AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Learning, Memory, and Cognition

ISSN: 0278-7393

https://doi.org/10.1037/xlm0001182

Formal Language Hierarchy Reflects Different Levels of Cognitive Complexity

Susana Silva[1], Filomena Inácio[2, 3], Daniel Rocha e Sousa[4], Nuno Gaspar[1], Vasiliki Folia[5],
and Karl Magnus Petersson[2, 5]
[1] Center for Psychology at University of Porto, Faculty of Psychology and Educational Sciences, Psychology Department, University of Porto
[2] Center for Research in Health Technologies and Information Systems (CINTESIS-UAlg), University of Algarve
[3] Faculty of Human and Social Sciences, Psychology Department, University of Algarve
[4] Faculty of Human and Social Sciences, Social Sciences department, Fernando Pessoa University
[5] Max Planck Institute for Psycholinguistics, Neurobiology of Language department, Nijmegen, the Netherlands

Formal language hierarchy describes levels of increasing syntactic complexity (adjacent dependencies, nonadjacent nested, nonadjacent crossed) of which the transcription into a hierarchy of cognitive complexity remains under debate. The cognitive foundations of formal language hierarchy have been contradicted by two types of evidence: First, adjacent dependencies are not easier to learn compared to nonadjacent; second, crossed nonadjacent dependencies may be easier than nested. However, studies providing these findings may have engaged confounds: Repetition monitoring strategies may have accounted for participants' high performance in nonadjacent dependencies, and linguistic experience may have accounted for the advantage of crossed dependencies. We conducted two artificial grammar learning experiments where we addressed these confounds by manipulating reliance on repetition monitoring and by testing participants inexperienced with crossed dependencies. Results showed relevant differences in learning adjacent versus nonadjacent dependencies and advantages of nested over crossed, suggesting that formal language hierarchy may indeed translate into a hierarchy of cognitive complexity.

Keywords: nonadjacent dependencies, eye tracking, implicit learning, syntax

Human languages display various types of syntactic dependencies between words or other constituents in sentences, and the resolution of these dependencies is crucial to generate adequate sentence-level interpretations. For example, processing the sentence "The mouse the cat chased was small" requires one to connect "the mouse" (Constituent A1) with "was small" (B1) and "the cat" (A2) with "chased" (B2). A basic distinction is often made between *adjacent* dependencies, where dependent constituents from different categories (A vs. B) are contiguous in space or time, and *nonadjacent* ones, where constituents are separated by intervening material (Wilson et al., 2020). In the example above, the A1B1 dependency was nonadjacent in *A1*A2 | B2*B1*. Suppressing A2B2 ("The mouse was small") would turn A1B1 into an adjacent dependency. Nonadjacency does not necessarily occur in a context of multiple AB dependencies as in the example above (A1A2|B2B1): It may be instantiated by single AB dependencies like AxB, where x is the intervening material (Petkov & Ten Cate, 2020; Wilson et al., 2020). When multiple nonadjacent dependencies coexist, at least two different types of organizations may be found: The first one corresponds to *nested* or center-embedded dependencies (as in the example above, A1A2|B2B1). It engages a mirror-like organization (12 is followed by 21), where dependent constituents A and B increase the distance between them as new dependencies are added (e.g., A1A2A3|B3B2B1). The other, less pervasive, type corresponds to *crossed* or interleaved dependencies. Here, the organization of As and Bs is copy-like (A1A2|B1B2, 12 followed by 12), with As and Bs keeping a constant distance between them whatever the number of dependencies (Joshi, 1990). Crossed dependencies—at least those between noun phrases and verbs—seem to be rare, and Swiss German and Dutch are the only languages known so far to have such dependencies (Kaan & Vasić, 2004; Stabler, 2004). In the present study, we investigated the learning of crossed versus nested nonadjacent dependencies and how this might relate to what is known about learning adjacent dependencies.

Adjacent, nested, and crossed dependencies have been associated to three different levels of complexity initially defined by

---

Susana Silva https://orcid.org/0000-0003-2240-1828
Karl Magnus Petersson https://orcid.org/0000-0002-8245-0392
Vasiliki Folia is now at Lab of Cognitive Neuroscience, School of Psychology, Aristotle University of Thessaloniki.

This work was supported by Fundação para a Ciência e a Tecnologia (PTDC/P SIPCO/110734/2009; PTDC/PSI-GER/32602/2017; CPUP UIDB/00050/2020; and PTDC/PSIGER/5845/2020). We are grateful to our participants. The stimulus sets, data sets, and code for analysis are available at https://osf.io/5n7ag/?view_only=bdf728d1bcf44cc283401cb84191cc91. The authors declare no conflict of interest.

Correspondence concerning this article should be addressed to Susana Silva, Center for Psychology at University of Porto, Faculty of Psychology and Educational Sciences, Psychology Department, University of Porto, Rua Alfredo Allen, s/n, 4200-135 Porto, Portugal. Email: susanamsilva@fpce.up.pt

Chomsky (1956, 1965) in his hierarchy of formal languages (see Figure 1): While adjacent dependencies can be realized in regular or finite-state grammars (the simplest level in formal language hierarchy), unbounded nested ones require context-free grammars (one level above), and unbounded crossed ones require context-sensitive grammars, which are more complex than nested (de Vries et al., 2011; Folia & Petersson, 2014; Jäger & Rogers, 2012; Öttl et al., 2015; Rohrmeier et al., 2012; Westphal-Fitch et al., 2018). Please note that a finite number of certain types of unbounded long-distance dependencies can be represented by regular grammars. Crucial work on formal language hierarchy (Jäger & Rogers, 2012; Joshi, 1985; Stabler, 2004) indicated that a subset of context-sensitive grammars should be enough to accommodate the characteristics of crossed dependencies, and one should refer to mildly context-sensitive grammars instead. In any case, the point is that crossed dependencies are expected to be formally more complex than nested ones.

Adjacent, Nonadjacent Nested, and Nonadjacent Crossed Dependencies: A Cognitive Hierarchy?

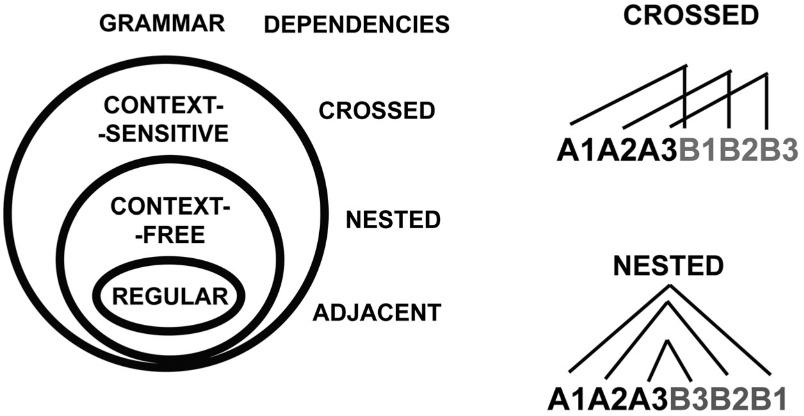
Formal language hierarchy is not specifically about human language but about different levels of computational power required by increasingly complex formal languages. A question that makes formal language hierarchy interesting to cognitive science—and that remains open—is whether these different levels capture discontinuities in human computational power for language (Fitch & Friederici, 2012), that is, (a) whether adjacent dependencies are unquestionably easier to process than nested and (b) whether nested are easier than crossed. In the present study, we addressed these two open questions, equating computational power (ease of processing) with *implicit learnability,* that is, the ability to learn without intention and without awareness of what has been learned (Ziori et al., 2014).

To investigate the implicit learnability of different types of dependencies, we used a visual artificial grammar learning (AGL) paradigm (Reber, 1967). AGL paradigms mimic the implicit learning process typical of natural syntax acquisition while controlling for participants' previous knowledge of the stimulus material (Fitch et al., 2012; but see Fitch & Friederici, 2012; Ojima & Okanoya, 2014). In the training phase of an AGL experiment, participants are exposed to examples of items (sequences of letters, sounds, etc.) formed according to a set of grammatical rules. Participants remain unaware of the existence of such rules until the moment they are asked to discriminate between grammatical and nongrammatical sequences in a subsequent classification phase. We used a learning design (Petersson et al., 1999a, 1999b) that employs an additional baseline classification test before training so that learning could be measured as the within-subjects difference between pre- and posttraining discrimination and potential pretraining biases could thus be controlled for. In this case, pretraining discrimination cannot be measured with questions relating to grammaticality (grammatical vs. nongrammatical, correct vs. incorrect). Therefore, pre- and posttraining preference classification tests are used instead, where participants state whether they like or dislike each sequence. Preference classification tests provide an accurate estimate of acquired knowledge (liking what one knows) while minimizing explicit influences (Forkstam et al., 2008; Manza & Bornstein, 1995; Newell & Bright, 2001). As for using a visual grammar, it matters to state that AGL paradigms were conceived to capture learning of syntax-related, amodal information. Visual AGL paradigms have been used along with auditory ones to mimic natural syntax acquisition (Silva et al., 2018; Stobbe et al., 2012; Westphal-Fitch et al., 2018), despite the fact that the first language is typically acquired via auditory learning. These visual paradigms seem to have succeeded in that language skills correlate with AGL outcomes when grammars are visual (Christiansen et al., 2010).

Current answers for the two open questions addressed here, namely (a) whether adjacent dependencies are unquestionably easier to process than nonadjacent and (b) whether nested are easier than crossed, include negative answers that challenge the idea of increased levels of complexity in these terms. Research findings indicate that (a) participants may show equivalent performance in nonadjacent and adjacent dependencies provided they are given enough time to learn nonadjacent ones (a "matter of time") and (b) crossed dependencies may be easier to learn than nested. However,

Figure 1

Formal Language Hierarchy (Left) and Structure of Nested Versus Crossed Nonadjacent Dependencies (Right)



these findings may reflect experimental confounds, as we outline below.

Adjacent Versus Nonadjacent Dependencies: Just a Matter of Time?

The first open question we addressed here—the extent to which multiple nonadjacent dependencies represent an increased learning challenge compared to adjacent ones—is not new, and the dominant answer points to relevant differences between the two dependency types based on human and animal studies. The hypothesis that nonadjacent dependencies may require more sophisticated cognitive resources goes back at least to 2004, when Fitch and Hauser (2004) showed that both humans and nonhuman primates could learn adjacent dependencies, but only humans were able to learn nested ones. Later research on macaque monkeys and songbirds pointing to encoding ability of nested structures (Abe & Watanabe, 2011; Gentner et al., 2006) has been criticized (Beckers et al., 2017; Fitch & Friederici, 2012; ten Cate & Okanoya, 2012; van Heijningen et al., 2009), while recently Jiang et al. (2018) could prove that nonhuman primates could learn supraregular structures. Thus, research on other species is not conclusive and remains a fascinating question (Dehaene et al., 2015). In humans, evidence that adjacent and nonadjacent dependencies recruit different brain substrates has gained strength (Calmus et al., 2020; Friederici et al., 2006; Uddén et al., 2020). Nonadjacency has been shown to be acquired later than adjacency in development (Gómez & Maye, 2005; Teinonen et al., 2009), and direct comparisons between adjacent and nonadjacent dependencies showed increased learning for adjacent ones (Friederici et al., 2006; Öttl et al., 2017). One detour from this scenario was made by Uddén and colleagues (Uddén et al., 2009, 2012), who showed that multiple nonadjacent dependencies could be learned as efficiently as adjacent dependencies, provided that participants were given more time to learn (9 days instead of 5; see also Uddén et al., 2017). The authors compared learning outcomes in their 9-day study on nonadjacent (crossed and nested) dependencies with outcomes from previous 5-day studies on adjacent dependencies using the same paradigm (Folia et al., 2008; Forkstam et al., 2006, 2008) and found a similar pattern of results. Thus, while most studies had pointed to fundamental (qualitative) differences in adjacent versus nonadjacent dependency learning, Uddén et al. (2009, 2012) indicated that these might be merely quantitative ("a matter of time"). Nonadjacent dependencies would not constitute a significantly increased learning challenge compared to adjacent, speaking against the cognitive translation of one aspect of formal language hierarchy.

