PHILOSOPHICAL TRANSACTIONS —OF—

THE ROYAL SOCIETY

B | **BIOLOGICAL SCIENCES**

Processing multiple non-adjacent dependencies: evidence from sequence learning

Meinou H. de Vries, Karl Magnus Petersson, Sebastian Geukes, Pienie Zwitserlood and Morten H. Christiansen

References	**This article cites 50 articles, 4 of which can be accessed free** http://rstb.royalsocietypublishing.org/content/367/1598/2065.full.html#ref-list-1
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. B* go to: **http://rstb.royalsocietypublishing.org/subscriptions**

Research

Processing multiple non-adjacent dependencies: evidence from sequence learning

Meinou H. de Vries[1,*], **Karl Magnus Petersson**[2–4],
Sebastian Geukes[5], **Pienie Zwitserlood**[5]
and Morten H. Christiansen[6,*]

[1]*Faculty of Psychology and Education, VU University, 1081 BT Amsterdam, The Netherlands*
[2]*Max Planck Institute for Psycholinguistics, 6525 XD Nijmegen, The Netherlands*
[3]*Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging,
6525 EN Nijmegen, The Netherlands*
[4]*Department of Psychology, IBB-CBME, University of Algarve, 8005-139 Faro, Portugal*
[5]*Department of Psychology, University of Münster, 48149 Münster, Germany*
[6]*Department of Psychology, Cornell University, Ithaca, NY 14853, USA*

Processing non-adjacent dependencies is considered to be one of the hallmarks of human language. Assuming that sequence-learning tasks provide a useful way to tap natural-language-processing mechanisms, we cross-modally combined serial reaction time and artificial-grammar learning paradigms to investigate the processing of multiple nested ($A_1A_2A_3B_3B_2B_1$) and crossed dependencies ($A_1A_2A_3B_1B_2B_3$), containing either three or two dependencies. Both reaction times and prediction errors highlighted problems with processing the middle dependency in nested structures ($A_1A_2A_3B_3\_B_1$), reminiscent of the 'missing-verb effect' observed in English and French, but not with crossed structures ($A_1A_2A_3B_1\_B_3$). Prior linguistic experience did not play a major role: native speakers of German and Dutch—which permit nested and crossed dependencies, respectively—showed a similar pattern of results for sequences with three dependencies. As for sequences with two dependencies, reaction times and prediction errors were similar for both nested and crossed dependencies. The results suggest that constraints on the processing of multiple non-adjacent dependencies are determined by the specific ordering of the non-adjacent dependencies (i.e. nested or crossed), as well as the number of non-adjacent dependencies to be resolved (i.e. two or three). Furthermore, these constraints may not be specific to language but instead derive from limitations on structured sequence learning.

Keywords: non-adjacent dependencies; sequence learning; artificial grammar learning;
serial reaction time

1. THEORETICAL BACKGROUND

The natural-language phenomenon of recursion and its potential underlying processing mechanisms have attracted great interest from scholars of different fields, including psycholinguistics, biology, computer science and cognitive neuroscience. Most research so far has focused on sequence-learning tasks, assuming that they share cognitive mechanisms with language processing. The focus in the past decade was to determine whether such tasks could capture processing differences between specific language structures, often with their processing complexity defined in terms of the Chomsky hierarchy [1]. In the current

paper, we identify two different sources of processing complexity, and we propose that processing differences are intrinsically tied to (i) the memory resources required and (ii) relevant processing experience [2]. We will empirically investigate this claim by focusing on non-adjacent dependency processing.

(a) *Non-adjacency in language*

Recursion has been suggested to be a hallmark of human language (cf. [3]). Recursion is an operation that permits a finite set of rules to generate an infinite number of expressions. In this paper, we concentrate on bounded recursive structures involving *multiple overlapping non-adjacent dependencies*. Their existence has been suggested by generative linguists to be one of the major challenges for empirically based approaches to language [4], as they may point to the limits of human language processing.

* Authors for correspondence (m.h.de.vries@vu.nl; meinou_de_vries@
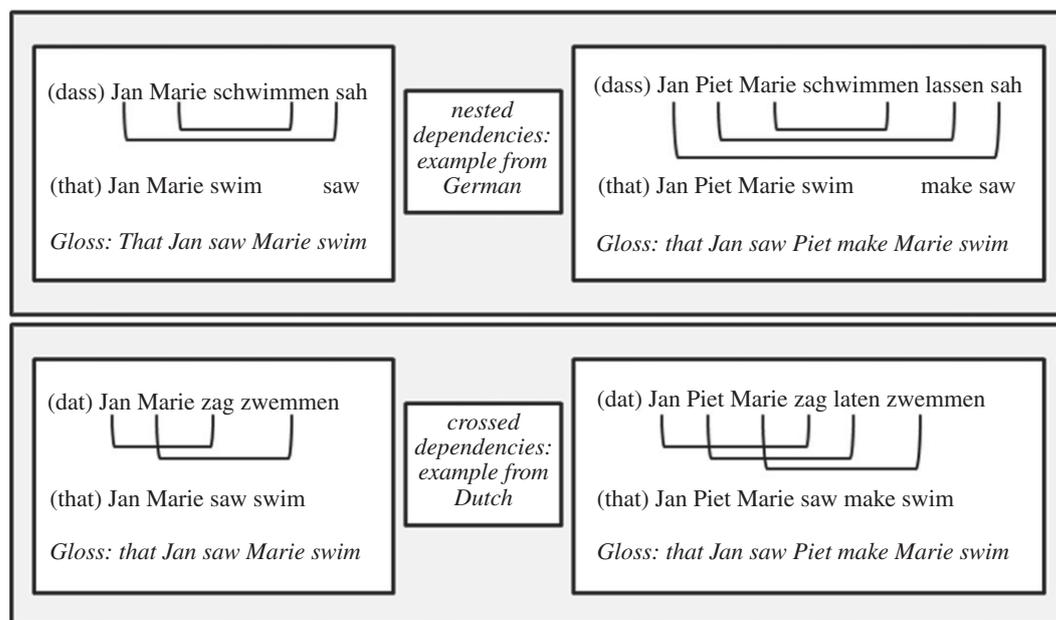hotmail.com; christiansen@cornell.edu).

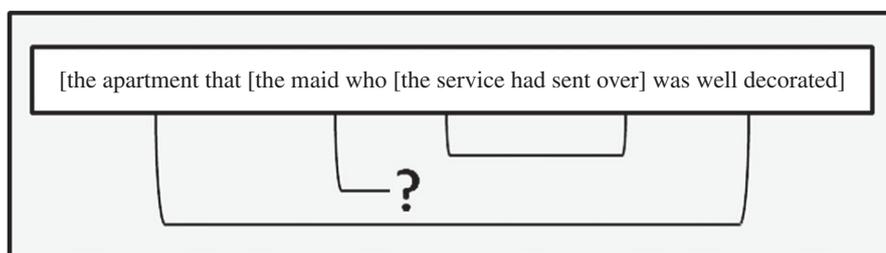Figure 1. Different types of non-adjacency.

Figure 2. The missing-verb effect.

Non-adjacent dependencies are dependencies between elements that are not contiguous. For instance, in the sentence *The dog that scared the cat ran away*, we need to link *the dog* to *ran away*, a verb phrase further down the sentence. Non-adjacency can be expressed in several ways across languages. One instantiation of non-adjacency involves *nested dependencies*, with dependencies nested within one another, exemplified in the structure $A_1A_2A_3B_3B_2B_1$, where element A_i needs to be linked to element B_i. Another instantiation involves *crossed dependencies*, where the dependencies between elements cross each other, exemplified in the structure $A_1A_2A_3B_1B_2B_3$. These types of non-adjacency are depicted in figure 1, demonstrating that non-adjacent dependencies are expressed differently across languages, in this case German and Dutch, which are otherwise closely related.

