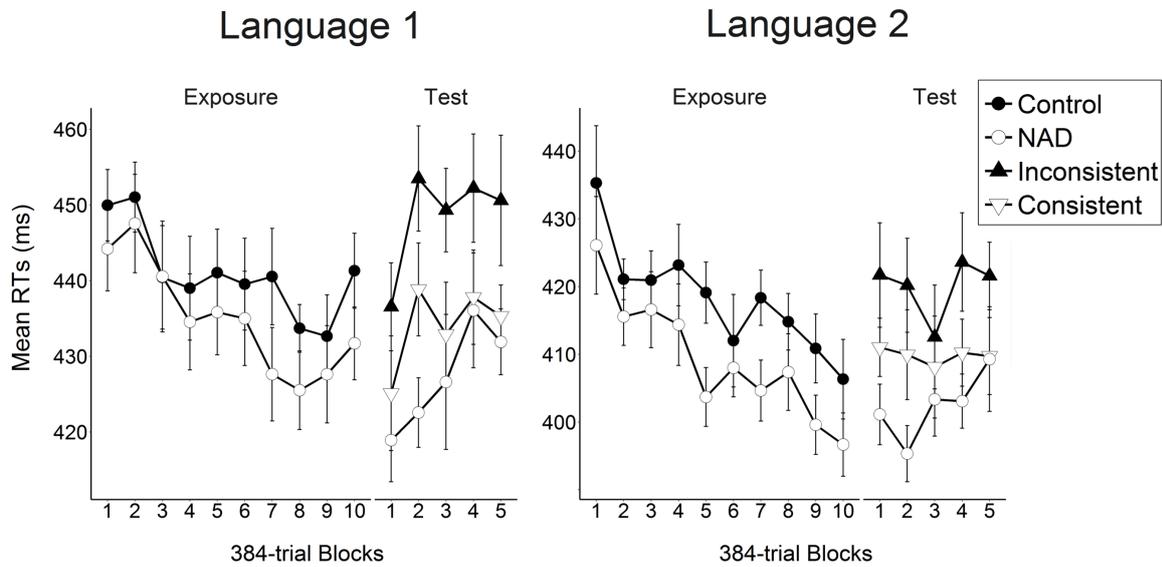
457

458   **Main Results**

459   Mean RTs are reported in Fig. 4 for each language, phase and condition. For the

460   exposure phase, the main effect of Condition was significant, $F (1, 13) = 16.28$, $p = .001$, $\eta_p^2 =$

461   .56, corresponding to faster response times in the NAD (Mean ± SD = 422 ± 44 ms) than in

462   the control condition (430 ± 43 ms). Also significant was the main effect of Block, $F(1, 13) =$

463   16.88, $p = .001$, $\eta_p^2 = .56$, showing that RTs decreased with practice. Finally, the main effect

464   of Language was also significant, $F(1, 13) = 6.60$, $p < .01$, $\eta_p^2 = .34$, indicating faster

465   response times on average for the second (414 ± 32 ms) compared to the first language (438 ±

466   50 ms). The Condition*Block interaction was close to significance level, $F(1, 13) = 3.21$, $p =$

467   .10, $\eta_p^2 = .20$. There were no other interactions (all $ps > .42$). The main effect of condition,

468   along with the absence of Condition*Language interaction, indicate that the baboons were

469   able to learn the NADs in both languages.

470

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 $ps > .54$). An ANOVA conducted on the first test block confirmed that a significant difference between the

27

488   consistent and inconsistent conditions was already present at the onset of that phase, as

489   revealed by a main effect of Condition, $F(1, 13) = 9.53$, $p < .01$, $\eta_p^2 = .42$. No other main

490   effect or interaction were found ($ps > 12$).

491

492   **Spatial factors**

493   An ANOVA revealed a significant interaction between Condition and Pair, $F(3, 39) =$

494   4.56, $p < .01$, $\eta_p^2 = .26$. Post-hoc tests revealed significantly longer RTs in the inconsistent

495   compared to the consistent condition for two pairs: A_2-B_2 (416 ± 68 and 389 ± 44 ms,

496   respectively, $d = .44$) and A_2-B_1 (467 ± 61 and 438 ± 53 ms, $d = .50$), but not for the A_1-B_1

497   (458 ± 38 and 463 ± 58 ms) and A_1-B_2 (379 ± 45 and 384 ± 43 ms) pairs. These results

498   suggest that baboons have been able to generalize two out of the four NADs.