What remains to be demonstrated is that participants in Uddén and colleagues' (2009, 2012) study acquired deep structural knowledge, and successful discrimination could not be explained by one specific surface-related strategy—repetition monitoring. The use of surface-related strategies for discriminating grammatical and nongrammatical sequences containing nonadjacent dependencies in AGL tasks has been vastly documented. Relying on memory for chunks (subsequences of adjacent items—letters or sounds) is one surface-related strategy (Poletiek & Lai, 2012). Other strategies, potentially more effective than chunking (Ling et al., 2016), relate to the overall configuration of As and Bs and take advantage of unbalanced test stimulus structures. For instance, Fitch and Hauser (2004) tested whether participants familiarized with nested sequences (A1A2A3|B3B2B1) could distinguish these from nongrammatical sequences of type A1B1A2B2A3B3 in the classification phase, with As differing from Bs in pitch. As shown in subsequent research (Hochmann et al., 2008; Perruchet & Rey, 2005), participants seem able to make this distinction without ever acquiring the agreement structure (A1 pairs with B1, etc.) that substantiates the rule set: Instead, they may simply rely on the pitch-based distinction between multiple As followed by multiple Bs (perceived as grammatical) and single As alternating with single Bs (perceived as nongrammatical). Friederici et al. (2006) exposed participants to nested sequences with an equal number of As and Bs (e.g., three As and three Bs) but tested them with nongrammatical sequences containing four As and two Bs. Again, as shown by de Vries et al. (2008), participants would not need to know the agreement rules to discriminate grammatical from nongrammatical sequences; they would just need to count. De Vries and colleagues (2008) went further and showed that repetition monitoring may be another shortcut for discrimination: When participants are presented with grammatical sequences containing letter repetitions, all repetitions within the A part are followed by repetitions in the B part (e.g., *A1A1*A2|B2*B1B1* in nested grammars). If nongrammatical sequences in the test phase do not preserve this symmetry in repetition structure (e.g., *A1A1*A2|B2*B1B2*), participants can easily spot nongrammatical sequences without deep structural knowledge. Uddén et al. (2009, 2012) controlled for the chunk shortcut to learning using measures of associative chunk strength (ACS; Bailey & Pothos, 2008) and used the same number of As and Bs in nongrammatical test sequences (for two as there would be two Bs, etc.). However, they did not balance symmetry in repetition structure across grammatical (symmetric) and nongrammatical sequences (nonsymmetric). Therefore, it is yet to be demonstrated that participants used the 9-day exposure period to acquire deep structural knowledge on nested and crossed dependencies instead of spotting nongrammatical sequences based on violations of symmetry in repetition structure.

The first goal of the present study was to determine whether Uddén et al.'s (2009, 2012) findings (nonadjacent dependency learning is just "a matter of time") replicate when controlling for differences between grammatical and nongrammatical sequences concerning repetition structure. As in Uddén et al. (2009, 2012), we ran a 9-day study on nested versus crossed dependencies, the results of which were compared to a previous 5-day behavioral and eye-tracking study of ours on adjacent dependencies (Silva et al., 2017) showing behavioral and eye-tracking signatures of successful implicit learning. A pattern of results similar to Silva et al. (2017) would strengthen the idea that nonadjacent dependencies do not represent a substantially increased learning challenge compared to adjacent ones, in line with Uddén et al. (2009, 2012). In contrast, unsuccessful learning of nonadjacent dependencies in the current study would be consistent with the possibility that participants' high discrimination performance in Uddén et al. (2009, 2012) was due to surface-related strategies such as repetition monitoring, thus weakening the hypothesis that adjacent and nonadjacent dependencies may be learned with similar levels of success. Given the importance of controlling for repetition structure across grammatical and nongrammatical sequences, this was a priority criterion in stimulus generation.

Are Crossed Dependencies Easier to Learn Than Nested Ones?

Concerning the second open question—relative implicit learnability of nested versus crossed nonadjacent types—available findings are mixed. According to formal language hierarchy, nested dependencies should be easier to acquire. Although there is empirical evidence in favor of this (Nakamura & Miyamoto, 2006), opposite findings abound: The study of Bach et al. (1986) compared the performance of Dutch versus German participants decoding nonadjacent dependencies typical of their native languages (crossed for Dutch, nested for German) and found performance advantages for Dutch speakers. They concluded that crossed dependencies require less computational resources than nested ones. Theoretical support for this came later with Gibson's (1998) syntactic prediction locality theory, explaining why more distant dependencies—such as those happening in nested grammars—may be more demanding. A few AGL studies have supported the increased learnability of crossed compared to nested dependencies (de Vries et al., 2012; Uddén et al., 2009, 2012), at least when sequences contain more than two dependent pairs (AAA|BBB; but see Bader, 2017). However, these results may have reflected a moderating influence of linguistic experience on cognitive complexity—a role that has been highlighted by usage-based accounts of natural-language recursion (Christiansen & Chater, 2015). Specifically, Uddén and colleagues (2009, 2012) recruited Dutch participants, and linguistic experience with crossed dependencies could have been responsible for the advantage of crossed over nested grammars (de Vries et al., 2011). Strengthening this possibility, de Vries et al. (2012) found advantages in reaction times for nested grammars in German-speaking (nested-based language) participants when compared to Dutch ones. Further studies with German participants (Öttl et al., 2015, 2017) showed a different pattern in that nested and crossed dependencies were equally difficult to learn and process. However, Öttl and colleagues did not use an additional baseline classification test before training, thus leaving open a complete characterization of learning outcomes and subsequent considerations on the role of linguistic experience.

The second goal of this study was to compare the implicit learnability of nested versus crossed nonadjacent dependencies without the limitations of previous studies. Overcoming the limitations of studies with Dutch participants, we recruited Portuguese-speaking participants, who have little or no experience with crossed dependencies between different words or constituents (but please see Martins, 2006, and Piechnik, 2015, for reduplication phenomena in Portuguese). Overcoming the limitations of Öttl and colleagues' (2015, 2017) studies with German participants (also unexperienced with crossed dependencies), we used a learning approach that included pre- and posttraining preference classification tests. If nested dependencies proved to be easier to learn than crossed ones, this would strengthen the idea that formal language hierarchy is compatible with the cognitive architecture of humans.

The Purpose of Eye-Tracking Measures

The previous study on adjacent dependencies (Silva et al., 2017) that we used here as a reference for comparison combined behavioral (participants' classification of nongrammatical vs. grammatical test sequences) with eye-tracking measures (eye movements on violation vs. control letter). To optimize the comparison, we kept both techniques in the current study. In Silva et al. (2017), both types of measures captured learning outcomes. Behavioral results indicated significantly increased discrimination from pre- to post-training tests. In eye tracking, indices relating to the total extension of each trial (whole-trial measures)—including increased dwell time, increased number of fixations, and increased return to the violation letter (second-pass reading) after training—were the most effective measures. First fixation duration—characterizing participants' first reaction when confronted with the violation letter—showed no pre-post training changes. Both behavioral indices and eye movements reflected learning free of explicit influences—as indicated by the learning outcomes of a subset of participants who were totally unable to generate grammatical sequences in a postexperimental questionnaire (purely implicit learners). One exception was second-pass reading (eye tracking), which lost sensitivity when we considered the learning outcomes of purely implicit learners. Therefore, we then raised the hypothesis that second-pass reading could reflect explicit influences on a process designed to be implicit but that is nevertheless susceptible to these (Ziori et al., 2014). Maximizing the identification of possible explicit-knowledge influences was, thus, another reason for using eye-tracking measures in the present study.

Control Experiment

We carried out two experiments. In Experiment 1, all grammatical and nongrammatical test sequences were equivalent regarding repetition structure, and nongrammatical sequences did not contain any violations in symmetry between the A and B parts. To illustrate, the nongrammatical counterpart of a nested grammatical sequence *A1A1*A2|B2*B1B1* was *A1A1*A2|B1*B2B2* (same repetition structure across A and B, though with nongrammatical pairings) and not *A1A1*A2|B2*B1B2* (different repetition structure working as extra cue and nongrammatical pairings). We compared the learning outcomes for our nonadjacent dependencies with those from Silva et al. (2017), referring to adjacent dependencies. Unlike Uddén et al. (2009, 2012), we found no behavioral evidence that performance with nonadjacent dependencies could match performance with adjacent ones. To ensure that the reason why Uddén et al.'s (2009, 2012) participants performed better than ours was that the former had access to an extra cue (violations in symmetry between A and B in nongrammatical strings), we ran a control experiment. Thus, in Experiment 2, we compared learning for nongrammatical sequences with versus without these repetition-related extra cues.

Experiment 1: Learning Nonadjacent Dependencies Without Extra Cues

Materials and Method

Participants

Sample size was calculated based on effect sizes for behavioral indices of implicit learning reported by Uddén et al. (2012). For crossed grammars, Cohen's *d* was 1.24, and for nested grammars it was 1.80. According to G*Power algorithms (Faul et al., 2007), a critical alpha level of .05 and power of .80 would require 18 participants for the crossed grammar and 10 for the nested one.

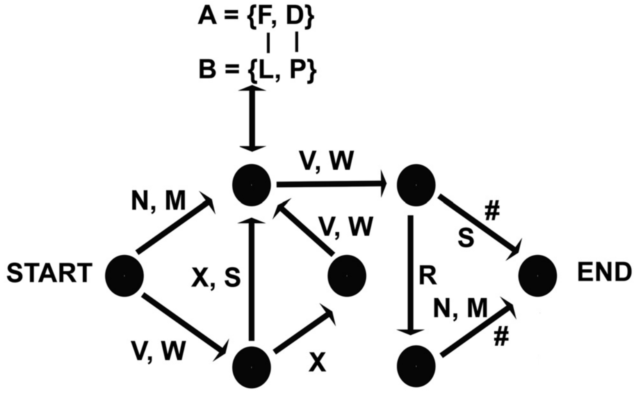
Forty-six native Portuguese speakers volunteered to take part in the experiment. Twenty-three were assigned to the crossed grammar, and the other 23 to the nested grammar. Due to excessive eye-tracking artifacts, six were excluded from crossed ($n = 17$) and five from nested ($n = 18$). Participants were matched for gender (crossed: 13 female; nested: 12 female), age (crossed: 25.6 ± 5.6; nested: 25.1 ± 4.9, $p > .74$), and years of schooling (crossed: 17.2 ± 2.5; nested: 17 ± 2, $p > .89$). All participants were prescreened for visual impairment, medication use, history of drug use, head trauma, neurological or psychiatric illness, and family history of neurological or psychiatric illness. None were bilingual, and none had Swiss German or Dutch as a second language. All gave written informed consent according to the declaration of Helsinki.

Stimulus Material

Sequences were generated with the grammar used by Uddén et al. (2009). The grammar includes a nonadjacent (crossed or nested) dependency part (see Figure 2), containing the letters [F, D, L, P]. The first half of the nonadjacent part (Part A) was selected from the set [F, D] and the second (Part B) from [L, P], F being paired with L and D with P. For the crossed grammar, pairings were made according to the rule A1A2A3B1B2B3 (e.g., FFDLLP) and, for nested, as A1A2A3B3B2B1 (e.g., FFDPLL). Sequences contained up to three nonadjacent pairs (three as in, e.g., FFDPLL; see also Table 1). The nonadjacent part was pre- and postfixed with adjacent-dependency parts from the alphabet [M, N, V, X, W, R, S] to add length and variance to the sequences and avoid start/end position effects (see Figure 2). An example of a full sequence containing nested dependencies is given in the first bulleted example below, with the nonadjacent part (A1A1A2B2B1B1) highlighted in bold. An analogous example is given in the second bullet for crossed dependencies.

Figure 2

Grammar Used in the Present Study



Note. See also Uddén et al. (2009). Grammatical sequences are generated by following the arrows that connect the circles and using the letter options available at each arrow (one letter per arrow). The start point begins with adjacent dependencies (letters N, M, V, W, X, S). Shorter or longer paths can be taken. Possible combinations could be, for example, N, M (shorter), VX, VS, VXV, VXW (longer). The top-left circle introduces the nonadjacent part (F, D as A part; L, P as B part) in the middle of the sequence. Once this part is formed, the sequence takes in adjacent dependencies again (letters N, M, V, W and also R, S). Possible combinations in this part could be, for example, VS, VRN, VRM, WS, etc. The symbol # indicates the last letter of the sequence.