Theoretically, an infinite number of embeddings can be added; however, since we are constrained by finite brains, sentences with three or more dependencies are notoriously hard to understand [5–9]. An interesting demonstration comes from Gibson & Thomas [10], who investigated the role of memory limitations on processing sentences that contained three nested dependencies. They observed that speakers of English would rate the grammatical sentence in 1a no better that its ungrammatical counterpart in 1b, which is missing the middle verb phrase:

— 1a. The apartment that the maid who the service had sent over was cleaning every week was well decorated.
— 1b. The apartment that the maid who the service had sent over was well decorated.

This *missing-verb effect* ([10]; figure 2) is likely related to real-time memory overload as the language system needs to keep track of too many unresolved dependencies [10]. Apparently, resolving three non-adjacent dependencies puts too much load on available memory. As a consequence, our memory system does not retain certain parts of the sentence (cf. the 'syntactic prediction memory cost' account, [10]), and a missing verb is not missed.

(b) Processing consequences of differences between languages

The missing-verb effect has been replicated in several languages such as French [11] and English, with both online and offline grammaticality judgement tasks and with control for sentence length and semantic plausibility [12]. Interestingly, a recent study by Vasishth *et al.* [13] suggests that Germans are less sensitive to this effect. Using two online measures (self-paced reading and eye-tracking), they found that Germans were sensitive to the ungrammaticality of sentences such as 1b, whereas English participants were not. The authors surmised that this might be due to prior experience with verb-final sentences, which are common in German and

require keeping track of dependency relations over a relatively long distance (see also Hurford [14], for an observation that German students have less difficulties with nested dependencies than English students). The absence of the missing-verb effect may thus have been caused by adaptation to the specific grammatical properties of German, and therefore German speakers may maintain predictions about upcoming sentence parts more robustly than English speakers. This is consistent with the proposal that working-memory constraints can be shaped by prior linguistic experience [12,15]. Moreover, several natural-language studies support the importance of prior experience for the processing of multiple overlapping dependencies. For instance, a training study by Roth [16] showed that experience with non-adjacent dependencies is the key factor in preschoolers' comprehension of such structures (and not cognitive development). Also, specific experience through education affects the ability to comprehend sentences containing different types of non-adjacency [17], possibly due to differential exposure both in terms of quantity and complexity of the linguistic material [12]. As for cross-linguistic data, it has been found that nested non-adjacent dependencies are more easily processed in Spanish than in English [18], and that there are substantial differences in perceived processing difficulty for the exact same types of non-adjacency in English, German, Japanese and Persian [19].

Could it be the case that the multiple nested and crossed dependencies are processed differently, depending on their prevalence in the language that one has been exposed to? For example, it has been claimed that some languages lack recursion altogether [20]—though see [21,22], and [14] for further discussion on these issues. The observed linguistic diversity suggests that language processing is not identical in all humans, but perhaps rather depends on the memory constraints that are shaped by, among other things, linguistic experience. As illustrated in figure 1, non-adjacent dependencies are expressed differently in German (where nested dependencies are more common) and Dutch (where crossed dependencies are more common), such that speakers of both languages have different linguistic experience. This may lead to cross-linguistic differences in processing nested and crossed dependencies. The aim of the current paper is, among other things, to investigate if a sequence-learning paradigm can produce an *analogue* to the missing-verb effect, assessed in both German and Dutch participants. Given the prevalent crossed-dependency structure of their language, Dutch participants should be more susceptible to our missing-verb-effect analogue than Germans.

(c) *Processing differences between nested and crossed dependencies*

Another source of processing differences may relate to the nested and crossed dependencies themselves, which differ qualitatively as a function of their inherent complexity. A seminal psycholinguistic study by Bach *et al.* [5] directly investigated the effect of complexity differences between nested and crossed dependencies on natural-language processing. Native German speakers provided comprehensibility ratings of German

sentences containing nested dependencies, and native Dutch speakers rated Dutch sentences containing crossed dependencies (see the examples of figure 1). The results indicated that crossed dependencies are easier to process than nested, but only when the sentences contained more than two dependencies. This pattern of results was replicated by Christiansen & MacDonald [12], who modelled the relative difficulty of nested versus crossed dependencies by training a simple recurrent network (SRN) [23] on sentences that included such dependencies. Their SRNs exhibited the same processing difficulties found in humans: crossed dependencies were easier to process, but only in the case of three dependencies, not of two (see also [24] for similar results).

A possible explanation as to why crossed dependencies are found easier to process than nested ones is offered by the Syntactic Prediction Locality Theory (SPLT) [25]. The SPLT associates different memory costs with different incomplete non-adjacent dependencies, primarily in proportion to the distance since the incomplete dependency was first encountered. Indeed, it has been shown that it is harder to retain items in short-term memory as more interfering items are processed (cf. [25]). Thus, according to SPLT, nested dependencies involve higher maximal complexity than crossed ones. The longest dependency is established earlier in crossed-dependency structures than in nested dependencies, resulting in lower overall complexity for crossed dependencies. As a purely memory-based account, however, SPLT cannot provide a complete explanation of the processing of non-adjacent dependencies because it does not capture the earlier-noted effects of experience and differences in the processing difficulty associated with the same linguistic structure across different languages.

In an extensive review, De Vries *et al.* [2] put forward an empirically based view on processing complexity, stressing the importance of this distinction between sentences with two (or less) and three (or more) dependencies. Thus, the processing benefit of crossed dependencies becomes relevant *only* when the number of to-be-resolved dependencies exceeds two. In sum, two important determinants of processing complexity are (i) the order of the non-adjacent dependencies (nested or crossed) and (ii) the number of the non-adjacent dependencies that need to be resolved simultaneously (two or three). Figure 3 provides a schematic of the different processing complexity levels [2].

Our current study implements an *analogue* of the missing-verb effect in a sequence-learning paradigm. We predict that there will be no 'missing-verb' effect when crossed dependencies have to be formed. Indeed, assuming that the missing-verb effect is caused by memory (over)load *and* that crossed dependencies are easier to process than nested, situations where crossed dependencies have to be established should be less prone to the missing-verb effect. This should be different for nested dependencies with more than two dependencies.

(d) *Goal of the current study*

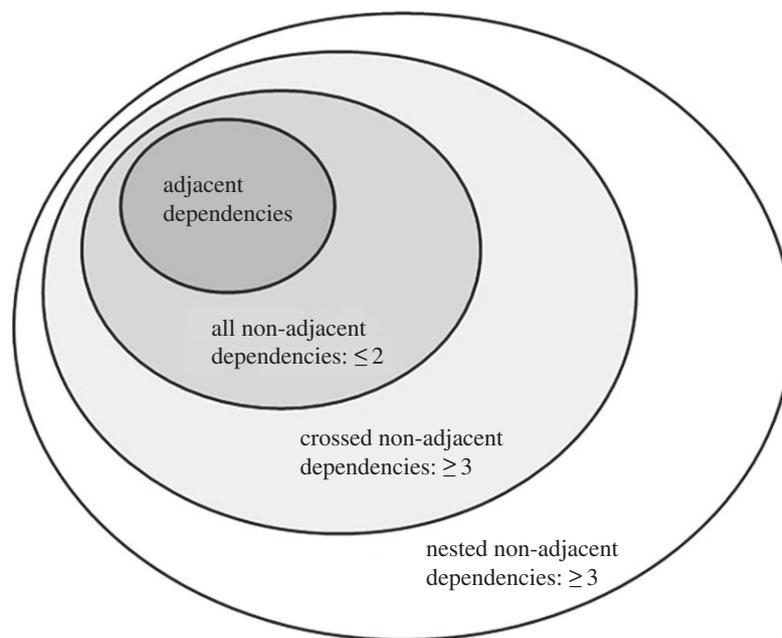To summarize, on the basis of the assumptions that processing differences stem from (i) memory load,