499   Table 3 reports the results of the one-tailed paired t-tests evaluating whether the

500   generalization obtained with the two NADs mentioned above occurred on every novel Y

501   location. RTs were longer in the inconsistent compared to the consistent condition for all Y

502   locations, and this difference was significant for four out of the six Y locations (i.e., Y_2, Y_3,

503   Y_4, Y_6), and approached significance for a fifth one (i.e., Y_5). Therefore, these results suggest

504   that the baboons generalized the NADs over several novel Y locations, with only one

505   exception for Y_1.

506

507

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

508

509   **Discussion**

510   Two results were obtained in Experiment 2. First, baboons successfully learned the

511   NADs during the exposure phase, replicating the findings already obtained in Experiment 1.

512   Second, for two NADs response times were slower for inconsistent than for consistent test

513   trials. This finding suggests generalization of these NADs over the novel intervening

514   locations. Importantly, these results suggesting generalization cannot be accounted for by

515   learning of the consistent sequences, since each beginning of a sequence (i.e., A_1-Y_{1-6} and A_2-

516   Y_{1-6}) was followed by a B_1 or B_2 location at equal frequency within each language session.

517   Analyses moreover confirmed that the speed advantage for consistent sequences was already

518   present at the onset of the test phase.

519   Interestingly, reliable generalization was only observed for the two NADs beginning

520   by A_2 which was located in the lower-left corner of the matrix (see Fig. 3). By contrast, the

521   baboons failed to properly generalize the two NADs beginning at the A_1 location, which was

522   not located in a corner. This result suggests that the saliency of the predictive target had an

523   impact on baboons' ability to generalize NADs in this task. We also note that generalization

29

524   of the above two NADs was not observed for the only Y element located in a corner of the

525   matrix (Y_1), possibly because of its greater saliency, or eccentricity in the matrix, compared to

526   the other Y locations which were more central. Taken together, these findings suggest that the

527   baboons can generalize NADs over novel locations, but that this ability may be hindered by

528   spatial factors.

529

530   **General discussion**

531   In Experiment 1, humans and baboons responded faster to the last target in the NAD

532   compared to the control sequences. Importantly, participants did not merely learn that the

533   sequences always finished by one of two locations (with a probability of .5 for each), but

534   processed the conditional probabilities of the third location, given the first location. This

535   result was replicated in baboons in Experiment 2. These data overall provide evidence of

536   baboons' ability to learn NADs. However, while humans presented a significant difference in

537   response times between the two conditions within the first exposure block, this difference

538   appeared after approximately 400 presentations of each NAD in baboons. This difference in

539   learning speed differs from what has been found for adjacent dependencies (Minier et al.,

540   2016; Rey et al., in press). In these earlier studies, baboons and humans learned at similar

541   speed (in less than 200 trials) the predictive relationships between two locations presented at

542   the end of three different triplets. This suggests that learning of NADs is more difficult for

543   baboons than humans, while these two species seem equally fast for learning adjacent

544   dependencies. Importantly, the amount of exposure required by the baboons to extract NADs

545   in the current experiment is in the same range as previously found with habituation-

546   dishabituation paradigms in other primate species (e.g., 672 presentations in Newport et al.,

547   2004). This amount may serve as a reference for designing future experiments investigating

548   related topics.

549   For two NADs we found that when presented with novel intervening locations the

550   baboons were slower to respond when the sequences were inconsistent with the learned NADs

551   than when the dependencies were preserved. To our knowledge, this result is the first

552   demonstration in a non-human species of a successful generalization of sequentially non-

553   adjacent dependencies over novel intervening elements. The current results therefore extend

554   to the case of sequential NADs Sonnweber et al.'s findings (2015), in which generalization of

555   NADs was tested between visual shapes presented concurrently. Along with previous findings

556   in humans (e.g., Frost & Monaghan, 2016), our study indicates that this ability is shared

557   across different primate species. However, successful generalization of the NADs in the

558   current SRT study appeared to be modulated by spatial features of the stimuli. Indeed, the