- VX**FFDPLL**VS
- VX**FFDLLP**VS

We generated 100 grammatical sequences for the training set of each grammar (nested/crossed). We created three additional classification sets per grammar, each comprising 32 grammatical and 32 nongrammatical sequences. Each classification set could be used in the baseline, final preference, or grammaticality classification test. In total, we had 100 + 192 (64 × 3) sequences per grammar (to see the full stimulus set, please go to the Open Science Framework [OSF] data repository).

To test whether participants acquired deep structural knowledge on nonadjacent dependencies, surface-related differences between grammatical and nongrammatical sequences had to be controlled for. More specifically, both grammatical and nongrammatical sequences belonging to classification sets should reflect the surface properties of the training set in similar ways; otherwise, it could be argued that discrimination was due to something other than deep structural knowledge. The main innovation of the present study was controlling for symmetry in repetition structure: Since training sequences containing letter repetitions were necessarily symmetric (repetitions in A leading to repetitions in B), we granted that this principle was never violated in nongrammatical classification sequences (i.e., letter repetitions in Part A [F, D] coexisted with letter repetitions in Part B [L, P], even though in a nonlegal way; Table 1, Property 1). Moreover, training, classification-grammatical, and classification-nongrammatical subsets had the same number of sequences with letter repetitions in the A part (Table 1, Property 2), and this was valid for both crossed and nested grammars. Meeting these two criteria was a priority in the creation of classification stimuli. For this reason, control over other stimulus properties was sometimes sacrificed and compensated with control analyses on the results.

As in previous studies (Silva et al., 2017), we also controlled for ACS (Bailey & Pothos, 2008). The ACS of a sequence presented during classification indicates the frequency of occurrence of its bigrams (groups of two consecutive letters) and trigrams (groups of three) during the training phase. Ideally, grammatical and nongrammatical sequences should not differ in ACS. This was true for our stimulus set, both for the crossed grammar, $t(95) = 1.57$, $p = .12$, and the nested one, $t(95) = -.47$, $p = .96$ (Table 1, Property 3).

Since ACS was computed for the whole sequence—collapsing adjacent and nonadjacent parts—we carried out a similar control analysis isolating the nonadjacent part. For each subset (train, classification-grammatical, and classification-nongrammatical), we identified the bigrams at the nonadjacent part (Tables 1, Property 4; for detailed counts, see Appendix A) and compared them across subsets. For crossed, all bigrams from the training subset were present in grammatical as well as in nongrammatical classification sequences. For nested, this was not the case: Bigrams FP and DL were absent in training sequences as well as in grammatical classification sequences, but they were present in nongrammatical sequences. This was difficult to avoid given that bigrams FP and

Table 1
Properties of the Stimulus Sets

Property	Grammar	Train (G) (*n* = 100)	Classification G (*n* = 96)	Classification NG (*n* = 96)
1. Symmetry between Parts A [D, F] and B [P, L] in repetition structure	Nested			Always
	Crossed			Always
2. Sequence distribution according to letter repetitions in the A part (% of no-yes, example)	Nested	30−70, FDPL-FFLL	31−69, DFLP-DDPP	31−69, DFPL-DDLL
	Crossed	26−74, FDLP-FFLL	23−77, DFPL-DDPP	23−77, DFLP-DDLL
3. Sequence associative chunk strength ($M \pm SD$)	Nested	—	.56 ± .04	.56 ± .02
	Crossed	—	.56 ± .02	.55 ± .02
4. Nonadjacent bigrams (DD, DF, DL, DP, FD, FF, FL, FP, LL, LP, PL, PP) included	Nested	Except DL and FP	Except DL and FP	All
	Crossed	All	All	All
5. Sequence length (number of letters, $M \pm SD$)	Nested	9.0 ± 1.7	9.2 ± 1.5	9.3 ± 1.2
	Crossed	9.3 ± 1.8	9.9 ± 0.8	10.0 ± 0.8
6. Sequence distribution according to number of nonadjacent dependencies (% of 1-2-3)	Nested	16-33-51	19-13-68	20-20-60
	Crossed	12-32-56	0-25-75	0-32-68

Note. G = grammatical; NG = nongrammatical. Stimulus control was dominated by Properties 1 and 2.

DL were illegal (see Figure 2): F should be far apart from P and D from L; at the innermost, adjacent pair (A1*B1*|*B2*A2), F should be followed by L and D by P. Even though we managed to create nongrammatical sequences without FP or DL (42 out of 96; see OSF data repository), some (54 out of 96) included these bigrams. To circumvent the effects of this potential extra cue for discrimination (facilitating nested compared to crossed), we ran a control analysis where we excluded nongrammatical sequences containing FP or DL bigrams (see Statistical Analysis).

Prioritizing repetition-structure-related criteria in stimulus creation (Table 1, Properties 1–2) attenuated control over other stimulus properties, and sequence length was one of these (Table 1, Property 5). Specifically, crossed and nested grammars differed in how the length of test sequences related to the length of training sequences. In the crossed grammar, *t* tests showed that both grammatical and nongrammatical classification sequences were significantly longer than training sequences, grammatical: $t(95) = -3.90$, $p < .001$; nongrammatical: $t(95) = -3.97$, $p < .001$. In nested sequences, both grammatical and nongrammatical sequences had the same length as training sequences (grammatical: $p = .39$; nongrammatical: $p = .12$). Thus, although discrimination between grammatical and nongrammatical could not be aided by length in any of the two grammars, crossed (but not nested) grammar learners had to deal with classification sequences overall longer than the ones they saw during training. Either because longer sequences could increase processing challenges or simply because they differed from the training set, crossed participants could be at a disadvantage compared to nested participants. To rule out this possibility, we tested whether longer sequences decreased learning.

Finally, we compared sequence types based on the number of nonadjacent pairs (sequences with one, two, or three pairs; Table 1, Property 6) across training, classification-grammatical, and classification-nongrammatical subsets. In all subsets and grammar types, three-pair sequences were the dominant type. However, there was again a distinction between nested and crossed that could favor nested grammar learners: In nested, all three sequence types (one, two, or three pairs) existed in training, classification-grammatical, and classification-nongrammatical; in crossed, the three types were present in training, but only two- and three-pair sequences existed in classification (grammatical or nongrammatical). To rule out this possible confound, we ran another control analysis (see Statistical Analysis) that excluded all one-pair nested classification sequences.

Procedure

The experiment spanned 9 days spread over 2 weeks, with one training session each day (see Table 2). On Day 1, a baseline preference classification test ("like or dislike the sequence?") was administered before the first training session (AGL1). On Day 9, subjects did another preference classification test (final preference, AGL2). Since preference classification tests are not standard in AGL literature, we ran a second posttraining discrimination test for control using the standard grammaticality classification (AGL3: "Is sequence correct or incorrect? Guessing should be based on gut feeling"). Right after AGL2, participants were informed about the existence of rules, but not the rules themselves, and they moved on to AGL3. Eye-tracking data were collected in the three classification sessions.

Table 2
Tasks Performed by Participants in Each Day of the 9-Day Experiment

Day	Tasks
Day 1	Baseline preference (AGL1), 15 min, 64 trials, using eye tracking; Short-term memory task, 100 trials, 15 min
Days 2−8	Short-term memory task, 100 trials, 15 min
Day 9	Short-term memory task, 100 trials, 15 min; Final preference (AGL2), 64 trials, 15 min, using eye tracking; Grammaticality classification (AGL3), 64 trials, 15 min, using eye tracking; Explicit knowledge questionnaire, 10 min

The 9-day acquisition task was presented as a short-term memory task. Each sequence was shown on a computer monitor for 4,000 ms, after which the subject typed it from memory on a keyboard, in a self-paced manner. No feedback was provided. In the three classification tests, sequences were presented for 4,000 ms, after which a question appeared on screen, and the subject responded by clicking one of two options with the mouse.

To assess participants' levels of explicit knowledge, we administered a postexperimental questionnaire combining verbal report and direct tests (Rebuschat, 2013; see Appendix B). It started with vague questions related to noticing patterns in the sequences and the use of strategies for classification. At this point, the subject was asked to generate grammatical sequences. The subject was then asked about knowledge of the grammar in a progressively focused manner, following the topics of (a) middle versus extreme (nonadjacent vs. adjacent) letters, (b) two halves of middle letters (A-B nonadjacent part), and (c) pairings between the elements of the two halves.

Eye-Tracking Data Recording and Preprocessing

Monocular eye movements were recorded at 1,250 Hz with an SMI hi-speed system (http://www.smivision.com). Participants placed their chin on a chin rest and held a mouse for response. They sat 80 cm away from a 22-in. monitor. At that distance, the interletter space subtended horizontally 1.8° of the visual angle (approximately the angle of the fovea). A 5-point calibration was performed twice in each test (before the test and during an interval between 32-item blocks), followed by validation. The tracking error was smaller than .5°.

Preprocessing was carried out with the SMI analysis software and then with MatLab (http://www.mathworks.com). Events were extracted based on a high-speed algorithm for saccade detection. A threshold of 30° peak velocity and a minimum duration of 22 ms defined a saccade. Fixations shorter than 50 ms were rejected. Trials were visually inspected, and those with excessive artifacts were rejected. Trials with a number of fixations 3 standard deviations above or below the subject's mean per condition were marked as outliers and also rejected. The mean rejection rates were between 0% and 2.6% in the crossed and between 1.6% and 3.6% in the nested grammar.

The analysis software computed the values of four eye features for critical letters as areas of interest. The critical letter was the first letter of the second-half nonadjacent part (first B), corresponding to the first violation the subject would encounter while reading from left to right. The four eye features were chosen to be the same as the those of our previous experiment (Silva et al., 2017), which included (a) first fixation duration; the (b) proportion of fixations on the critical letter (a number) relative to number of fixations in the whole string; the (c) proportion of dwell time on the critical letter, computed in similar terms; and the (d) ratio between dwell time (absolute) on the critical letter and first fixation duration. The first measure characterized the first confrontation participants had with the violation letter, and the other three contemplated their ocular movements during the full trial time (whole-trial measures). Among these three, the proportions of fixations and the proportion of dwell time characterized how the violation letter attracted the initiation of fixations and the time spent on the violation letter. Both measures indicate the extent to which the violation grabbed participants' attention. The last measure emphasized the amount of second-pass reading, return to the violation letter area after the first fixation, suggesting revision processes.

Statistical Analysis

For each grammar type (crossed vs. nested), we considered the effects of test (AGL1, AGL2, AGL3) and grammatical status of the sequence (grammatical vs. nongrammatical). At the behavioral level, we analyzed endorsement rates (classification of sequences as likeable/correct per grammatical status and test). Behavioral learning was indexed by increased endorsement of grammatical sequences and decreased endorsement/liking of nongrammatical after training (AGL1 vs. AGL2, two tests with the same instruction). Further comparisons between AGL2 and AGL3 served to analyze instruction effects (state preference, AGL2 vs. classify as correct or incorrect, AGL3) in posttraining classification. Grammaticality classification tests like AGL3 tend to provide similar (Forkstam et al., 2008) or stronger learning outcomes than AGL2 (Silva et al., 2017), with the reverse being unexpected. Differences between AGL2 and AGL3 are often assigned to increased explicit-knowledge influences in AGL3 (Forkstam et al., 2008; Manza & Bornstein, 1995; Newell & Bright, 2001): Once participants are informed about the existence of rules, they may start drawing conscious hypotheses about the content of these rules based on their previous exposure. In this sense, comparisons between AGL2 and AGL3 could also provide hints on uses of explicit knowledge and their possible eye-tracking correlates. For eye-tracking data, we analyzed four eye features per grammatical status and test: first fixation duration, proportion of dwell time, proportion of fixations, and dwell-to-first-fixation ratio—all on the critical letter. Eye-movement indices of learning should consist of increased values for nongrammatical (longer processing times for violations) and/or decreased ones for grammatical when comparing AGL1 with AGL2. Similar to behavioral analyses, we also compared AGL2 with AGL3 to address instruction effects. The main eye-tracking data analysis included all trials—with or without accurate behavioral responses. We did a cross-check analysis with accurate trials only (see Appendix C).

In all analyses, we used linear mixed models as implemented in the lme4 package (Bates et al., 2015; lmerTest package used for significance values, sjPlot for tables) for R (R core team, 2013). Test, grammatical status, and grammar type (crossed vs. nested) were entered as fixed factors and participants as a random factor (random intercept). We first focused on the interactions between test and grammatical status (learning nonadjacent dependencies) and that between test, grammatical status, and grammar type (learning across nonadjacent grammar types), considering the two test-related comparisons (AGL1&2, AGL2&3) one at a time. For significant Test × Grammatical Status × Grammar Type interactions, we broke down the analysis to consider grammar types (crossed vs. nested) one at a time. Critical alpha levels were set to .05.