Figure 3. A schematic of the suggested levels of processing complexity—see also [2]. Each level of the hierarchy denotes a decisive factor that adds to processing complexity. Thus, it is not the case that three nested dependencies are harder to process than, say, six crossed dependencies. Instead, this figure emphasizes that the presence of more than two dependencies is a sine qua non condition for measurable differences in processing complexity between nested and crossed dependencies.

such that crossed dependencies are easier to process than nested, but only when the number of dependencies exceeds two, and (ii) processing experience, such that experience with a specific type of non-adjacency benefits the processing of structures similar to that specific non-adjacency type, we advance the following predictions with respect to the missing-verb-effect analogue:

— the missing-verb effect is specific to nested dependencies, and will be less strong or absent in crossed dependencies;
— the missing-verb effect in nested-dependency situations is specific to native speakers of Dutch, as Germans may benefit from their prior experience with nested dependencies; and
— the missing-verb effect is absent in cases of two nested or crossed dependencies, and no differences will be observed between them, for both German and Dutch participants.

2. MEASURING PROCESSING DIFFERENCES WITH SEQUENCE-LEARNING TASKS

Language processing crucially involves the extraction of regularities from highly complex sequential input, as the relations between units such as words, syllables and morphemes adhere to structural characteristics typical of language [26]. This may point to a clear similarity between sequence learning and language, as both involve the extraction and processing of discrete elements occurring in complex structured sequences [27]. Different sequence-learning paradigms such as *serial reaction time* (SRT) [28] and *artificial-grammar learning* (AGL) [29] all have in common that they involve extracting sequence regularities (for a review see, [30]). In the SRT

paradigm, participants typically have to respond to each element in a fixed, repeating sequence. The reaction times of these responses are measured. As training progresses, participants begin to anticipate upcoming elements, resulting in decreased reaction times. When participants are then presented with a randomly ordered sequence, reaction times increase again. In contrast, AGL is a paradigm that usually involves training on a set of symbol sequences generated by a set of rules. After a training phase, participants are asked to decide for a set of novel sequences if they are 'grammatical', i.e. following the same set of rules as the training sequences. The AGL paradigm has been used to assess the neural correlates of sequence learning, by means of functional neuroimaging (e.g. [31–33]; for an overview, see [34]), brain stimulation [35,36] and special populations such as individuals with Parkinson's disease [37,38], autism spectrum disorders [39], agrammatic aphasia [40] and dyslexia (e.g. [41,42]; for a review, see [43]). These studies have pointed to a general involvement of frontal–striatal–cerebellar circuits [44,45] that are also involved in the acquisition of grammatical regularities [45]. This suggests that such sequence-learning tasks are useful to investigate natural-language processing [34].

Sequence-learning tasks are often used to investigate the learning of non-adjacent dependencies, which are assumed to be harder to learn than adjacent dependencies [46–48]. Gomez [49] showed that the relative variability of the intervening material in sequences such as AXB, where X denotes the intervening material and A needs to be linked to B, determines the degree to which non-adjacent dependencies are learned, both in adults and in infants. When X is varied to a large degree, the dependency between A and B stands out, which benefits learning. Moreover, Newport & Aslin [47] and Onnis *et al.* [50] found that the similarity

between A and B items may further facilitate the learning of non-adjacent dependencies. Thus, the learning of non-adjacent dependencies is possible, but under more restrictive conditions than for adjacent dependencies [51]. In a recent study, Misyak *et al.* [52] used a combined SRT–AGL paradigm, where participants were presented with a grid on a computer screen with six non-sense words. They would then hear a sequence of three non-sense words (following the AXB rule), and their task was to click as fast as possible on each of the corresponding non-sense words on the screen. The results demonstrated that individual differences in learning non-adjacent dependencies strongly correlate with the processing of natural-language sentences containing complex non-adjacent dependencies in the form of nested relative clauses. This again suggests a strong link between sequence-learning tasks and natural-language processing (for a review, see [53]).

The combined SRT–AGL task combines the strengths of both SRT and AGL paradigms: the SRT paradigm uses reaction times as the dependent variable, which makes it possible to study the exact nature of the learning trajectory of non-adjacent dependencies, and as such may be more informative than standard AGL tasks. The benefits of AGL, compared with SRT, are the language-like nature of the sequences, the smaller number of training exemplars and the greater transparency compared to natural-language structure [52]. In the current study, we implemented a combined SRT–AGL task, assessing both nested and crossed-dependency structures, containing two or three dependencies.

3. METHODS
(a) *Participants*
A total of 136 participants were included, 68 native speakers of Dutch (41 female, students from the Radboud University Nijmegen, The Netherlands), and 68 native speakers of German (55 female, students from the Westfälische-Wilhelms-Universität Münster, Germany). All participants signed informed consent forms and were paid for their participation. Each participant was assigned to one of the experimental conditions, resulting in the following design:

language	ordering of dependencies	number of dependencies	n
German	nested	2	17
German	nested	3	17
German	crossed	2	17
German	crossed	3	17
Dutch	nested	2	17
Dutch	nested	3	17
Dutch	crossed	2	18
Dutch	crossed	3	16

(b) *Materials*
A training block and a recovery block, each with 48 unique syllable strings, were created. Depending on the experimental condition, these training strings denoted either nested or crossed dependencies. The

dependencies were marked by shared vowels, such that a crossed dependencies string $A_1A_2A_3B_1B_2B_3$ is exemplified by ba-no-mi-la-yo-di for a sequence with three dependencies, or no-mi-yo-di for a sequence with two dependencies. The syllables used were the same as those used by Fitch & Hauser [54] but forming the following pairs: [ba, la], [yo, no], [mi, di] and [wu, tu]. The syllables were spoken by a female voice with durations ranging from 300 to 500 ms and could occur as both A and B elements. Additionally, two blocks of 30 violations were created: 10 with the first dependent element violated (i.e. $A_1A_2A_3\_B_2B_3$ for crossed, and $A_1A_2A_3\_B_2B_1$ for nested dependencies, where the _ was replaced by a syllable that would occur in one of the subsequent B-positions (for the outer position, this was not possible; so we used either a syllable that had occurred in the sequence already, or a syllable that had not been presented in that sequence yet), or by a beep tone), 10 with the middle dependent element violated, and 10 with the last dependent element violated.

(c) *Procedure*
The paradigm included the training block (48 sequences), the violation block, (30 sequences), the recovery block (48 sequences) and, finally, the violation block for the beep task (30 sequences).

Participants were seated in a sound-proof booth in front of a computer screen. A target syllable was presented through headphones, and simultaneously a target and a foil were presented on the computer screen in a vertical column format (figure 4), using the presentation experiment control program (www.neurobs.com). Participants were instructed to click as fast as possible on the syllable that they just had heard through headphones, using the computer mouse. For the beep task, they were told that when they heard a beep, they had to click on the syllable they thought was replaced by the beep. For all violations, the correct syllable was always displayed as an option in the response pair. The positions of all syllables (i.e. in the upper or lower box) were counter-balanced, such that the structure could not be learned from the position within the response pair. See figure 4 for an example of a training sequence, a violating sequence and a beep sequence. Participants were not informed about the patterns underlying the sequences.