559   baboons only generalized the two NADs whose predictive target was located in a corner of

560   the matrix and generalization was not observed for one of the novel intervening location

561   which was also located in a corner. These data present some similarities with previous

562   findings in the auditory domain. For instance, Endress, Scholl, and Mehler (2005, see also

563   Peña et al., 2002) reported that human participants could properly extract some specific

564   regularities from auditory sequences only when the relevant stimuli were presented at an edge

565   of the sequences. Moreover, several studies manipulating the phonological parameters of the

566   stimuli suggested that the relative saliency of the non-adjacent and intervening elements

567   impacts the extraction of NADs (Creel et al., 2004; Onnis et al., 2005; Toro & Trobalón,

568   2005). Taken together, these findings suggest (1) that perceptual factors constrain the

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

570   domains.

571   The exact nature of the processes involved in NADs extraction and recognition is still

572   debated in the literature, and at least three theories are proposed. According to a first theory,

573   the mechanisms that support NADs extraction are the same associative mechanisms as those

574   allowing the learning of adjacent dependencies (Perruchet, Tyler, Galland, & Peereman,

575   2004). This single-process account proposes that the joint attentional processing of two

576   elements is a necessary and sufficient condition to associate them, even when these are distant

577   in space or time (Pacton & Perruchet, 2008). A second and complementary theory proposes

578   that the extraction of adjacent and non-adjacent dependencies involves similar statistical

579   learning mechanisms, but emphasize the role played by working memory for NADs

580   extraction (Conway, Deocampo and Smith, 2016; Wilson et al., 2017). According to this dual-

581   system theory, the detection of adjacent relationships would be sustained by modality-specific

582   processes involving perceptual-motor neural networks (Conway et al., 2016), or ventral

583   regions of the frontal cortex (Wilson, Kikuchi, et al., 2015), while the extraction of non-

584   adjacent relationships would additionally recruit more dorsal areas (Wilson et al., 2017).

585   According to a last hypothesis (Endress & Bonatti, 2007; Endress, Nespor, & Mehler, 2009;

586   Peña et al., 2002), humans and possibly other animals (Endress, Carden, Versace, & Hauser,

587   2010) may be endowed with some rule-based mechanisms dedicated to fast learning of

588   "classes" of elements occurring at the edges of sequences (e.g., here, learning that A_1 and A_2

589   always occur in the first position), as well as statistical learning mechanisms for the

590   computation of transitional probabilities (e.g. $A_1\_B_1$). The current experiment did not aim at

591 assessing the edge-based positional learning hypothesis but examined the extraction time

592 course of those transitional probabilities.

593   In humans, learning occurred early in the experiment, and was strengthened

594 throughout the entire exposure phase. In baboons, learning of the NADs occurred later in the

595 experiment, but a similar progressive emergence was observed. In previous SRT experiments,

596 we found that the learning of adjacent dependencies in triplets (Minier et al., 2016), or pairs of

597 shapes (Fagot, Malassis, & Medam, 2017) gave rise to similar progressive (albeit earlier)

598 decrease in response times throughout the exposure to these regularities. These data, along

599 with previous findings (Frost & Monaghan, 2016; Romberg & Saffran, 2013; Vuong et al.,

600 2016), suggest that adjacent and non-adjacent dependencies might be extracted by similar

601 statistical learning mechanisms. However, they do not allow disentangling between the single-

602 process and the dual-system proposals described above. Indeed, the difference in speed

603 learning that is observed between baboons and humans might reflect an involvement of

604 working memory in NADs extraction, as working memory capacity differ between these

605 species (e.g., Fagot & De Lillo, 2011). However, addressing this question would require

606 additional behavioral and brain imaging studies. Regarding behavioral studies, correlational

607 designs in humans and non-human primates would allow examining individual differences in

608 statistical learning ability across different types of dependencies (i.e., adjacent and non-

609 adjacent, Siegelman, Bogaerts, Christiansen, & Frost, 2017; Siegelman & Frost, 2015), as

610 well as exploring the relationships between working memory capacity and NADs extraction.