Additional analyses were run to address potential stimulus-related confounds (see Stimulus Material). To rule out the possibility that crossed grammar learners were at a disadvantage compared to nested grammar learners due to crossed classification sequences being longer than training sequences, we tested whether behavioral accuracy decreased with sequence length. To rule out the other two potential sources of advantage to nested grammar learners—absence of one-pair nonadjacent sequences in crossed classification subsets and nongrammatical-exclusive bigrams in the nested but not in the crossed grammar—learning-related analyses were performed three times: first, with the full data set (all crossed and nested sequences); second, excluding nested sequences with a single pair of letters at the nonadjacent part (see Table 1); and third, excluding nested nongrammatical sequences containing nongrammatical-exclusive bigrams (FP or DL; Table 1).

Responses to questionnaires were classified to provide an explicit-knowledge score. This score ranged from 0 to 4, and it was computed as follows: Participants were first assigned 1, 2, or 3 points according to their explicit knowledge on middle versus extreme (nonadjacent vs. adjacent) letters (Level 1), letters in each of the two halves of the nonadjacent part (Level 2, implying 1), and pairings between the elements of the two halves (Level 3, implying 1 and 2). The number of correctly generated strings (only nonadjacent part considered) was then added, with participants scoring plus 1 point if all 5 generated sequences were correct and the corresponding score in case of partial hit (e.g., .4 for two correct sequences). Explicit knowledge scores were compared across grammar types. To test for explicit-learning trends in our sample, we correlated the cross-test increase in behavioral discrimination (increase in *d*-prime) with explicit knowledge scores. *D*-prime values (Stanislaw & Todorov, 1999) for a given test [Z(hits) − Z(false alarms)] index are essentially the same as differences in endorsement rates across grammatical and nongrammatical sequences (grammatical endorsed − nongrammatical endorsed), which we used in behavioral analyses (for a *d*-prime-based analysis of behavioral results, please visit the OSF data repository associated with this study).

Results

Behavioral Discrimination Across Tests

Increases in discrimination from AGL1 to AGL2 did not reach significance (nonsignificant Test × Grammatical Status or Test × Grammatical Status × Grammar interaction; Table 3, Figure 3) for nonadjacent dependencies, whatever the stimulus set (full, nested without one nonadjacent pair, nested without FP or DL bigrams). Comparisons between AGL2 and AGL3 (Table 4, Figure 3) showed nonsignificant results too. This contrasts with the results of Silva et al. (2017) on adjacent dependencies, where discrimination increased from AGL1 to AGL2 and from AGL2 to AGL3. As suggested in Figure 3, the poor learning outcomes for nonadjacent dependencies seem related to participants' inability to reject nongrammatical sequences after training.

Behavioral Accuracy and Sequence Length (Control Analysis)

This control analysis aimed to rule out the possible disadvantage of crossed compared to nested grammar learners due to crossed classification sequences being longer than crossed training sequences. For crossed and nested sequences altogether, we found that accurately classified sequences were longer (accuracy predicts length; $B = .16$, confidence interval [95% CI; .10, .21], $p < .001$). This pattern prevailed for nested sequences alone ($B = .12$, [.04, .20], $p < .003$). For crossed sequences, there was no significant association ($B = .01$, [−.03, .06], $p = .564$). Therefore, the increased length of crossed classification sequences compared to nested does not seem able to decrease learning.

Table 3

Grammatical Status (G vs. NG) × Test (AGL1 vs. AGL2) Interactions on Behavioral (Endorsement Rates) and Eye-Tracking Learning Indices

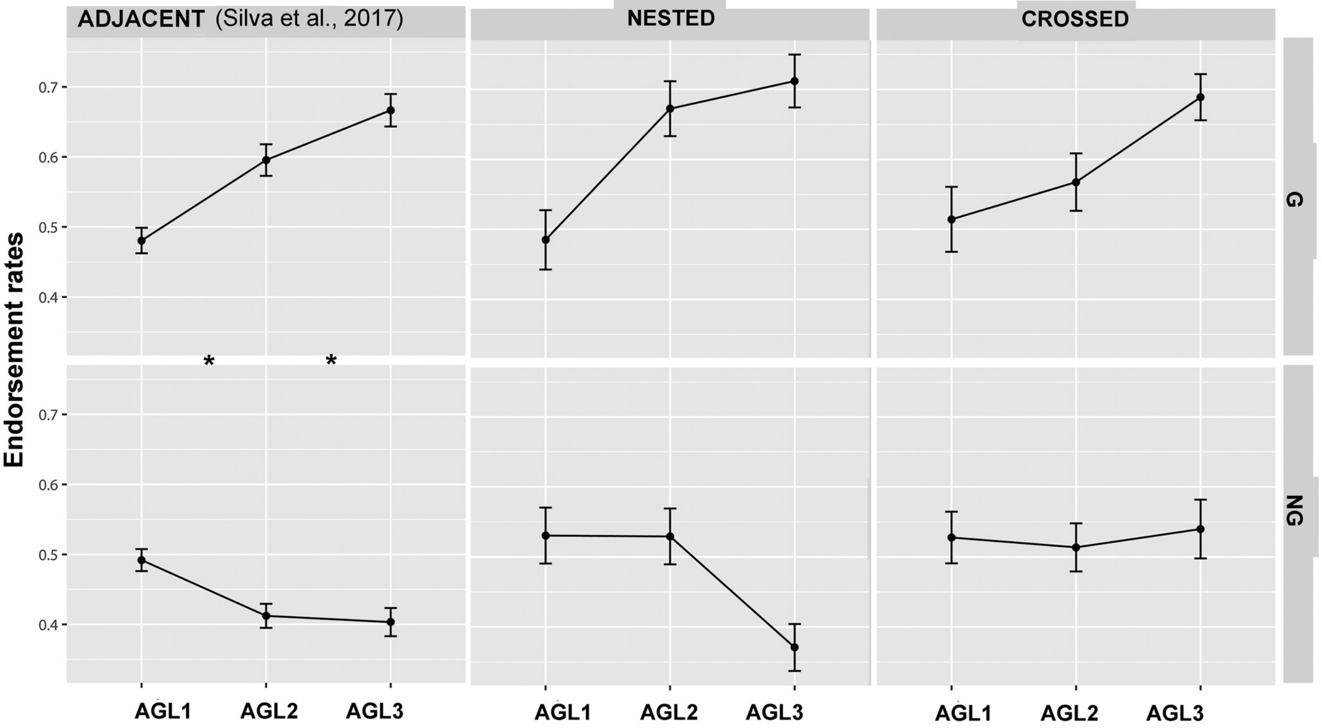
Test: AGL1 vs. AGL2	GS × Test B; [95% CI]; p; (R^2)	GS × Test × Grammar B; [95% CI]; p; (R^2)	GS × Test, Nested B; [95% CI]; p; (R^2)	GS × Test, Crossed B; [95% CI]; p; (R^2)
Endorsement rates				
All NE	−0.09; [−20.00, .02]; .120	−0.06; [−22.00, .10]; .433		
NE without 1 NA pair	−0.09; [−20.00, .02]; .120	−0.06; [−21.00, .10]; .467		
NE without FP/DL	−0.09; [−20.00, .05]; .193	−0.16; [−35.00, .03]; .096		
First fixation duration				
All NE	11.27; [−17.70, 40.23]; .446	−3.49; [−43.86, 36.87]; .865		
NE without 1 NA pair	11.26; [−16.71, 39.24]; .430	15.53; [−25.62, 56.68]; .459		
NE without FP/DL	11.26; [−17.32, 39.85]; .440	6.85; [−39.10, 52.80]; .770		
Proportion of dwell time				
All NE	[a]	0.04; [.01, .06]; **.001**; (.082)[b]	0.03; [.01, .05]; <**.001**; (.069)	−0.01; [−.02, .01]; .493
NE without 1 NA pair		0.04; [.01, .06]; **.001**; (.071)	0.03; [.01, .05]; <**.001**; (.059)	
NE without FP/DL		0.03; [.00, .05]; **.030**; (.074)	0.02; [.00, .04]; **.030**; (.071)	
Proportion of fixations				
All NE		0.03; [.01, .05]; <**.001**; (.084)	0.03; [.02, .04]; <**.001**; (.076)	−0.01; [−.02, .01]; .430
NE without 1 NA pair		0.03; [.01, .05]; **.002**; (.072)	0.03; [.01, .04]; <**.001**; (.067)	
NE without FP/DL		0.03; [.00, .05]; **.018**; (.081)	0.02; [.00, .04]; **.015**; (.082)	
Dwell/first fixation				
All NE		0.54; [.23, .85]; <**.001**; (.079)	0.39; [.18, .60]; <**.001**; (.083)	−0.15; [−.38, .07]; .181
NE without 1 NA pair		0.44; [.13, .76]; **.006**; (.071)	0.29; [.07, .51]; **.009**; (.071)	
NE without FP/DL	−0.15; [−.37, .07]; .168	0.35; [.00, .69]; .051		

Note. GS × Test × Grammar (nested vs. crossed) interactions are broken down when significant. Analyses were run three times: all nested (NE) sequences, NE sequences with more than one nonadjacent (NA) pair, and NE sequences not containing exclusive bigrams (FP or DL). GS = grammatical status; G = grammatical; NG = nongrammatical; AGL1 = baseline preference; AGL2 = final preference; CI = confidence interval.

[a] GS × Test interactions are not presented when GS × Test × Grammar (nested vs. crossed) interactions were significant. [b] Significant interactions highlighted in bold.

Figure 3

Endorsement Rates (Proportion of Liked/Correct Sequences) Across Grammatical Status (G = Grammatical; NG = Nongrammatical) and Test (AGL1 = Baseline Preference Classification; AGL2 = Final Preference Classification; AGL3 = Grammaticality Classification)



Note. Bars indicate the standard error of the mean. Asterisks indicate significant increase in discrimination from AGL1 over AGL2 or from AGL2 over AGL3. The leftmost panel plots results from Silva et al. (2017) on adjacent dependencies.

Explicit Knowledge

The internal consistency of the explicit questionnaire was satisfactory (Cronbach's α = .71). Participants showed an average score of 1.54 (*SD* = 1.06). Values ranged between 0 and 3, indicating that the maximum score of explicit knowledge (4) was not reached by any participant; although some participants were able to identify the A and B halves of the nonadjacent part (2 points), and some generated valid grammatical sequences (plus 1 point), none were able to state the pairings between letters from the A and B parts (Figure 4A). Sequence generation was at chance level in both grammars (nested: $p > .89$; crossed: $p > .69$; see Figure 4B).

Although participants from the two grammar groups did not differ in explicit scores ($p > .90$), the correlations between explicit scores and learning (posttraining increase in *d*-prime) seemed to indicate differences: For nested, correlations were nonsignificant (AGL1&2 learning: $p > .76$; AGL2&3 learning: $p > .22$); for crossed, correlations between learning and explicit scores were marginal (AGL1&2: $r = .457$, $p = .065$; AGL2&3: $r = .433$, $p = .082$). Nevertheless, Fisher *Z* tests for comparing correlations (Lenhard & Lenhard, 2014) did not reach significance—neither for AGL1&2 ($p > .13$) nor for AGL2&3 ($p > .34$). The contrast between crossed and nested was more pronounced when we correlated sequence generation scores (a component of explicit scores) with learning: For nested, correlations were nonsignificant (AGL1&2 learning: $p = .10$; AGL2&3 learning: $p = .990$), while for crossed they reached significance (AGL1&2: $r = .518$, $p = .033$; AGL2&3: $r = .555$, $p = .021$). Here, Fisher *Z* tests showed significant differences across nested and crossed for AGL2&3 ($z = 1.69$, $p = .045$), though they did not for AGL1&2 ($p > .34$). Thus, the amount of acquired explicit knowledge was similar in the two grammars, but only crossed grammar learners may have used this to some extent, namely in the final classification task (AGL3).