Between each sequence, there was an interval of 500 ms. The whole experiment lasted about 35 min.

(d) *Analysis*
SPSS 17.0 was used for the statistical analyses. For the analysis with reaction time as the dependent variable, the 48 training sequences were divided into two blocks to analyse the learning trajectory. Block 1 was the first half of the training (i.e. sequences 1–24); block 2 the second half of the training (i.e. sequences 25–48); block 3 contained all violations (i.e. 30 sequences). Only the reaction times on elements of the second half of the sequence were included in the analysis, as they indicate the establishing of the particular dependencies. Thus, in a sequence of the form $A_1A_2A_3B_1B_2B_3$, only the reaction times on
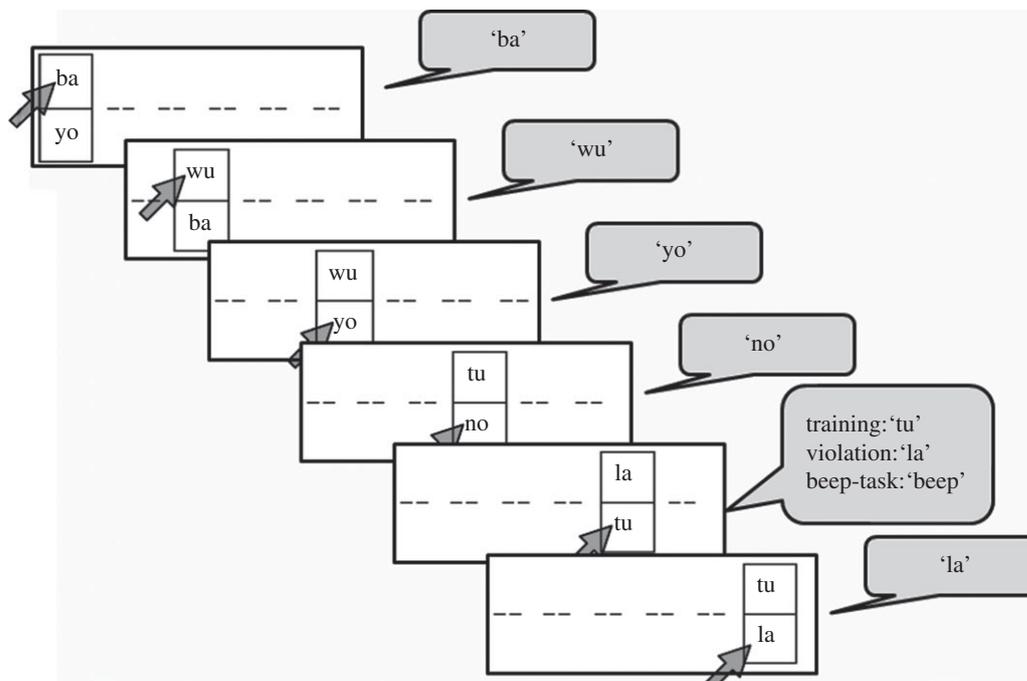
Figure 4. A schematic of the experimental task and possible sequences in the condition involving three nested dependencies. The arrow indicates the correct mouse click upon hearing a syllable through the headphones (here, presented in the text balloons).

B-elements were taken into account. These B-elements are referred to by their first, middle and last positions in the string. Two mixed ANOVAs were performed: one for three-dependency sequences, with block (1,2,3) and position (first, middle, last) as within-subjects factors, and structure (nested, crossed) and language (German, Dutch) as between-subjects factors; and one ANOVA for two-dependency sequences, with only two levels of position (first, last). For the beep task, only those sequences that included a beep (i.e. excluding the recovery block) were analysed, with the number of errors as the dependent variable. This resulted in a mixed ANOVA with position (first, middle, last) as within-subjects factor, and structure (nested, crossed) and language (German, Dutch) as between-subjects factors in the case of three dependencies, and, again, with only two levels (first, last) of the factor position in the case of two dependencies.

4. RESULTS
(a) *Three-dependency sequences: reaction times*
The mixed ANOVA with block (1,2,3) and position (first, middle, last) as within-subjects factors, and structure (nested, crossed) and language (German, Dutch) as between-subjects factors, with reaction times as dependent variable, revealed a significant main effect of block ($F_{2,128} = 14.51$, $p < 0.001$), and a significant main effect of position ($F_{2,128} = 8.50$, $p < 0.01$). Furthermore, significant interactions were observed for position $\times$ block ($F_{4,256} = 6.39$, $p < 0.001$), position $\times$ structure ($F_{2,128} = 19.14$, $p < 0.001$), and position $\times$ block $\times$ structure ($F_{4,256} = 26.38$, $p < 0.001$). Figure 5a illustrates these interactions, showing that the reaction times for the middle position in nested sequences were not affected by the violations (as was also the case for the missing-

verb effect discussed in §1). This was different for crossed dependency sequences, where reaction times at all positions increased in the violation block. This was supported by a separate analysis on the violation block only. Here, post-hoc paired *t*-tests showed that the middle position differed significantly from the first and last positions, irrespective of native language, for the nested sequences only ($p < 0.001$ for both first versus middle and last versus middle, Bonferroni-corrected with $\alpha = 0.0125$). As for the between-subjects factor, a significant main effect of language ($F_{1,64} = 8.18$, $p < 0.01$) was observed. This was caused by overall faster reaction times of the German compared with the Dutch participants (average 671.80 ms for Dutch participants, and 620.26 ms for German participants), and did not interact with the factors block, position and structure.

The earlier-mentioned analysis included all correct responses elicited by *all* training and violation trials per sequence position; thus, violations at a specific position could have affected responses in the subsequent positions in that trial. We ran a second analysis on only those responses elicited by the specific violated position, while eliminating the responses at the other positions in that specific sequence. This resulted in a main effect of position ($F_{2,128} = 5.65$, $p < 0.01$), a main effect of block ($F_{2,128} = 14.35$, $p < 0.001$), a main effect of language ($F_{1,64} = 6.42$, $p < 0.05$), an interaction between position and structure ($F_{2,128} = 7.65$, $p = 0.001$) and an interaction for position, structure and block ($F_{4,256} = 28.58$, $p < 0.001$). Post-hoc paired *t*-tests for the specific positions per sequence type revealed significant differences between the middle and last position for nested sequences (Bonferroni-corrected with $\alpha = 0.0125$; $t_{33} = -3.24$, $p < 0.01$), and between the first and middle and middle and last positions for crossed sequences ($\alpha = 0.0125$;
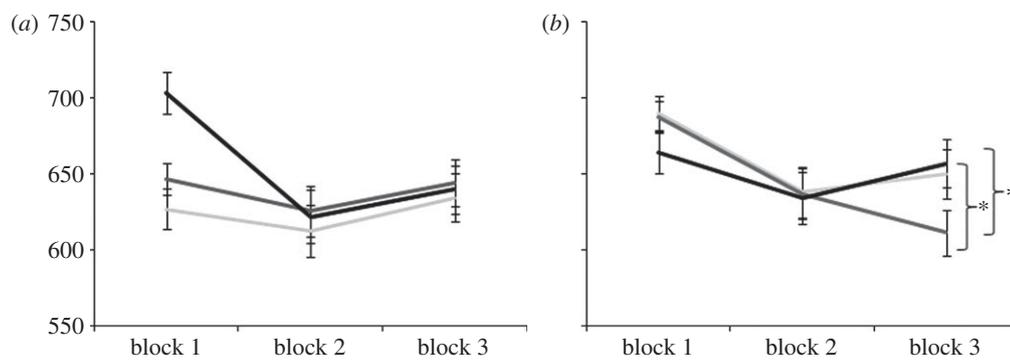
Figure 5. Reaction times in milliseconds for each syllable position (light grey lines, first; dark grey lines, middle; black lines, last), in each experimental block (1, 2, 3), for both (*a*) crossed and (*b*) nested dependencies, irrespective of the participants' native language. Error bars indicate s.e. of the mean.
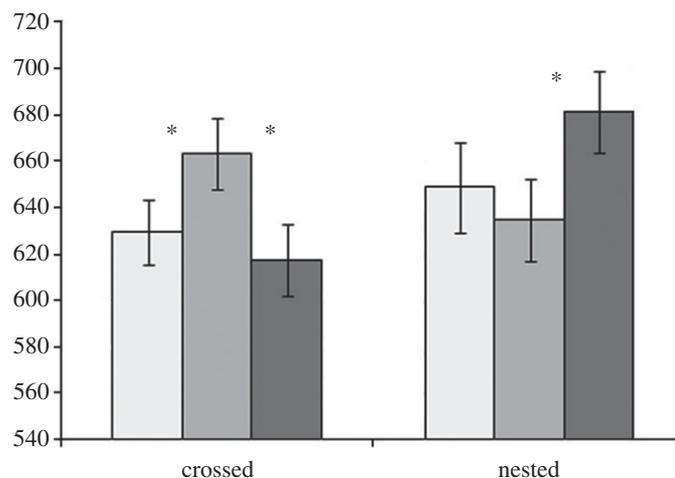