611 Comparative neuroimaging studies have already suggested that adjacent dependencies are

612 processed by evolutionary conserved brain structures, including the frontal operculum and

613 anterior insula (Wilson, Kikuchi, et al., 2015). However, further researches are also needed in

614   this domain to discover whether NADs processing is sustained by similar or different neural

615   substrates in human and non-human primates (Milne et al., 2016; Wilson et al., 2017).

616   Another important question is the domain generality of the mechanisms involved

617   during the learning of NADs. For instance, are the extraction of NADs in the current SRT task

618   and the extraction of NADs in natural languages (such as in *is* read*ing*) supported by the same

619   learning mechanisms? Evidences accumulated so far reveal a contrasted pattern (Frost,

620   Armstrong, Siegelman, & Christiansen, 2015). On the one hand, individuals with specific

621   language impairment have been found to exhibit poor performance in auditory statistical

622   learning tasks (e.g., Evans, Saffran, & Robe-Torres, 2009) as well as in SRT tasks (Lum,

623   Conti-Ramsden, Morgan, & Ullman, 2014), compared with typically-developing individuals.

624   They also express greater difficulties in processing NADs in an AGL-SRT task (Hsu,

625   Tomblin, & Christiansen, 2014). In another study, Misyak, Christiansen, and Tomblin (2010)

626   found a positive correlation between typically-developing individuals' ability to process long-

627   distance dependencies from a natural language and NADs in an AGL-SRT task. On the other

628   hand, statistical learning appears to be subject to modality and stimulus specificity (e.g.,

629   Conway & Christiansen, 2005, 2006). Siegelman and Frost (2015) reported for instance that

630   the performance of their participants did not correlate between any of four statistical learning

631   tasks they have used (auditory-verbal, auditory-nonverbal, visual, and SRT tasks). These

632   results suggest the existence of modality and stimulus specific constraints on statistical

633   learning mechanisms, which prevent drawing strong conclusions regarding the similarity

634   between NADs extraction in natural languages and in our visuo-spatial task.

635   From an evolutionary standpoint, the ability to keep track of predictive relationships

636   between non-adjacent events might present various functional advantages, for instance for

34

637 planning complex behavioral or motor sequences, or for monitoring social interactions and

638 their outcomes (Sonnweber et al., 2015). In baboons, this process may allow for instance to

639 detect that a threat behavior from an individual A is followed by a scream from an individual

640 B, even if other individuals emit unrelated vocalizations in between. One general advantage of

641 the learning of NADs might also be a reduction in memory demands. For instance in song

642 birds, admitting an optional or variable song element between two others (thereby being non-

643 adjacent) could limit the number of pairwise transitions needed to be memorized (Petkov &

644 Wilson, 2012). In our experiments, the extraction of two NADs may be cognitively less

645 demanding than the learning and storage of ten sequences.

646

647 **Acknowledgments**

648

655

656
657

658   **References**

659

660   Cate, C. ten, & Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals:

661   natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal*

662   *Society of London B: Biological Sciences*, *367*(1598), 1984–1994.

663   https://doi.org/10.1098/rstb.2012.0055

664   Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences.

665   *Journal of Experimental Psychology. General*, *120*(3), 235–253.

666   Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates.

667   *Trends in Cognitive Sciences*, *5*(12), 539–546. https://doi.org/10.1016/S1364-6613(00)01800-

668   3

669   Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of

670   tactile, visual, and auditory sequences. *Journal of Experimental Psychology. Learning,*

671   *Memory, and Cognition*, *31*(1), 24–39. https://doi.org/10.1037/0278-7393.31.1.24

672   Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between

673   modalities: pitting abstract against stimulus-specific representations. *Psychological Science*,

674   *17*(10), 905–912. https://doi.org/10.1111/j.1467-9280.2006.01801.x

675   Conway, C. M., Deocampo, J., A., & Smith, G..N.L. (2016). Multiple Routes to Implicit

676   Statistical Learning? A Dual-System Perspective. Abstracts of the Psychonomic Society, 57th

677   annual meeting, Boston.