Eye-Movement Discrimination Across Tests

Significant training-related differences in eye-movement discrimination (AGL1, baseline preference vs. AGL2, final preference classification) were restricted to the nested grammar (Table 3, Figure 5). Paralleling the results we had for adjacent dependencies in Silva et al. (2017; see Figure 5, left), effects showed up for whole-trial measures (proportion of dwell time, proportion of fixations, and dwell/first fixation on critical letter). The control analysis without nongrammatical-exclusive bigrams (potential aid in nested grammar learning) rendered dwell/first fixation indices nonsignificant (see Table 3). No other changes were introduced by control analyses, with proportion of dwell time and proportion of fixations responding consistently to training regardless of potential confounds (see Table 3).

Comparisons between final preference (AGL2) and grammaticality classification (AGL3) showed increased discrimination

Table 4

Grammatical Status (G vs. NG) × Test (AGL2 vs. AGL3) Interactions on Behavioral (Endorsement Rates) and Eye-Tracking Learning Indices

Test: AGL2 vs. AGL3	GS × Test *B*; [95% CI]; *p*; (R^2)	GS × Test × Grammar *B*; [95% CI]; *p*; (R^2)	GS × Test, Nested *B*; [95% CI]; *p*; (R^2)	GS × Test, Crossed *B*; [95% CI]; *p*; (R^2)
Endorsement rates				
All NE	−0.13; [−.27, .01]; .077	−0.06; [−.26, .13]; .521		
NE without 1 NA pair	−0.13; [−.28, .02]; .099	−0.03; [−.18, .24]; .765		
NE without FP/DL	−0.22; [−.29, −.04]; .141	0.02; [−.22, .25]; .886		
First fixation duration				
All NE	11.37; [−15.30, 38.04]; .403	19.10; [−18.15, 56.34]; .315		
NE without 1 NA pair	11.37; [−14.10, 36.84]; .382	6.96; [−30.62, 44.54]; .717		
NE without FP/DL	11.37; [−15.09, 37.83]; .400	1.68; [−40.90, 44.27]; .938		
Proportion of dwell time				
All NE	0.01; [.00, .03]; .098	0.00; [−.03, .02]; .696		
NE without 1 NA pair	0.01; [.00, .03]; .080	0.00; [−.03, .02]; .734		
NE without FP/DL	0.01; [.00, .03]; .080	−0.02; [−.05, .02]; .050		
Proportion of fixations				
All NE	0.01; [.00, .02]; .158	−0.01; [−.02, .01]; .609		
NE without 1 NA pair	0.01; [.00, .02]; .136	0.00; [−.02, .02]; .844		
NE without FP/DL	0.01; [.00, .02]; .134	−0.02; [−.04, .00]; .124		
Dwell/first fixation				
All NE	0.22; [.01, .43]; **.038**[a]; (.069)	−0.23; [−.53, .06]; .121		
NE without 1 NA pair	0.22; [.02, .43]; **.034**; (.066)	−0.16; [−.47, .14]; .299		
NE without FP/DL		−0.40; [−.72, .08]; **.015**	−0.18; [−.42, .07]; .159	0.22; [.02, .43]; .031

Note. GS × Test × Grammar (nested vs. crossed) interactions are broken down when significant. Analyses were run three times: all nested (NE) sequences, NE sequences with more than one nonadjacent (NA) pair, and NE sequences not containing exclusive bigrams (FP or DL). GS = grammatical status; G = grammatical; NG = nongrammatical; AGL2 = final preference; AGL3 = grammaticality classification; CI = confidence interval.

[a] Significant interactions highlighted in bold.

related to dwell/first fixation in both grammar types (Figure 5, Table 4). Again, the control analysis without nongrammatical-exclusive bigrams in nested sequences rendered nested-related results nonsignificant. Accurate-only trials (see Appendix C) showed the same pattern as all trials considered (correct and incorrect), except that dwell/first fixation indices of learning were never significant—neither for the AGL1&2 comparison nor for AGL2&3.

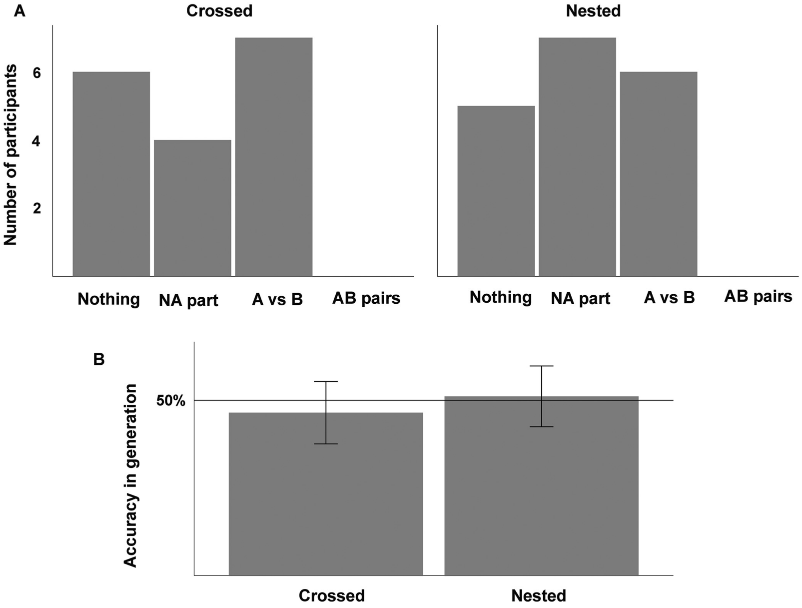
Discussion

Previous studies (Uddén et al., 2009, 2012) have suggested that learning nested and crossed nonadjacent dependencies may be as easy as learning adjacent ones, provided that participants are given more time to learn. Our first goal was to determine whether the high performance of participants in Uddén et al.'s (2009, 2012) study was due to learning of the internal structural patterns based on increased exposure time (9 days instead of 5) or, on the contrary, if participants might have taken advantage of an uncontrolled surface distinction that allowed discrimination but not necessarily deep structural learning. The hypothesized surface distinction was the violation of symmetry between A and B concerning letter repetitions in nongrammatical but not in grammatical sequences. To achieve our goal, we created a stimulus set for classification in which both grammatical and nongrammatical nested and crossed sequences had symmetric repetitions (e.g., the nongrammatical counterpart of FFDLLP would be FFDPPL, and never FFDPLP), we ran a 9-day AGL study, and we compared the results with those from a previous behavioral and eye-tracking 5-day study of ours (Silva et al., 2017) on adjacent dependencies. We controlled for other potential confounds, such as ACS, sequence length, number of nonadjacent dependencies, and bigrams exclusive to nongrammatical sequences.

Results pointed to lower performance in nonadjacent dependencies, despite the 9-day exposure: Unlike in Silva et al. (2017), participants' behavioral indices of learning in the present study were nonsignificant for both crossed and nested nonadjacent dependencies. Critically, the pattern of results indicated that—although participants increased endorsement of grammatical sequences after training—they were unable to increase rejection of nongrammatical items. This is consistent with the possibility that, in previous studies, they might have relied on symmetry violations, instead of grammaticality violations, to recognize nongrammatical sequences. Concerning eye tracking, we saw significant learning indices for nested dependencies, engaging the same measures as in Silva et al. (2017)—proportion of dwell time, proportion of fixations, and dwell/first fixation—but dwell/first fixation indices did not survive control analyses addressing the influence of bigrams exclusive to nongrammatical sequences. Crossed dependencies showed no eye-tracking evidence of implicit learning. Overall, our findings do not support the idea that learning nonadjacent dependencies may be just "a matter of time" (Uddén et al., 2009, 2012): When participants are deprived of surface cues for discrimination between grammatical and nongrammatical nonadjacent sequences—as they were in our study—prolonged training does not seem to grant learning, at least at a level comparable to learning adjacencies.

Our second goal related to the comparative learnability of nested versus crossed sequences. Formal language hierarchy allows predicting increased difficulty for crossed sequences, but several studies pointed to increased difficulty for nested. Since most of these studies have been conducted with Dutch participants—whose native language contains crossed dependencies—we

Figure 4
Explicit Knowledge Scores



Note. Panel A: Distribution of scores reached in each group, according to three levels of knowledge (1 = NA part, knowledge of letters in the middle vs. extremes; 2 = A vs. B, knowing which letters belong to A and B in the nonadjacent part; 3 = AB pairs, knowing how letters from A and B pair with each other). Panel B: Mean accuracy in the attempt to generate correct sequences (0–100%). Sequence generation was at chance level (50%) in both grammars.

hypothesized that the advantage of crossed dependencies could be due to linguistic experience (Christiansen & Chater, 2015), which would override cognitive-architectural constraints. To address this hypothesis, we ran our study on Portuguese-speaking participants, inexperienced with crossed dependencies. We saw an advantage of nested over crossed dependencies in eye-tracking learning indices, suggesting that linguistic experience may have played a role in previous studies that showed advantages of crossed over nested. As for the possibility that nested sequences may have been learned more efficiently than crossed ones due to influences other than grammar type itself—decreased length of sequences, presence of one-dependency-pair sequences in classification, presence of bigrams exclusive to nongrammatical sequences—these were ruled out by control analyses.

It has been argued that nested dependencies have one characteristic that lends them potential advantages over crossed ones: In nested, but not crossed, grammars, there is always one pair that works like an adjacent dependency (Lai & Poletiek, 2011). For instance, in A1A2A3|B3B2B1, the inner pair A3B3 is part of the nested structure, but its element letters are contiguous. Is it possible that this accounted for the advantage of nested over crossed dependencies in the present study? The short answer is no: As shown in Table 1 and Appendix A, middle bigrams of nested sequences (FL or DP) were not exclusive to grammatical items from the stimulus sets. We managed to avoid such exclusiveness by using crossed-like sequences in nested nongrammatical subsets (e.g., VXDD*FL*LPWS, VSFF*DP*PLWS). As we pointed out throughout this article, the only bigram-exclusiveness scenario regarded FP and DL in nongrammatical, but not in grammatical, sequences. As our control analyses showed, evidence of learnability for nested sequences remained after excluding nongrammatical sequences with these two bigrams.

The poor behavioral performance we saw in this experiment—specifically, the inability of participants to reject nongrammatical sequences—suggests that the significant learning of nonadjacent dependencies found by Uddén et al. (2009, 2012) was due to the presence of the extra cue that we did not allow here: violations in repetition structure for nongrammatical, but not for grammatical, sequences. Nevertheless, we cannot be certain of this until we make a direct comparison between learning with and without such extra cue. We did it in Experiment 2.