Figure 6. Reaction times in milliseconds per sequence type (crossed or nested) per violated sequence position (first (light grey bars), middle (mid-grey bars) and last (dark grey bars) position), collapsed over language. Error bars denote s.e. of the mean.

$t_{33} = -3.59$, $p = 0.001$ and $t_{33} = 4.46$, $p < 0.001$, respectively). These post-hoc paired *t*-tests indicate that violations at the middle (missing verb) position for crossed sequences elicited slower responses than did those at the two other positions. For nested sequences, responses to violations in the middle position were faster compared with the last, but not with the first position (figure 6).

(b) *Three-dependency sequences: beep task*

The above analysis took reaction times as a dependent variable. The second task of our experiment was the beep task, where participants had to choose the syllable that they thought was replaced by the beep. Here, the number of errors is the dependent variable. We expected the pattern of results to reflect the above pattern of reaction times: for nested-dependency sequences, errors should be more frequent on the middle position than on the other positions, as participants will tend to neglect this position and instead opt for the alternative syllable (which was the syllable that would come at the last position). A mixed ANOVA with position (first, middle, last) as within-subjects factor, and structure (nested, crossed) and language (German, Dutch) as between-subjects factors revealed main effects of position ($F_{2,128} = 23.03$, $p < 0.001$) and structure ($F_{1,64} = 25.28$, $p < 0.001$), and a

significant interaction between these two ($F_{2,128} = 8.92$, $p < 0.001$), indicating that for nested structures, the error pattern was indeed, as we had expected, based on the reaction time data: most errors were elicited by beeps on the middle position, which significantly differed from the error rate on the last and first elements (post-hoc paired *t*-tests, Bonferroni-corrected with $\alpha = 0.0125$; first versus middle position, $t_{33} = -4.04$, $p < 0.001$, middle versus last position, $t_{33} = 4.40$, $p < 0.001$). For crossed structures, this was not the case (post-hoc paired *t*-tests, Bonferroni-corrected with $\alpha = 0.0125$; first versus middle position, $t_{33} = 1.33$, $p = 0.19$, middle versus last position, $t_{33} = 5.50$, $p < 0.001$; see also figure 8).

(c) *Two-dependency sequences: reaction times*

After removing one outlier (more than three times s.d. (95.62 ms) with an average reaction time of 899 ms), the mixed ANOVA with block (1,2,3) and position (first, last) as within-subjects factors, and structure (nested, crossed) and language (German, Dutch) as between-subjects factors revealed a main effect of block (Greenhouse–Geisser corrected, $F_{2,128} = 44.38$, $p < 0.001$). Furthermore, a significant position × block interaction ($F_{2,128} = 14.73$, $p < 0.001$) was found (figure 7*a*). As for the between-subjects effects,
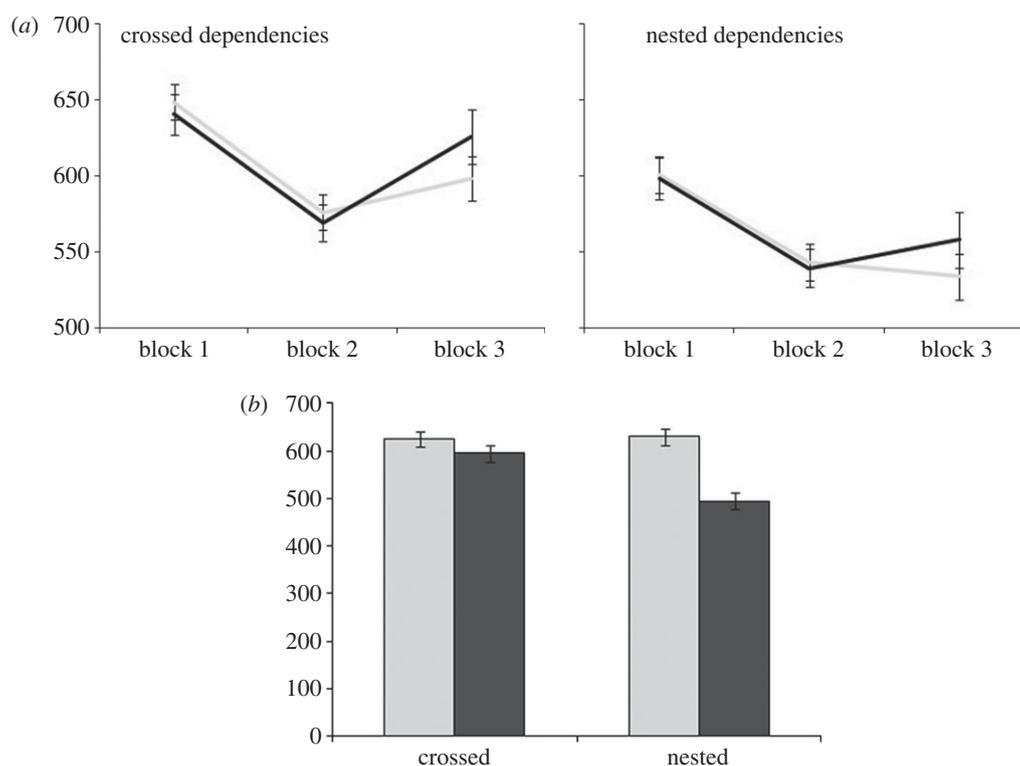
Figure 7. (a) Reaction times in milliseconds for each syllable position (grey lines, first; black lines, last), in each experimental block (1, 2, 3), for both nested and crossed dependencies, irrespective of the participants' mother tongue. (b) Reaction times in milliseconds for language (dark grey bars, German; light grey bars, Dutch) and structure (nested and crossed dependencies). The Germans responded faster when encountering a nested structure compared with a crossed structure. This difference is not seen in the Dutch participants. Error bars represent s.e. of the mean.
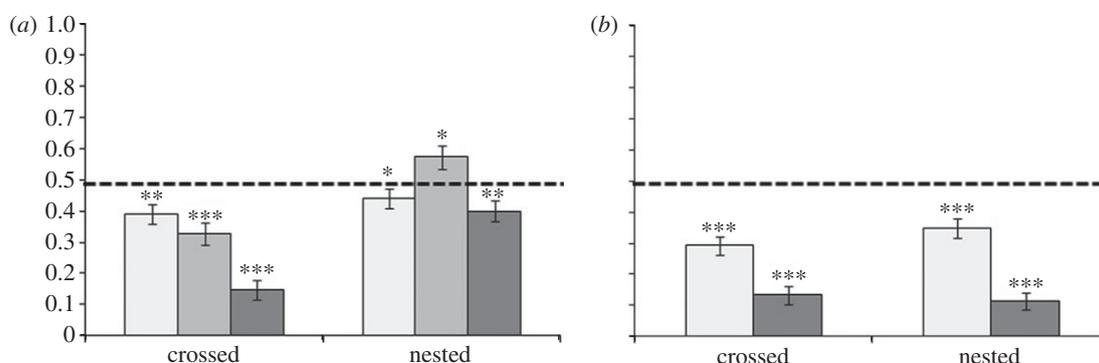