678 Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: statistical learning of

679 nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology. Learning,*

680 *Memory, and Cognition*, *30*(5), 1119–1130. https://doi.org/10.1037/0278-7393.30.5.1119

681 de la Mora, D. M., & Toro, J. M. (2013). Rule learning over consonants and vowels in a non-

682 human animal. *Cognition*, *126*(2), 307–312. https://doi.org/10.1016/j.cognition.2012.09.015

683 Endress, A. D. (2010). Learning melodies from non-adjacent tones. *Acta Psychologica*,

684 *135*(2), 182–190. https://doi.org/10.1016/j.actpsy.2010.06.005

685 Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually

686 continuous speech stream. *Cognition*, *105*(2), 247–299.

687 https://doi.org/10.1016/j.cognition.2006.09.010

688 Endress, A. D., Carden, S., Versace, E., & Hauser, M. D. (2010). The apes' edge: positional

689 learning in chimpanzees and humans. *Animal Cognition*, *13*(3), 483–495.

690 https://doi.org/10.1007/s10071-009-0299-8

691 Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on

692 language acquisition. *Trends in Cognitive Sciences*, *13*(8), 348–353.

693 https://doi.org/10.1016/j.tics.2009.05.005

694 Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of

695 algebraic rules. *Journal of Experimental Psychology. General*, *134*(3), 406–419.

696 https://doi.org/10.1037/0096-3445.134.3.406

697   Endress, A. D., & Wood, J. N. (2011). From movements to actions: two mechanisms for

698   learning action sequences. *Cognitive Psychology*, *63*(3), 141–171.

699   https://doi.org/10.1016/j.cogpsych.2011.07.001

700   Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical Learning in Children With

701   Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *52*(2),

702   321–335. https://doi.org/10.1044/1092-4388(2009/07-0189)

703   Fagot, J., & Bonté, E. (2010). Automated testing of cognitive performance in monkeys: use of

704   a battery of computerized test systems by a troop of semi-free-ranging baboons (Papio papio).

705   *Behavior Research Methods*, *42*(2), 507–516. https://doi.org/10.3758/BRM.42.2.507

706   Fagot, J., & De Lillo, C. (2011). A comparative study of working memory: Immediate serial

707   spatial recall in baboons (Papio papio) and humans. *Neuropsychologia*, *49*(14), 3870–3880.

708   https://doi.org/10.1016/j.neuropsychologia.2011.10.003

709   Fagot, J., Malassis, R., & Medam, T. (2017). The processing of positional information in a

710   two-item sequence limits the emergence of symmetry in baboons (Papio papio), but not in

711   humans (Homo sapiens). *Learning & Behavior*. https://doi.org/10.3758/s13420-017-0290-1

712   Fagot, J., & Paleressompoulle, D. (2009). Automatic testing of cognitive performance in

713   baboons maintained in social groups. *Behavior Research Methods*, *41*(2), 396–404.

714   https://doi.org/10.3758/BRM.41.2.396

715   Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial

716   structures from visual scenes. *Psychological Science*, *12*(6), 499–504.

717   Friederici, A. D. (2004). Processing local transitions versus long-distance syntactic

718   hierarchies. *Trends in Cognitive Sciences*, *8*(6), 245–247.

719   https://doi.org/10.1016/j.tics.2004.04.013

720   Froehlich, A. L., Herbranson, W. T., Loper, J. D., Wood, D. M., & Shimp, C. P. (2004).

721   Anticipating by Pigeons Depends on Local Statistical Information in a Serial Response Time

722   Task. *Journal of Experimental Psychology: General*, *133*(1), 31–45.

723   https://doi.org/10.1037/0096-3445.133.1.31

724   Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality

725   versus modality specificity: the paradox of statistical learning. *Trends in Cognitive Sciences*,

726   *19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010

727   Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of

728   non-adjacent dependencies from continuous speech. *Cognition, 147*, 70–74.

729   https://doi.org/10.1016/j.cognition.2015.11.010

730   Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and

731   nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review*,

732   *16*(3), 486–490. https://doi.org/10.3758/PBR.16.3.486

733   Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*,

734   *13*(5), 431–436.

735   Gómez, R., & Maye, J. (2005). The Developmental Trajectory of Nonadjacent Dependency

736   Learning. *Infancy, 7*(2), 183–206. https://doi.org/10.1207/s15327078in0702_4

737   Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a

738   non-human primate: statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53–B64.