Experiment 2: Learning Nonadjacent Dependencies With Versus Without Extra Cues

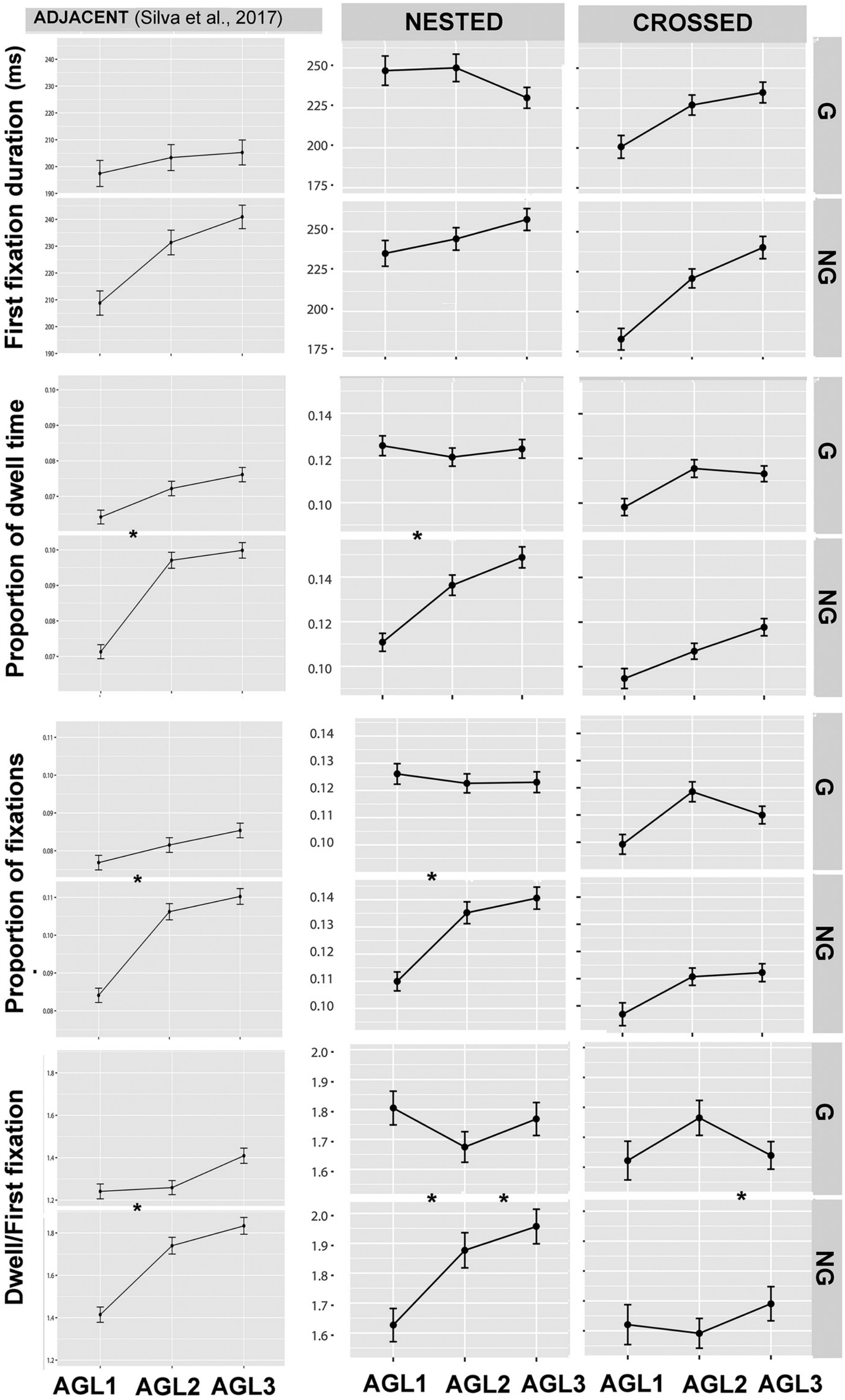
Materials and Method

Participants

Forty-six native Portuguese speakers volunteered to take part in the experiment. Twenty-three were assigned to the crossed grammar and the other 23 to the nested grammar. Participants were matched for gender (crossed: 19 female; nested: 17 female), age (crossed: 22.9 ± 7.3; nested: 19.6 ± 3.9, $p = .069$), and years of schooling (crossed: 13.8 ± 2.5; nested: 12.8 ± 1.3, $p = .11$).

Figure 5

Eye-Movement Indices (First Fixation Duration, Proportion of Dwell Time, Proportion of Fixations, Dwell/First Fixation on Critical Letter) Across Grammatical Status (G = Grammatical; NG = Nongrammatical) and Test (AGL1 = Baseline Preference Classification; AGL2 = Final Preference Classification; AGL3 = Grammaticality Classification)



Note. Bars indicate the standard error of the mean. Asterisks indicate significant increase in discrimination from AGL1 over AGL2 or from AGL2 over AGL3. The leftmost panel plots results from Silva et al. (2017) on adjacent dependencies.

Exclusion criteria were the same as in Experiment 1. All gave written informed consent according to the declaration of Helsinki.

Stimulus Material

We kept the training stimulus set of Experiment 1 ($n = 100$) and expanded the classification set ($n = 96/96$ for nested/crossed) so that it also included NG sequences with violations of symmetry in repetition structure. The original classification stimulus set contained 66/74 nongrammatical sequences (69%/74% of 96) with letter repetitions for nested/crossed grammars (see Table 5), all of them symmetric. For the present experiment, these sequences with letter repetitions were duplicated and then edited in the last nonadjacent position (last letter of Part B) such that symmetry in repetition structure was broken. For instance, the sequence MDFD*LPL*VR was duplicated and transformed into MDFD*LPP*VR; NDDD*LLL*VR was transformed into NDDD*LLP*VR (for the full set of nongrammatical sequences, see the OSF data repository**).**

This process generated 66/74 additional nongrammatical sequences for nested/crossed, enlarging the nongrammatical test set to 162 sequences (66 with symmetry violations + 66 without symmetry violations + 30 without letter repetitions; Table 5) in nested and to 170 sequences in crossed (74 with symmetry violations + 74 without symmetry violations + 22 without letter repetitions). To keep the 50/50 proportion of grammatical/nongrammatical sequences during test, we duplicated the grammatical sequences with letter repetitions (66/74 in nested/crossed). As in Experiment 1, the full classification set was divided into three blocks, to be used at the three different classification tasks of the experiment (baseline preference, final preference, grammaticality classification).

Procedure

Apart from the fact that the experiment did not register eye movements, the procedure was the same as in Experiment 1.

Statistical Analysis

Nongrammatical sequences without symmetry violations and without letter repetitions were grouped together as sequences without extra cues (96/96 in nested/crossed, same as in Experiment 1), as opposed to the new nongrammatical sequences with symmetry violations (extra cue present, 66/74 in nested/crossed). Each subset of nongrammatical sequences was compared to all grammatical sequences (162/170 for nested/crossed) for endorsement rates.

Table 5
Distribution of Classification Sequences in Experiment 2

Sequences	Nested	Crossed
Nongrammatical		
Experiment 1		
Without letter repetitions	30	22
With letter repetitions, no symmetry violations	66	74
Added		
With letter repetitions and symmetry violations	66	74
Total nongrammatical	162	170
Grammatical		
Experiment 1		
Without letter repetitions	30	30
With letter repetitions	66	74
Added		
With letter repetitions (duplicate)	66	74
Total grammatical	162	170

The analysis followed the same principles and procedures as in Experiment 1, with the exception that symmetry violation (no vs. yes) was added to the model as a fixed factor. Explicit scores were also analyzed in the same way as in Experiment 1.

Results

Behavioral Discrimination Across Tests

For the comparison between baseline preference (AGL1) and final preference (AGL2), the model (marginal R^2/conditional R^2 = .174/.621) showed a significant interaction between grammatical status, test, and symmetry violations ($\beta = -.19$, 95% CI [−.35, −.40], $p = .016$). All other terms were nonsignificant. Breaking down the interaction into symmetry violation levels (see Figure 6), we found significant learning outcomes when symmetry breaks were present at nongrammatical strings ($\beta = -.30$, [−.42, −.17], $p < .001$, marginal R^2/conditional R^2 = .198/.565) but not when they were absent ($\beta = -.10$, [− .21, −.00], $p = .059$, marginal R^2/conditional R^2 = .140/.602). In both cases, further interactions with grammar type were null ($p > .40$), indicating no differences between nested and crossed grammars. Comparisons between final preference and grammaticality classification showed no significant interactions engaging Grammatical Status × Test (*p*s > .08), indicating no evidence of boosted learning outcomes in the grammaticality classification test.

Explicit Knowledge

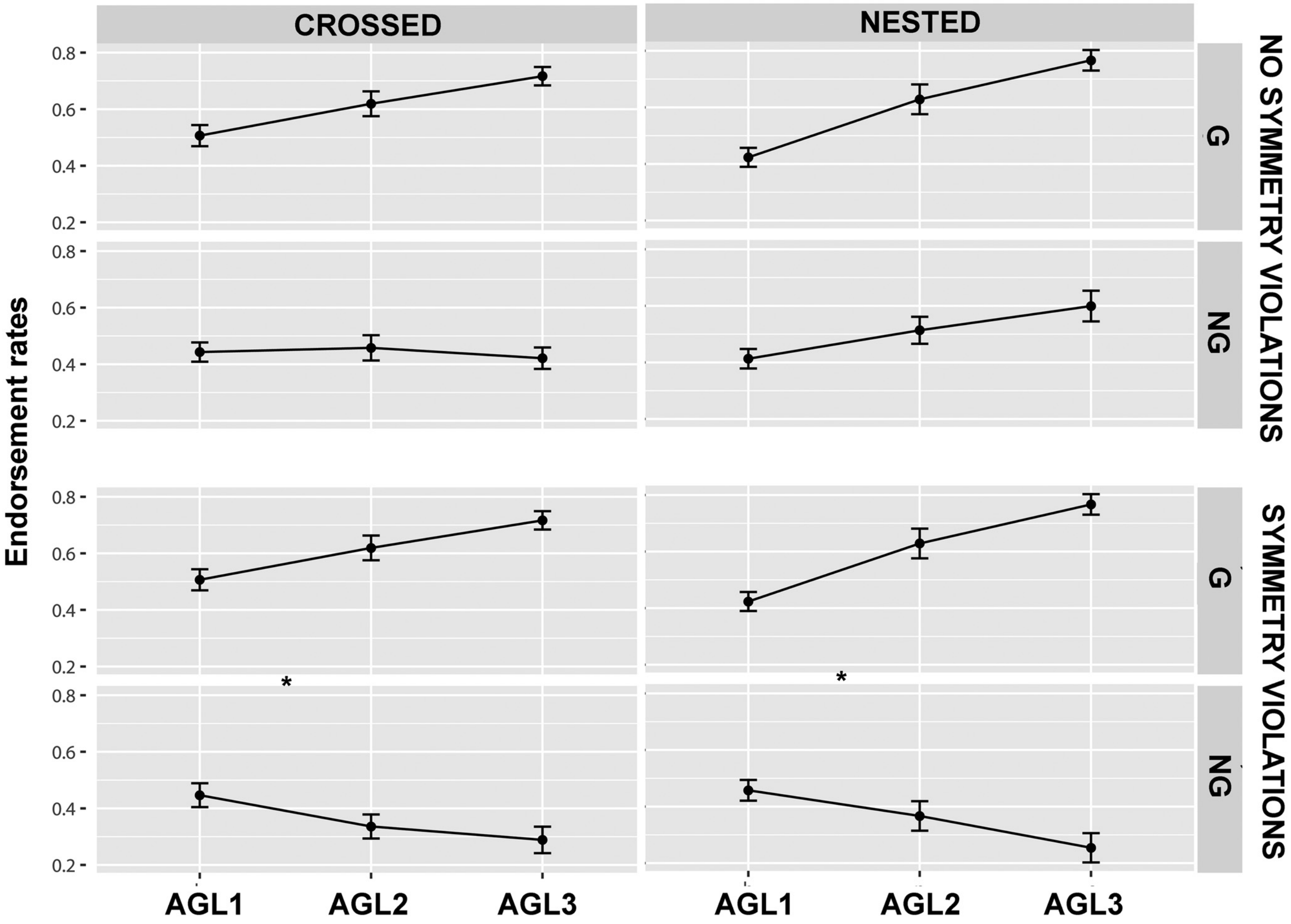
The internal consistency of the explicit questionnaire was satisfactory (Cronbach's $\alpha = .76$). Participants showed an average score of 1.61 ($SD = 1.33$) in a scale ranging from 0 to 4. Sequence generation was at chance level in both grammars (nested: $p > .49$; crossed: $p > .95$). As in Experiment 1, participants from the two grammars did not differ in explicit scores ($p > .23$).

For sequences like the ones in Experiment 1 (no symmetry violations in nongrammatical strings), correlations between explicit scores and *d*-prime increase did not reach significance in any grammar (nested: AGL1&2, $p > .060$, AGL2&3, $p > .16$; crossed: AGL1&2, $p = .070$, AGL2&3, $p > .15$). When correlating *d*-prime increase with sequence generation scores (one component of explicit scores), values were nonsignificant for nested (AGL1&2: $p > .59$; AGL2&3: $p > .33$) and significant for crossed regarding AGL1&2 ($r = .440$, $p = .035$; AGL2&3: $p > .12$). Nevertheless, the Fisher *Z* test comparing the correlations between AGL1&2 learning and generation scores for nested versus crossed did not reach significance ($p > .13$). Comparing these results with those from Experiment 1, we conclude that the increased reliance on explicit knowledge in crossed than in nested when going from AGl2 to AGL3 no longer applied.

For sequences with symmetry violations, the correlation between explicit scores and *d*-prime change was significant at AGL2&3 ($r = .472$, $p = .023$; AGL1&2: $p > .84$) in nested sequences, and both correlations were null (AGL1&2: $p > .20$; AGL2&3: $p > .13$) in crossed. However, direct comparisons showed no cross-grammar differences. Differences between nested and crossed emerged when we correlated sequence generation scores with *d*-prime change. Correlations did not reach significance for nested (AGL1&2: $p > .29$;

Figure 6

Endorsement Rates (Proportion of Liked/Correct Sequences) Across Symmetry Violation Levels, Grammatical Status (G = Grammatical; NG = Nongrammatical), and Test (AGL1 = Baseline Preference Classification; AGL2 = Final Preference Classification; AGL3 = Grammaticality Classification)



Note. Bars indicate the standard error of the mean. Asterisks indicate significant increase in discrimination from AGL1 over AGL2 or from AGL2 over AGL3.