Figure 8. Error rates per sequence position in both crossed and nested structures with (a) three and (b) two dependencies. Error bars denote s.e. of the mean. The dotted line denotes the chance level (0.5). Significance levels are denoted by asterisks, $*p < 0.05$, $**p < 0.01$ and $***p < 0.001$. Light grey bars, first; mid-grey bars, middle; dark grey bars, last.

a main effect of language was observed ($F_{1,64} = 23.20$, $p < 0.01$), again due to the significantly faster overall reaction times of the German (mean 544.75 ms) compared with the Dutch participants (mean 627.69 ms). A significant interaction was found for structure × language ($F_{1,64} = 9.14$, $p < 0.01$), explained by faster overall responses to the nested structure compared with the crossed structure by the German participants but not by the Dutch (figure 7b).

As for the three-dependency structures, we additionally ran an analysis with only those responses elicited by the specific violated sequence positions in the violation trials. This resulted in a main effect of position ($F_{1,64} = 4.86$, $p < 0.05$), a main effect of block ($F_{2,128} = 43.50$, $p < 0.001$), an interaction

between position and block ($F_{2,128} = 19.16$, $p < 0.001$), a main effect of structure ($F_{1,64}) = 6.61$, $p < 0.05$), a main effect of language ($F_{1,64} = 21.8$. $p < 0.001$) and an interaction between language and structure $F_{1,64} = 8.14$, $p < 0.01$. This again indicates that Germans respond faster to nested than to crossed structures, relative to Dutch participants.

(d) *Two-dependency sequences: beep task*
A mixed ANOVA with position (first, last) as within-subjects factor, and structure (nested, crossed) and language (German, Dutch) as between-subjects factors revealed a main effect of position with $F_{165} = 116.26$, $p < 0.001$, and a significant interaction

between position and structure $F_{1,65} = 4.35$, $p < 0.05$. Post-hoc paired t-tests showed that for both nested and crossed sequences, error rates on the last position are lower than on the first (Bonferroni-corrected with $\alpha = 0.0125$; first versus last position, $t_{34} = 7.27$, $p < 0.001$ and $t_{34} = 8.84$, $p < 0.001$ for nested and crossed sequences, respectively), and error rate per structure type did not differ for both positions (first position (equal variances not assumed), $t_{65.22} = -1.27$, $p = 0.21$; last position, $t_{67} = 0.50$, $p < 0.62$) (see also figure 7).

(e) *Processing limitations: beep task against chance level*

In order to further investigate potential limitations on sequence processing and thus, by hypothesis, language processing, we conducted one-sample t-tests against chance level (0.5), for the beep task in each condition. For the three-dependency crossed structures, the error rate was significantly below the chance level (i.e. participants produced significantly fewer errors than would be expected by chance) for all sequence positions (first position versus chance, $t_{33} = -3.26$, $p < 0.01$; middle position versus chance, $t_{33} = -4.50$, $p < 0.001$; last position versus chance $t_{33} = -12.75$, $p < 0.001$). For the three-dependency nested structures, only the first and last positions were below the chance level, whereas error rates on the middle position were significantly higher than that predicted by chance (first position versus chance $t_{33} = -2.11$, $p < 0.05$; middle position versus chance, $t_{33} = 2.10$, $p < 0.05$; third position versus chance, $t_{33} = -2.75$, $p < 0.01$). For the two-dependency structures, both crossed and nested, error rates at all sequence positions were below the chance level (crossed structures: first position versus chance, $t_{34} = -6.15$, $p < 0.001$, last position versus chance, $t_{34} = -14.57$, $p < 0.001$; nested structures: first position versus chance, $t_{33} = -5.39$, $p < 0.001$; last position versus chance, $t_{33} = -12.26$, $p < 0.001$). When collapsing across sequence positions, the nested three-dependency structure was the only structure type that elicited error rates significantly above the chance level (three dependencies: crossed versus chance level, $t_{33} = -8.52$, $p < 0.001$; nested versus chance level, $t_{33} = -1.19$, $p = 0.24$; two dependencies: crossed versus chance level, $t_{34} = -10.13$, $p < 0.001$; nested versus chance level, $t_{33} = 10.72$, $p < 0.001$; figure 8).

5. DISCUSSION

In §1, we proposed two possible sources of processing difficulty for different types of non-adjacent dependency structures, namely (i) memory load due to the inherent complexity of the structure and (ii) processing experience. We discuss our findings in light of this proposal.

(a) *Inherent complexity of nested and crossed dependencies*

The results clearly show that, in accord with our prediction, the missing-verb-effect analogue is specific to nested dependencies. Both reaction times and error rate in the beep task showed that this is absent in crossed dependencies, even in situations where three

dependencies have to be resolved: the second reaction time analysis showed a rather opposite result with the slowest responses elicited by the middle position. This is the case for both reaction times and error rate in the beep task. The general tendency was that responses at the middle position in three-dependency nested structures were insensitive to the violation trials—the predicted 'missing-verb' position, resulting in a continuing decrease in reaction times from block 1 to block 3, as well as an error rate significantly higher than chance for this position. Reaction times for the first and last positions slowed down in the violation block, which is typical for SRT tasks. The insensitivity to violation trials at the middle position can be seen as analogous to the missing-verb effect in natural-language situations, and thus underscores the appropriateness of sequence-learning tasks for investigating natural-language processing. The same holds for the beep task: when completing the sequence $A_1A_2A_3B_1$, with B_1 followed by a beep, participants opted for B_3 instead of B_2. Interestingly, this was true for nested, but not for crossed dependencies. As argued in §1, this may indicate that crossed sequences tax memory less than nested sequences, such that crossed sequences suffer less from memory saturation (a key factor in the missing-verb effect) than nested sequences. Again, this is supported both by reaction time patterns and the error rates in the beep task.

The results also supported our prediction that no differences between nested and crossed dependency structures would emerge when only two non-adjacent dependencies had to be learned. For the structures involving two non-adjacent dependencies, no effects of structure type, or of place of violation, were found on reaction time. That is, violations at all positions caused a slowing down in the reaction times, for both nested and crossed dependencies, and irrespective of the participants' native language. In the beep task, error rates at all positions were lower than would be expected by chance, suggesting that crossed structures, whether involving two or three dependencies, still lie within the limits of sequence processing. This is not the case for nested structures: whereas two nested dependencies are still within our processing limits, three nested dependencies appear to be beyond what we can process, which is also predicted by SPLT [25]. This supports our hypothesis that the ordering of dependencies (i.e. nested or crossed) becomes relevant only when the number of dependencies exceeds two [2]. As discussed in §1, this is in line with the natural-language data from Bach *et al.* [5], and the modelling data from Christiansen & MacDonald [12]. Moreover, analyses of natural-language corpora by Karlsson [55] further underscore our sequential-learning findings: he found greater occurrence of two nested dependencies compared with three, which are very rare in spoken language, and argues that 'only one cycle of centre-embedding is in really productive use' (p. 24). Unfortunately, corpus analyses determining the frequency of crossed-dependency structures are scarce, though it has been observed that copying operations (relevant for crossed dependencies) are much more common in natural language than mirror operations (relevant for nested dependencies) [56].