739   https://doi.org/10.1016/S0010-0277(00)00132-3

740   Heijningen, C. A. A. van, Visser, J. de, Zuidema, W., & Cate, C. ten. (2009). Simple rules can

741   explain discrimination of putative recursive syntactic structures by a songbird species.

742   *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0908113106

743   Hsu, H. J., Tomblin, J. B., & Christiansen, M. H. (2014). Impaired statistical learning of non-

744   adjacent dependencies in adolescents with specific language impairment. *Frontiers in*

745   *Psychology*, *5*, 175. https://doi.org/10.3389/fpsyg.2014.00175

746   Kovács, Á. M., & Mehler, J. (2009). Flexible Learning of Multiple Speech Structures in

747   Bilingual Infants. *Science*, *325*(5940), 611–612. https://doi.org/10.1126/science.1173947

748   Lum, J. A. G., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. T. (2014). Procedural

749   learning deficits in specific language impairment (SLI): A meta-analysis of serial reaction

750   time task performance. *Cortex; a Journal Devoted to the Study of the Nervous System and*

751   *Behavior*, *51*(100), 1–10. https://doi.org/10.1016/j.cortex.2013.10.011

752   Marchetto, E., & Bonatti, L. L. (2013). Words and possible words in early language

753   acquisition. *Cognitive Psychology*, *67*(3), 130–150.

754   https://doi.org/10.1016/j.cogpsych.2013.08.001

755   Marchetto, E., & Bonatti, L. L. (2015). Finding words and word structure in artificial speech:

756   the development of infants' sensitivity to morphosyntactic regularities. *Journal of Child*

757   *Language*, *42*(4), 873–902. https://doi.org/10.1017/S0305000914000452

758   Milne, A. E., Mueller, J. L., Männel, C., Attaheri, A., Friederici, A. D., & Petkov, C. I.

759   (2016). Evolutionary origins of non-adjacent sequence processing in primate brain potentials.

760   *Scientific Reports*, *6*, 36259. https://doi.org/10.1038/srep36259

761   Minier, L., Fagot, J., & Rey, A. (2016). The Temporal Dynamics of Regularity Extraction in

762   Non-Human Primates. *Cognitive Science*, *40*(4), 1019–1030.

763   https://doi.org/10.1111/cogs.12279

764   Misyak, J. B., Christiansen, M. H., & Bruce Tomblin, J. (2010). Sequential Expectations: The

765   Role of Prediction-Based Learning in Language. *Topics in Cognitive Science*, *2*(1), 138–153.

766   https://doi.org/10.1111/j.1756-8765.2009.01072.x

767   Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-

768   adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162.

769   Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II.

770   Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive*

771   *Psychology*, *49*(2), 85–117. https://doi.org/10.1016/j.cogpsych.2003.12.002

772   Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from

773   performance measures. *Cognitive Psychology*, *19*(1), 1–32. https://doi.org/10.1016/0010-

774   0285(87)90002-8

775   Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2005). Variability is the spice of

776   learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies

777   [Proceedings paper].

778   Onnis, Luca, Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts

779   segmentation in online speech processing. *Journal of Memory and Language*, *53*(2), 225–237.

780   https://doi.org/10.1016/j.jml.2005.02.011

781   Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and

782   nonadjacent dependency learning. *Journal of Experimental Psychology. Learning, Memory,*

783   *and Cognition*, *34*(1), 80–96. https://doi.org/10.1037/0278-7393.34.1.80

784   Pacton, S., Sobaco, A., & Perruchet, P. (2015). Is an attention-based associative account of

785   adjacent and nonadjacent dependency learning valid? *Acta Psychologica*, *157*, 195–199.

786   https://doi.org/10.1016/j.actpsy.2015.03.002

787   Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in

788   speech processing. *Science (New York, N.Y.)*, *298*(5593), 604–607.

789   https://doi.org/10.1126/science.1072901

790   Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent

791   dependencies: no need for algebraic-like computations. *Journal of Experimental Psychology.*

792   *General*, *133*(4), 573–583. https://doi.org/10.1037/0096-3445.133.4.573

793   Petkov, C. I., & Wilson, B. (2012). On the pursuit of the brain network for proto-syntactic

794   learning in non-human primates: conceptual issues and neurobiological hypotheses.