AGL2&3: $p > .09$), but they did so for crossed (AGL1&2: $r = .446$, $p = .033$; AGL2&3: $r = .473$, $p = .023$). Direct comparisons showed differences between nested and crossed at AGL1&2 ($Z = -2.25$, $p = .012$; AGL2&3: $p > .32$), suggesting that symmetry violations may have induced some degree of explicit learning in the crossed grammar.

Discussion

Within-subjects comparisons between learning with and learning without the extra cue of violations in repetition structure (present in Uddén et al., 2009, 2012, and absent in Experiment 1) showed significant learning when the extra cue was present but not when it was absent (replicating Experiment 1 in this matter). This strengthens our claim that learning nonadjacent dependencies is not just "a matter of time," and Uddén's findings may have been confounded by the presence of an extra cue for discrimination. According to analysis on explicit scores, the extra cue may have fostered explicit learning, at least in the nested grammar.

General Discussion

Our study provided support for two ideas. First, behavioral results from Experiments 1 and 2 showed that nonadjacent dependencies may be harder to learn implicitly than adjacent ones, even when participants are given more training time. Second, judging by the eye-tracking results of Experiment 1, crossed dependencies may be harder to learn than nested ones. Together, these ideas suggest that the hierarchy of formal languages initially proposed by Chomsky may overlap with a hierarchy of cognitive complexity.

Marginal to our two research questions, the present study brought methodological advances related to the use of eye tracking in AGL experiments. First, our previous study (Silva et al., 2017) was novel in showing eye-tracking signatures of implicitly acquired knowledge.

Since then, to our knowledge, no other study attempted to replicate it. In the present study, the eye-movement signatures of violation detection in the nested grammar were highly convergent with our previous findings on adjacent dependencies (Silva et al., 2017) in that whole-trial (proportion of dwell time, fixations, and dwell/first fixation on critical letter) but not first-pass measures (first fixation duration) signaled posttraining changes in discrimination in both preference and grammaticality classification tests. The fact that we replicated this pattern suggests that this may be a robust pattern of eye-tracking signatures of implicit learning. Second, the fact that nested dependencies showed no significant learning indices at the behavioral level, but did show them in eye tracking, suggests that eye-tracking measures may be more sensitive than behavioral ones. Third and final, among the three measures that captured learning of nested grammars, dwell/first fixation was the least stable: (a) it was the only measure showing significant results for crossed grammars (change from AGL2 to AGL3), (b) it lost sensitivity when NG-exclusive bigrams were removed from nested NG stimulus sets, and (c) it was no longer responsive after trials with inaccurate behavioral responses were removed. Altogether, these three sources of instability are consistent with the possibility that dwell/first fixation reflects, at least partly, influences from explicit knowledge. Concerning (a), we saw in Experiment 1 that crossed grammar learners showed changes in dwell/first fixation—but no other measure—from AGL2 over AGL3, and this co-occurred with indices that participants used to explicitly acquire knowledge to boost discrimination from AGL2 over AGL3. As for the fact that (b) dwell/first fixation reflected knowledge on all nested sequences, but it no longer did so when sequences containing non-grammatical-exclusive bigrams (chunk-based shortcuts) were excluded, this may also mean that dwell/first fixation reflects explicit knowledge—if we accept that chunk-based shortcuts represent explicit strategies (Kürten et al., 2012; but see Pothos & Wood, 2009). Finally, the fact that (c) dwell/first fixation was responsive to all sequences, but not to the subset of accurately classified ones, is not necessarily related to explicitness. However, it is possible that inaccurate responses were caused by attempts to use explicit knowledge, as has been demonstrated in some studies (e.g., Reber, 1976). The hypothesis that dwell/first fixation reflects explicit knowledge could be addressed in future studies comparing implicit with explicit learning.

References

Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience*, *14*(8), 1067–1074. https://doi.org/10.1038/nn.2869

Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes*, *1*(4), 249–262. https://doi.org/10.1080/01690968608404677

Bader, M. (2017). The limited role of number of nested syntactic dependencies in accounting for processing cost: Evidence from German simplex and complex verbal clusters. *Frontiers in Psychology*, *8*, Article 2268. https://doi.org/10.3389/fpsyg.2017.02268

Bailey, T. M., & Pothos, E. M. (2008). AGL StimSelect: Software for automated selection of stimuli for artificial grammar learning. *Behavior Research Methods*, *40*(1), 164–176. https://doi.org/10.3758/BRM.40.1.164

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., & Singmann, H. (2015). *Package ‘lme4’* [Computer software]. https://cran.rproject.org/web/packages/lme4/index.html

Beckers, G. J. L., Berwick, R. C., Okanoya, K., & Bolhuis, J. J. (2017). What do animals learn in artificial grammar studies? *Neuroscience and Biobehavioral Reviews*, *81*(Part B), 238–246. https://doi.org/10.1016/j.neubiorev.2016.12.021

Calmus, R., Wilson, B., Kikuchi, Y., & Petkov, C. I. (2020). Structured sequence processing and combinatorial binding: Neurobiologically and computationally informed hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1791), Article 20190304. https://doi.org/10.1098/rstb.2019.0304

Chomsky, N. (1956). Three models for the description of language. *IEEE Transactions on Information Theory*, *2*(3), 113–124. https://doi.org/10.1109/TIT.1956.1056813

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Christiansen, M. H., & Chater, N. (2015). The language faculty that wasn’t: A usage-based account of natural language recursion. *Frontiers in Psychology*, *6*, Article 1182. https://doi.org/10.3389/fpsyg.2015.01182

Christiansen, M. H., Louise Kelly, M., Shillcock, R. C., & Greenfield, K. (2010). Impaired artificial grammar learning in agrammatism. *Cognition*, *116*(3), 382–393. https://doi.org/10.1016/j.cognition.2010.05.015

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees. *Neuron*, *88*(1), 2–19. https://doi.org/10.1016/j.neuron.2015.09.019

de Vries, M., Christiansen, M. H., & Petersson, K. M. (2011). Learning recursion: Multiple nested and crossed dependencies. *Biolinguistics*, *5*(1–2), 10–35. https://doi.org/10.5964/bioling.8825

de Vries, M. H., Monaghan, P., Knecht, S., & Zwitserlood, P. (2008). Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures. *Cognition*, *107*(2), 763–774. https://doi.org/10.1016/j.cognition.2007.09.002

de Vries, M. H., Petersson, K. M., Geukes, S., Zwitserlood, P., & Christiansen, M. H. (2012). Processing multiple non-adjacent dependencies: Evidence from sequence learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 2065–2076. https://doi.org/10.1098/rstb.2011.0414

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fitch, W. T., & Friederici, A. D. (2012). Artificial grammar learning meets formal language theory: An overview. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 1933–1955. https://doi.org/10.1098/rstb.2012.0103

Fitch, W. T., Friederici, A. D., & Hagoort, P. (2012). Pattern perception and computational complexity: Introduction to the special issue. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 1925–1932. https://doi.org/10.1098/rstb.2012.0099

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

Folia, V., & Petersson, K. M. (2014). Implicit structured sequence learning: An fMRI study of the structural mere-exposure effect. *Frontiers in Psychology*, *5*, Article 41. https://doi.org/10.3389/fpsyg.2014.00041

Folia, V., Uddén, J., Forkstam, C., Ingvar, M., Hagoort, P., & Petersson, K. M. (2008). Implicit learning and dyslexia. *Annals of the New York Academy of Sciences*, *1145*(1), 132–150. https://doi.org/10.1196/annals.1416.012

Forkstam, C., Elwér, A., Ingvar, M., & Petersson, K. M. (2008). Instruction effects in implicit artificial grammar learning: A preference for grammaticality. *Brain Research*, *1221*, 80–92. https://doi.org/10.1016/j.brainres.2008.05.005

Forkstam, C., Hagoort, P., Fernandez, G., Ingvar, M., & Petersson, K. M. (2006). Neural correlates of artificial syntactic structure classification. *NeuroImage*, *32*(2), 956–967. https://doi.org/10.1016/j.neuroimage.2006.03.057

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(7), 2458–2463. https://doi.org/10.1073/pnas.0509389103

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, *440*(7088), 1204–1207. https://doi.org/10.1038/nature04675

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*(2), 183–206. https://doi.org/10.1207/s15327078in0702_4

Hochmann, J.-R., Azadpour, M., & Mehler, J. (2008). Do humans really learn A(n) B(n) artificial grammars from exemplars? *Cognitive Science*, *32*(6), 1021–1036. https://doi.org/10.1080/03640210801897849

Jäger, G., & Rogers, J. (2012). Formal language theory: Refining the Chomsky hierarchy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 1956–1970. https://doi.org/10.1098/rstb.2012.0077

Jiang, X., Long, T., Cao, W., Li, J., Dehaene, S., & Wang, L. (2018). Production of supra-regular spatial sequences by macaque monkeys. *Current Biology*, *28*(12), 1851–1859.e4. https://doi.org/10.1016/j.cub.2018.04.047

Joshi, A. K. (1985). How much context sensitivity is necessary for characterizing structural descriptions: Tree adjoining grammars. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational and theoretical perspectives* (pp. 206–250). Cambridge University Press. https://doi.org/10.1017/CBO9780511597855.007

Joshi, A. K. (1990). Processing crossed and nested dependencies: An automation perspective on the psycholinguistic results. *Language and Cognitive Processes*, *5*(1), 1–27. https://doi.org/10.1080/01690969008402095

Kaan, E., & Vasić, N. (2004). Cross-serial dependencies in Dutch: Testing the influence of NP type on processing load. *Memory & Cognition*, *32*(2), 175–184. https://doi.org/10.3758/BF03196850

Kürten, J., De Vries, M. H., Kowal, K., Zwitserlood, P., & Flöel, A. (2012). Age affects chunk-based, but not rule-based learning in artificial grammar acquisition. *Neurobiology of Aging*, *33*(7), 1311–1317. https://doi.org/10.1016/j.neurobiolaging.2010.10.008

Lai, J., & Poletiek, F. H. (2011). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition*, *118*(2), 265–273. https://doi.org/10.1016/j.cognition.2010.11.011

Lenhard, W., & Lenhard, A. (2014). *Hypothesis tests for comparing correlations*. Psychometrica. https://www.psychometrica.de/correlation.html

Ling, X., Li, F., Qiao, F., Guo, X., & Dienes, Z. (2016). Fluency expresses implicit knowledge of tonal symmetry. *Frontiers in Psychology*, *7*, Article 57. https://doi.org/10.3389/fpsyg.2016.00057

Manza, L., & Bornstein, R. F. (1995). Affective discrimination and the implicit learning process. *Consciousness and Cognition*, *4*(4), 399–409. https://doi.org/10.1006/ccog.1995.1047

Martins, A. (2006, March). *Emphatic verb reduplication in European Portuguese. European Dialect Syntax Project* [Paper presentation]. Workshop on Syntactic Doubling, Amsterdam. http://www.meertens.knaw.nl/projecten/edisyn/Online_proceedings/Paper_Martins.pdf

Nakamura, M., & Miyamoto, E. T. (2006). Crossed dependencies and plausibility factors in the interpretation of double-gap relative clauses in Japanese. *Cognitive Studies*, *13*(3), 369–391. https://doi.org/10.11225/jcss.13.369

Newell, B. R., & Bright, J. E. H. (2001). The relationship between the structural mere exposure effect and the implicit learning process. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *54*(4), 1087–1104. https://doi.org/10.1080/713756009

Ojima, S., & Okanoya, K. (2014). The non-hierarchical nature of the Chomsky hierarchy-driven artificial-grammar learning. *Biolinguistics*, *8*, 163–180. https://doi.org/10.5964/bioling.8997

Öttl, B., Jäger, G., & Kaup, B. (2015). Does formal complexity reflect cognitive complexity? Investigating aspects of the Chomsky hierarchy in an artificial language learning study. *PLoS ONE*, *10*(4), Article e0123059. https://doi.org/10.1371/journal.pone.0123059

Öttl, B., Jäger, G., & Kaup, B. (2017). The role of simple semantics in the process of artificial grammar learning. *Journal of Psycholinguistic Research*, *46*(5), 1285–1308. https://doi.org/10.1007/s10936-017-9494-y

Perruchet, P., & Rey, A. (2005). Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin & Review*, *12*(2), 307–313. https://doi.org/10.3758/BF03196377

Petersson, K. M., Elfgren, C., & Ingvar, M. (1999a). Dynamic changes in the functional anatomy of the human brain during recall of abstract designs related to practice. *Neuropsychologia*, *37*(5), 567–587. https://doi.org/10.1016/S0028-3932(98)00152-3

Petersson, K. M., Elfgren, C., & Ingvar, M. (1999b). Learning-related effects and functional neuroimaging. *Human Brain Mapping*, *7*(4), 234–243. https://doi.org/10.1002/(SICI)1097-0193(1999)7:4<234::AID-HBM2>3.0.CO;2-O

Petkov, C. I., & Ten Cate, C. (2020). Structured sequence learning: Animal abilities, cognitive operations, and language evolution. *Topics in Cognitive Science*, *12*(3), 828–842. https://doi.org/10.1111/tops.12444

Piechnik, I. (2015). Reduplicative syllables in Romance languages. *Romanica Cracoviensia*, *15*(1), 30–55. https://doi.org/10.4467/20843917RC.15.003.4000

Poletiek, F. H., & Lai, J. (2012). How semantic biases in simple adjacencies affect learning a complex structure with non-adjacencies in AGL: A statistical account. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 2046–2054. https://doi.org/10.1098/rstb.2012.0100

Pothos, E. M., & Wood, R. L. (2009). Separate influences in learning: Evidence from artificial grammar learning with traumatic brain injury patients. *Brain Research*, *1275*, 67–72. https://doi.org/10.1016/j.brainres.2009.04.019

R Core Team. (2013). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing.