To summarize, our findings fit nicely with the suggested complexity levels shown in figure 4, and support the prediction of De Vries *et al.* [2] that two decisive factors add to processing complexity: (i) the number of non-adjacent dependencies to be resolved and (ii) if this number exceeds two, then the ordering of non-adjacent dependencies becomes important, such that crossed are easier than nested dependencies.

(b) *Processing differences due to prior linguistic experience*

Our prediction that German participants would not be sensitive to the missing-verb effect analogue was not borne out by our findings; instead, both speakers of Dutch and German exhibited this effect. One possible explanation for this finding may be that we did not use natural-language materials, such as real words, but used non-sense syllables. By adding semantics to a structure, establishing non-adjacent dependencies is easier, as additional information about the underlying structure is revealed. Furthermore, when using natural-language material, the learner may be more likely to engage the statistical regularities used in real-life language processing.

Interestingly though, some differences between native Dutch and German speakers did emerge. First, we observed that German participants were overall faster than Dutch participants. We believe that this is an uninteresting main effect owing to our use of different populations and thus has no real theoretical implications. Importantly, this population difference has no effect on the experimental manipulations. A more interesting difference between Dutch and German participants emerged when *two* nested dependencies had to be resolved: Germans exhibited overall faster responses to nested compared with crossed dependencies, whereas Dutch participants did not differ between the two. Possibly this difference reflects prior linguistic experience with nested structures for the Germans. Given that two nested dependencies are much more frequent and natural-language-like than three nested dependencies (see Karlsson [55] for an extensive review), it may be again an indication that the closer the used materials are to natural language, the better one is able to replicate natural-language data. Our results, however, show that sequence-learning tasks nonetheless can provide a useful way of testing the limits of human sequence learning and thus, by extension, of natural-language processing.

To summarize, our findings of effects related to prior linguistic experience are relatively weak. As we have argued, this may be due to the materials not resembling natural-language elements. A very fruitful line for future research therefore is to investigate how we can fine-tune sequence-learning paradigms such that their sensitivity to capture specific natural-language phenomena can be improved. As a follow-up study, we would suggest using noun–verb pairs instead of syllables as stimuli, to investigate if prior language experience would affect the results in a situation that comes closer to a natural-language situation.

(c) *Conclusion*

Our results support the hypothesis that there are two sources of processing complexity, namely the memory resources required by the particular sequence structure and, to a lesser extent, processing experience. As for the former, the results show that the processing benefit of crossed dependencies is only prevalent in situations where there are more than two non-adjacent dependencies to be resolved [2]. Indeed, an analogue of the missing-verb effect was observed in our sequence-learning task, but only for nested, not for crossed dependencies. We suggest that these results illustrate the upper limits of both human sequence learning and natural-language processing.

REFERENCES

1 Chomsky, N. 1956 Three models for the description of language. *IRE Trans. Inf. Theory* **2**, 113–124. (doi:10.1109/TIT.1956.1056813)

2 De Vries, M. H., Christiansen, M. H. & Petersson, K. M. 2011 Learning recursion: multiple nested and crossed dependencies. *Biolinguistics* **5**, 10–35.

3 Hauser, M. D., Chomsky, N. & Fitch, W. T. 2002 The faculty of language: what is it, who has it, and how did it evolve. *Science* **298**, 1569–1579. (doi:10.1126/science.298.5598.1569)

4 Tallerman, M., Newmeyer, F., Bickerton, D., Nouchard, D., Kann, D. & Rizzi, L. 2009 What kinds of syntactic phenomena must biologists, neurobiologists, and computer scientists try to explain and replicate. In *Biological foundations and origin of syntax* (eds D. Bickerton & E. Szathmáry), pp. 135–157. Cambridge, MA: MIT Press.

5 Bach, E., Brown, C. & Marslen-Wilson, W. 1986 Crossed and nested dependencies in German and Dutch: a psycholinguistic study. *Lang. Cogn. Process.* **1**, 249–262. (doi:10.1080/01690968608404677)

6 Blaubergs, M. S. & Braine, M. D. 1974 Short-term memory limitations on decoding self-embedded sentences. *J. Exp. Psychol.* **102**, 745–748. (doi:10.1037/h0036091)

7 Hakes, D. T., Evans, J. S. & Brannon, L. L. 1976 Understanding sentences with relative clauses. *Mem. Cogn.* **4**, 283–290. (doi:10.3758/BF03213177)

8 Hamilton, H. W. & Deese, J. 1971 Comprehensibility and subject–verb relations in complex sentences. *J. Verbal Learn. Verbal Behav.* **10**, 163–170. (doi:10.1016/S0022-5371(71)80008-7)

9 Wang, M. D. 1970 The role of syntactic complexity as a determiner of comprehensibility. *J. Verbal Learn. Verbal Behav.* **9**, 398–404. (doi:10.1016/S0022-5371(70)80079-2)

10 Gibson, E. & Thomas, J. 1999 Memory limitations and structural forgetting: the perception of complex ungrammatical sentences as grammatical. *Lang. Cogn. Process.* **14**, 225–248. (doi:10.1080/016909699386293)

11 Gimenes, M., Rigalleau, F. & Gaonac'h, D. 2009 The effect of noun phrase type on working memory saturation during sentence comprehension. *Eur. J. Cogn. Psychol.* **21**, 980–1000. (doi:10.1080/09541440802469523)

12 Christiansen, M. H. & MacDonald, M. C. 2009 A usage-based approach to recursion in sentence processing. *Lang. Learn.* **59**, 126–161. (doi:10.1111/j.1467-9922.2009.00538.x)

13 Vasishth, S., Suckow, K., Lewis, R. L. & Kern, S. 2010 Short-term forgetting in sentence comprehension: cross-linguistic evidence from verb-final structures. *Lang. Cogn. Process.* **25**, 533–567. (doi:10.1080/01690960903310587)

14 Hurford, J. R. 2011 *The origins of grammar. Language in the light of evolution*, vol. xiii. Oxford, UK/New York, NY: Oxford University Press.

15 MacDonald, M. C. & Christiansen, M. H. 2002 Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychol. Rev.* **109**, 35–54. (doi:10.1037/0033-295X.109.1.35)

16 Roth, F. P. 1984 Accelerating language learning in young children. *Child Lang.* **11**, 89–107. (doi:10.1017/S0305000900005602)

17 Dabrowska, E. 1997 The LAD goes to school: a cautionary tale for nativists. *Linguistics* **35**, 735–766. (doi:10.1515/ling.1997.35.4.735)

18 Hoover, M. L. 1992 Sentence processing strategies in Spanish and English. *J. Psycholinguist. Res.* **21**, 275–299. (doi:10.1007/BF01067514)

19 Hawkins, J. A. 1994 *A performance theory of order and constituency.* Cambridge, UK: Cambridge University Press.

20 Everett, D. L. 2005 Cultural constraints on grammar and cognition in Piraha: another look at the design features of human language. *Curr. Anthropol.* **46**, 621–646. (doi:10.1086/431525)

21 Nevins, A., Pesetsky, D. & Rodrigues, C. 2009 Piraha exceptionality: a reassessment. *Language* **85**, 355–404. (doi:10.1353/lan.0.0107)

22 Everett, D. L. 2007 Cultural constraints on grammar in Pirahã: a reply to Nevins, Pesetsky, and Rodrigues. See http://ling.auf.net/lingBuzz/000427.