795   *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 2077–

796   2088. https://doi.org/10.1098/rstb.2012.0073

797   Ravignani, A., & Sonnweber, R. (2017). Chimpanzees process structural isomorphisms across

798   sensory modalities. *Cognition*, *161*, 74–79. https://doi.org/10.1016/j.cognition.2017.01.005

799 Ravignani, A., Sonnweber, R.-S., Stobbe, N., & Fitch, W. T. (2013). Action at a distance:

800 dependency sensitivity in a New World primate. *Biology Letters*, *9*(6), 20130852.

801 https://doi.org/10.1098/rsbl.2013.0852

802 Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and*

803 *Verbal Behavior*, *6*(6), 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

804 Rey, A., Minier, L., Malassis, R., Bogaerts, L., & Fagot, J. (in press). Regularity extraction

805 across species: associative learning mechanisms shared by human and non-human primates.

806 *Topics in Cognitive Science*.

807 Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley*

808 *Interdisciplinary Reviews. Cognitive Science*, *1*(6), 906–914. https://doi.org/10.1002/wcs.78

809 Romberg, A. R., & Saffran, J. R. (2013). All Together Now: Concurrent Learning of Multiple

810 Structures in an Artificial Language. *Cognitive Science*, *37*(7), 1290–1320.

811 https://doi.org/10.1111/cogs.12050

812 Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old

813 infants. *Science (New York, N.Y.)*, *274*(5294), 1926–1928.

814 Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of

815 individual differences in statistical learning. *Philosophical Transactions of the Royal Society*

816 *of London. Series B, Biological Sciences*, *372*(1711). https://doi.org/10.1098/rstb.2016.0059

817 Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical

818 perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105–120.

819 https://doi.org/10.1016/j.jml.2015.02.001

820   Sonnweber, R., Ravignani, A., & Fitch, W. T. (2015). Non-adjacent visual dependency

821   learning in chimpanzees. *Animal Cognition*, *18*(3), 733–745. https://doi.org/10.1007/s10071-

822   015-0840-x

823   Takahasi, M., Yamada, H., & Okanoya, K. (2010). Statistical and Prosodic Cues for Song

824   Segmentation Learning by Bengalese Finches (Lonchura striata var. domestica). *Ethology*,

825   *116*(6), 481–489. https://doi.org/10.1111/j.1439-0310.2010.01772.x

826   Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. L. (2008). Finding words and rules in a

827   speech stream: functional differences between vowels and consonants. *Psychological Science*,

828   *19*(2), 137–144. https://doi.org/10.1111/j.1467-9280.2008.02059.x

829   Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a

830   rodent. *Perception & Psychophysics*, *67*(5), 867–875. https://doi.org/10.3758/BF03193539

831   Vuong, L. C., Meyer, A. S., & Christiansen, M. H. (2016). Concurrent Statistical Learning of

832   Adjacent and Nonadjacent Dependencies. *Language Learning*, *66*(1), 8–30.

833   https://doi.org/10.1111/lang.12137

834   Wilson, B., Kikuchi, Y., Sun, L., Hunter, D., Dick, F., Smith, K., … Petkov, C. I. (2015).

835   Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in

836   macaques and humans. *Nature Communications*, *6*, 8901.

837   https://doi.org/10.1038/ncomms9901

838   Wilson, B., Marslen-Wilson, W. D., & Petkov, C. I. (2017). Conserved Sequence Processing

839   in Primate Frontal Cortex. *Trends in Neurosciences*, *0*(0).

840   https://doi.org/10.1016/j.tins.2016.11.004

841   Wilson, B., Slater, H., Kikuchi, Y., Milne, A. E., Marslen-Wilson, W. D., Smith, K., &

842   Petkov, C. I. (2013). Auditory Artificial Grammar Learning in Macaque and Marmoset

843   Monkeys. *The Journal of Neuroscience*, *33*(48), 18825–18835.

844   https://doi.org/10.1523/JNEUROSCI.2414-13.2013

845   Wilson, B., Smith, K., & Petkov, C. I. (2015). Mixed-complexity artificial grammar learning

846   in humans and macaque monkeys: evaluating learning strategies. *The European Journal of*

847   *Neuroscience*, *41*(5), 568–578. https://doi.org/10.1111/ejn.12834

848