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

Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(1), 88–94. https://doi.org/10.1037/0278-7393.2.1.88

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*(3), 595–626. https://doi.org/10.1111/lang.12010

Rohrmeier, M., Fu, Q., & Dienes, Z. (2012). Implicit learning of recursive context-free grammars. *PLoS ONE*, *7*(10), Article e45885. https://doi.org/10.1371/journal.pone.0045885

Silva, S., Folia, V., Inácio, F., Castro, S. L., & Petersson, K. M. (2018). Modality effects in implicit artificial grammar learning: An EEG study. *Brain Research*, *1687*, 50–59. https://doi.org/10.1016/j.brainres.2018.02.020

Silva, S., Inácio, F., Folia, V., & Petersson, K. M. (2017). Eye movements in implicit artificial grammar learning. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *43*(9), 1387–1402. https://doi.org/10.1037/xlm0000350

Stabler, E. P. (2004). Varieties of crossing dependencies: Structure dependence and mild context sensitivity. *Cognitive Science*, *28*(5), 699–720. https://doi.org/10.1207/s15516709cog2805_4

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Stobbe, N., Westphal-Fitch, G., Aust, U., & Fitch, W. T. (2012). Visual artificial grammar learning: Comparative research on humans, kea (Nestor notabilis) and pigeons (Columba livia). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 1995–2006. https://doi.org/10.1098/rstb.2012.0096

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*(1), Article 21. https://doi.org/10.1186/1471-2202-10-21

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

Uddén, J., Araújo, S., Forkstam, C., Ingvar, M., Hagoort, P., & Petersson, K. M. (2009). A matter of time: Implicit acquisition of recursive sequence structures. In N. Taatgen & H. Van Rijn (Eds.), *Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society* (pp. 2444–2449). Cognitive Science Society.

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

Uddén, J., Ingvar, M., Hagoort, P., & Petersson, K. M. (2012). Implicit acquisition of grammars with crossed and nested non-adjacent dependencies: Investigating the push-down stack model. *Cognitive Science*, *36*(6), 1078–1101. https://doi.org/10.1111/j.1551-6709.2012.01235.x

Uddén, J., Ingvar, M., Hagoort, P., & Petersson, K. M. (2017). Broca's region: A causal role in implicit processing of grammars with crossed non-adjacent dependencies. *Cognition*, *164*, 188–198. https://doi.org/10.1016/j.cognition.2017.03.010

van Heijningen, C. A. A., de Visser, J., Zuidema, W., & ten Cate, C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(48), 20538–20543. https://doi.org/10.1073/pnas.0908113106

Westphal-Fitch, G., Giustolisi, B., Cecchetto, C., Martin, J. S., & Fitch, W. T. (2018). Artificial grammar learning capabilities in an abstract visual task match requirements for linguistic syntax. *Frontiers in Psychology*, *9*, Article 1210. https://doi.org/10.3389/fpsyg.2018.01210

Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., van der Kant, A., Smith, K., & Rey, A. (2020). Non-adjacent dependency learning in humans and other animals. *Topics in Cognitive Science*, *12*(3), 843–858. https://doi.org/10.1111/tops.12381

Ziori, E., Pothos, E. M., & Dienes, Z. (2014). Role of prior knowledge in implicit and explicit learning of artificial grammars. *Consciousness and Cognition*, *28*, 1–16. https://doi.org/10.1016/j.concog.2014.06.003

Appendix A

Nonadjacent-Part Bigrams

	CR-train		CR-G		CR-NG		NE-train		NE-G		NE-NG	
Bigram	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
DD	48	0.12	40	0.09	47	0.11	43	0.12	43	0.11	43	0.12
DF	30	0.08	55	0.13	41	0.10	24	0.06	58	0.15	28	0.08
DL	27	0.07	33	0.08	22	0.05	0	0.00	0	0.00	32	0.09
DP	37	0.09	12	0.03	31	0.07	59	0.16	26	0.07	26	0.07
FD	35	0.09	42	0.10	40	0.10	30	0.08	21	0.05	47	0.13
FF	37	0.09	31	0.07	33	0.08	38	0.10	22	0.06	22	0.06
FL	14	0.04	5	0.01	33	0.08	41	0.11	70	0.18	13	0.04
FP	22	0.06	46	0.11	10	0.02	0	0.00	0	0.00	22	0.06
LL	37	0.09	31	0.07	53	0.13	38	0.10	22	0.06	40	0.11
LP	35	0.09	42	0.10	41	0.10	24	0.06	58	0.15	31	0.09
PL	30	0.08	55	0.13	40	0.10	30	0.08	21	0.05	41	0.11
PP	48	0.12	40	0.09	27	0.06	43	0.12	43	0.11	19	0.05
Count	400		432		418		370		384		364	

Note. CR = crossed; NE = nested; G = grammatical; NG = nongrammatical.

(*Appendices continue*)

Appendix B

Postexperimental Questionnaire on Explicit Knowledge

Questions

Referring to Day 1 (Baseline Preference Test, AGL1)

When you responded whether you liked or not each sequence, did you do it based only on your “gut feeling,” or were you using some other criterion?

Referring to Day 9 (Final Preference Test, AGL2)

When you responded whether you liked or not each sequence, did you do it based only on your “gut feeling,” or were you using some other criterion?

Referring to Day 9 (Final Grammaticality Test, AGL3)

1. Did you notice anything special or strange about the sequences? If so, what?
2. Did you use any strategy to distinguish between grammatical and nongrammatical sequences?
3. Please write down five grammatical (correct) sequences.
4. Some letters appeared only in the inner part of the sequence, and other letters only in the tails (begin and end of sequence). Which letters appeared in the inner part? Which ones were at the tails?
5. At the inner part of the sequence, you could see the letters D, F, L, P. Do you have any idea on how they were organized? Did some letters come before others?
6. The inner part of the sequence, containing D, F, L, P, was divided into two halves. Some of these letters appeared in the first half, and others in the second half. Which letters were at the first half, and which at the second?
7. At the inner part, there was a correspondence between letters in the first half and letters in the second half. How was the correspondence?

(*Appendices continue*)

Appendix C

Eye-Tracking Results for Accurate Trials Only

Table C1

Grammatical Status (G vs. NG) × Test (AGL1 vs. AGL2) Interactions on Eye-Tracking Learning Indices for Accurate Trials

Test: AGL1 vs. AGL2	GS × Test *B*; [95% CI]; *p*; (R^2)	GS × Test × Grammar *B*; [95% CI]; *p*; (R^2)	GS × Test, Nested *B*; [95% CI]; *p*; (R^2)	GS × Test, Crossed *B*; [95% CI]; *p*; (R^2)
First fixation duration				
All NE	16.2; [−20.2, 52.69]; .383	9.06; [40.00, 58.12]; .717		
NE without 1 NA pair	16.15; [−20.3, 52.33]; .382	23.34; [−28.19, 74.88]; .375		
NE without FP/DL	16.12; [−18.63, 50.88]; .363	43.81; [−13.61, 101.22]; .135		
Proportion of dwell time				
All NE	[a]	0.04; [.01, .06]; **.013**[b]; (.091)	0.04; [.02, .06]; <**.001**; (.085)	0.01; [.01, .03]; .507
NE without 1 NA pair		0.04; [.01, .07]; **.013**; (.084)	0.04; [.02, .06]; <**.001**; (.084)	
NE without FP/DL		0.03; [.00, .07]; **.037**; (.070)	0.04; [.01, .07]; **.002**; (.057)	
Proportion of fixations				
All NE		0.03; [.01, .06]; **.015**; (.095)	0.04; [.02, .05]; <**.001**; (.094)	0.00; [−.01, .02]; .615
NE without 1 NA pair		0.03; [.01, .06]; **.024**; (.080)	0.03; [.02, .05]; <**.001**; (.093)	
NE without FP/DL		0.03; [.00, .06]: **.047**; (.073)	0.03; [.01, .06]; **.003**; (.080)	
Dwell/first fixation				
All NE	0.12; [−.17, .41]; .426	0.30; [−.09, .68]; .136		
NE without 1 NA pair	0.12; [−.16, .40]; .408	0.25; [−.15, .65]; .218		
NE without FP/DL	0.12; [−.16, .40]; .408	0.05; [−.41, .50]; .832		

Note. GS = grammatical status; G = grammatical; NG = nongrammatical; AGL1 = baseline preference; AGL2 = final preference; CI = confidence interval; NE = nested; NA = nonadjacent.

[a] GS × Test interactions are not presented when GS × Test × Grammar (nested vs. crossed) interactions were significant. [b] Significant interactions highlighted in bold.

Table C2

Grammatical Status (G vs. NG) × Test (AGL2 vs. AGL3) Interactions on Eye-Tracking Learning Indices for Accurate Trials

Test: AGL2 vs. AGL3	GS × Test *B*; [95% CI]; *p*; (R^2)	GS × Test × Grammar *B*; [95% CI]; *p*; (R^2)	GS × Test, Nested *B*; [95% CI]; *p*; (R^2)	GS × Test, Crossed *B*; [95% CI]; *p*; (R^2)
First fixation duration				
All NE	17.87; [−16.33, 52.07]; .306	8.17; [−38.26, 54.60]; .730		
NE without 1 NA pair	17.81; [−16.21, 51.84]; .305	−1.53; [−50.55, 47.49]; .951		
NE without FP/DL	17.85; [−15.48, 51.19]; .294	−23.03; [−79.39, 33.32]; .423		
Proportion of dwell time				
All NE	0.01; [−.01, .03]; .393	0.00; [−.02, .03]; .728		
NE without 1 NA pair	0.01; [−.01, .03]; .366	0.00; [−.02, .03]; .720		
NE without FP/DL	0.01; [−.01, .03]; .354	−0.01; [−.05, .02]; .362		
Proportion of fixations				
All NE	0.00; [−.02, .02]; .793	0.01; [−.01, .03]: .414		
NE without 1 NA pair	0.00; [−.01, .02]; .779	0.01; [−.01, .04]; .371		
NE without FP/DL	0.00; [−.01, .02]; .775	0.00; [−.03, .03]; .865		
Dwell/first fixation				
All NE	0.12; [−.14, .39]; .365	−0.07; [−.46, .30]; .720		
NE without 1 NA pair	0.12; [−.14, .38]; .351	−0.02; [−.36, .39]; .930		
NE without FP/DL	0.12; [−.12, .37]; .324	−0.13; [−.54, .28]; .538		

Note. GS = grammatical status; G = grammatical; NG = nongrammatical; AGL2 = final preference; AGL3 = grammaticality classification; CI = confidence interval; NE = nested; NA = nonadjacent.

Received June 27, 2021
Revision received July 10, 2022
Accepted July 24, 2022 ■