23 Elman, J. L. 1991 Distributed representation, simple recurrent networks, and grammatical structure. *Mach. Learn.* **7**, 195–225.

24 Christiansen, M. H. & Chater, N. 1999 Toward a connectionist model of recursion in human linguistic performance. *Cogn. Sci.* **23**, 157–205. (doi:10.1207/s15516709cog2302_2)

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

26 Conway, C. M. & Pisoni, D. B. 2008 Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Learn. Skill Acquis. Read. Dyslexia* **1145**, 113–131. (doi:10.1196/annals.1416.009)

27 Christiansen, M. H. & Chater, N. 2008 Language as shaped by the brain. *Behav. Brain Sci.* **31**, 489–558. (doi:10.1017/S0140525X08004998)

28 Nissen, M. J. & Bullemer, P. 1987 Attentional requirements of learning: evidence from performance-measures. *Cogn. Psychol.* **19**, 1–32. (doi:10.1016/0010-0285(87)90002-8)

29 Reber, A. S. 1967 Implicit learning of artificial grammars. *J. Verbal Learn. Verbal Behav.* **6**, 855–863. (doi:10.1016/S0022-5371(67)80149-X)

30 Forkstam, C. & Petersson, K. M. 2005 Towards an explicit account of implicit learning. *Curr. Opin. Neurol.* **18**, 435–441. (doi:10.1097/01.wco.0000171951.82995.c4)

31 Lieberman, M. D., Chang, G. Y., Chiao, J., Bookheimer, S. Y. & Knowlton, B. J. 2004 An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *J. Cogn. Neurosci.* **16**, 427–438. (doi:10.1162/089892904322926764)

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

33 Bahlmann, J., Gunter, T. C. & Friederici, A. D. 2006 Hierarchical and linear sequence processing: an electrophysiological exploration of two different grammar types. *J. Cogn. Neurosci.* **18**, 1829–1842. (doi:10.1162/jocn.2006.18.11.1829)

34 Petersson, K. M., Forkstam, C. & Ingvar, M. 2004 Artificial syntactic violations activate Broca's region. *Cogn. Sci.* **28**, 383–407. (doi:10.1016/j.cogsci.2003.12.003)

35 De Vries, M. H., Barth, A. C. R., Maiworm, S., Knecht, S., Zwitserlood, P. & Flöel, A. 2010 Electrical stimulation of Broca's area enhances implicit learning of an artificial grammar. *J. Cogn. Neurosci.* **22**, 2427–2436. (doi:10.1162/jocn.2009.21385)

36 Udden, J., Folia, V., Forkstam, C., Ingvar, M., Fernandez, G., Overeem, S., van Elswijk, G., Hagoort, P. & Petersson, K. M. 2008 The inferior frontal cortex in artificial syntax processing: an rTMS study. *Brain Res.* **1224**, 69–78. (doi:10.1016/j.brainres.2008.05.070)

37 Knowlton, B. J., Mangels, J. A. & Squire, L. R. 1996 A neostriatal habit learning system in humans. *Science* **273**, 1399–1402. (doi:10.1126/science.273.5280.1399)

38 Reber, P. J. & Squire, L. R. 1999 Intact learning of artificial grammars and intact category learning by patients with Parkinson's disease. *Behav. Neurosci.* **113**, 235–242. (doi:10.1037/0735-7044.113.2.235)

39 Brown, J., Aczel, B., Jiménez, L., Kaufman, S. B. & Grant, K. P. 2010 Intact implicit learning in autism spectrum conditions. *Q. J. Exp. Psychol.* **63**, 1789–1812. (doi:10.1080/17470210903536910)

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

41 Pavlidou, E. V., Williams, J. M. & Kelly, L. M. 2009 Artificial grammar learning in primary school children with and without developmental dyslexia. *Ann. Dyslexia* **59**, 55–77. (doi:10.1007/s11881-009-0023-z)

42 Rüsseler, J., Gerth, I. & Münte, T. 2006 Implicit learning is intact in adult developmental dyslexic readers: Evidence from the serial reaction time task and artificial grammar learning. *J. Clin. Exp. Neuropsychol.* **28**, 808–827. (doi:10.1080/13803390591001007)

43 Folia, V., Uddén, J., Forkstam, C. & Petersson, K. M. 2008 Implicit learning and dyslexia. *Learn. Skill Acquis. Read. Dyslexia* **1145**, 132–150. (doi:10.1196/annals.1416.012)

44 Packard, M. G. & Knowlton, B. J. 2002 Learning and memory functions of the basal ganglia. *Annu. Rev. Neurosci.* **25**, 563–593. (doi:10.1146/annurev.neuro.25.112701.142937)

45 Ullman, M. T. 2004 Contributions of memory circuits to language: the declarative/procedural model. *Cognition* **92**, 231–270. (doi:10.1016/j.cognition.2003.10.008)

46 Gebhart, A. L., Newport, E. L. & Aslin, R. N. 2009 Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychon. Bull. Rev.* **16**, 486–490. (doi:10.3758/PBR.16.3.486)

47 Newport, E. L. & Aslin, R. N. 2004 Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cogn.*

Psychol. **48**, 127–162. (doi:10.1016/S0010-0285(03)00128-2)

48 Uddén, J., Araújo, S., Forkstam, C., Ingvar, M., Haggoort, P. & Petersson, K. M. 2009 A matter of time: implicit acquisition of recursive sequence structures. In *Proc. 31st Annual Conf. Cognitive Science Society, Amsterdam, The Netherlands, 29 July–1 August 2009* (eds N. Taatgen & N. Van Rijn), pp. 2444–2449. Cognitive Science Society Inc.

49 Gomez, R. L. 2002 Variability and detection of invariant structure. *Psychol. Sci.* **13**, 431–436. (doi:10.1111/1467-9280.00476)

50 Onnis, L., Monaghan, P., Richmond, K. & Chater, N. 2005 Phonology impacts segmentation in online speech processing. *J. Mem. Lang.* **53**, 225–237. (doi:10.1016/j.jml.2005.02.011)

51 Perruchet, P. & Pacton, S. 2006 Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn. Sci.* **10**, 233–238. (doi:10.1016/j.tics.2006.03.006)

52 Misyak, J. B., Christiansen, M. H. & Tomblin, J. B. 2010 Sequential expectations: the role of prediction-based learning in language. *Topics Cogn. Sci.* **2**, 138–153. (doi:10.1111/j.1756-8765.2009.01072.x)

53 Petersson, K. M., Folia, V., Uddén, J., De Vries, M. & Forkstam, C. 2010 Artificial language learning in adults and children. *Lang. Learn.* **60**, 188–220. (doi:10.1111/j.1467-9922.2010.00606.x)

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

55 Karlsson, F. 2007 Constraints on multiple center-embedding of clauses. *J. Linguist.* **43**, 365–392. (doi:10.1017/S0022226707004616)

56 Manaster-Ramer, A. 1986 Copying in natural languages, context freeness, and queue grammars. In *Proc. 24th Annual Meeting on Association for Computational Linguistics, New York, NY, 10–13 July 1986*, pp. 85–89. Stroudsburg, PA: Association for Computational Linguistics.