



Cognitive Science 36 (2012) 1078–1101
Copyright © 2012 Cognitive Science Society, Inc. All rights reserved.
ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2012.01235.x

Implicit Acquisition of Grammars With Crossed and Nested Non-Adjacent Dependencies: Investigating the Push-Down Stack Model

Julia Uddén,^{a,b,c} Martin Ingvar,^b Peter Hagoort,^{a,c} Karl M. Petersson^{a,b,c,d}

^a*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

^b*Cognitive Neurophysiology Research Group, Stockholm Brain Institute, Department of Clinical Neuroscience, Karolinska Institute*

^c*Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging, Radboud University Nijmegen*

^d*Cognitive Neuroscience Research Group, Institute of Biotechnology & Bioengineering, Centre for Molecular and Structural Biomedicine, University of the Algarve*

Received 17 August 2010; received in revised form 22 August 2011; accepted 24 August 2011

Abstract

A recent hypothesis in empirical brain research on language is that the fundamental difference between animal and human communication systems is captured by the distinction between finite-state and more complex phrase-structure grammars, such as context-free and context-sensitive grammars. However, the relevance of this distinction for the study of language as a neurobiological system has been questioned and it has been suggested that a more relevant and partly analogous distinction is that between non-adjacent and adjacent dependencies. Online memory resources are central to the processing of non-adjacent dependencies as information has to be maintained across intervening material. One proposal is that an external memory device in the form of a limited push-down stack is used to process non-adjacent dependencies. We tested this hypothesis in an artificial grammar learning paradigm where subjects acquired non-adjacent dependencies implicitly. Generally, we found no qualitative differences between the acquisition of non-adjacent dependencies and adjacent dependencies. This suggests that although the acquisition of non-adjacent dependencies requires more exposure to the acquisition material, it utilizes the same mechanisms used for acquiring adjacent dependencies. We challenge the push-down stack model further by testing its processing predictions for nested and crossed multiple non-adjacent dependencies. The push-down stack model is partly supported by the results, and we suggest that stack-like properties are some among many natural properties characterizing the underlying neurophysiological mechanisms that implement the online memory resources used in language and structured sequence processing.

Correspondence should be sent to Julia Uddén, Stockholm Brain Institute, A2:3, Retzius väg 8, 171 72 Stockholm, Sweden. E-mail: julia.udden@ki.se

Keywords: Artificial grammar learning; Non-adjacent dependencies; Crossed; Nested; Implicit learning

1. Introduction

Human language is one of the most complex computational biological systems. In psycholinguistic and neurobiological research on the language system, the concept of complexity is currently revisited with a focus on the relative processing difficulties of different types of syntactic structures, for instance, using artificial grammar learning (AGL) paradigms (see, e.g., de Vries, Monaghan, Knecht, & Zwitserlood, 2008; Fitch & Hauser, 2004; for a review, see de Vries, Christiansen, & Petersson, 2011). In order to understand syntactic complexity, recent research has investigated different types of sentence-level dependencies and their relative processing difficulties (Gómez & Maye, 2005; Newport & Aslin, 2004). For instance, in sentences with non-adjacent dependencies, the computational process of syntactic unification (Hagoort, 2005; Vosse & Kempen, 2000) is extended in time and this requires online processing memory. One possibility is that this memory is a specialized memory resource. The issue of processing difficulties is intimately connected with how memory resources are integrated into language processing. From the point of view of computability theory (Cutland, 1980; Davis, Sigal, & Weyuker, 1994), all computational devices need to store intermediate results in a memory (e.g., on a tape in a Turing machine or registers in a register machine) in order to compute (except in pure feedforward computations). However, from a mathematical point of view, the dynamical variables allocated to processing and memory in computational devices (e.g., actual computers or real neural networks) are mainly different in terms of which time scale they live on. Thus, in any physical realization of a computational device, there is a continuum between processing and more stable representations or memory, where processing variables typically will have faster dynamics than memory variables. In this study, we investigate syntactic complexity from the memory perspective by exploring predictions of different memory architectures available in the theoretical computational science literature.

This line of thinking started in the 1950s with the Chomsky hierarchy for formal grammars, which shows that more complex grammar classes require more powerful memory architectures. Thus, the Chomsky hierarchy is a memory hierarchy (Chomsky, 1963) and computational complexity is indistinguishable from the characteristics of processing memory (assuming finite-state control; Minsky, 1967). In an early version, this hierarchy consists of regular (finite-state; T3), context-free (T2), context-sensitive (T1), and general phrase-structure grammars (Turing-Thue; T0). All classes but the regular grammars require infinite memory resources in order to realize their full computational power. Neurobiological systems are finite, not only with respect to computational control but also with respect to memory, and thus cannot process, produce, or parse arbitrarily long sequences. It follows that the infinite aspect of the Chomsky hierarchy is not directly relevant for neurobiological language research, whereas the different memory architectures themselves might be (Folia,

Forkstam, Ingvar, Hagoort, & Petersson, 2011; Levelt, 1974; Petersson, 2005, 2008; Petersson, Folia, & Hagoort, 2010; Pullum & Scholz, 2009, 2010). One of the memory architectures in the hierarchy is the push-down stack and it is conceivable that a bounded push-down stack is used in language processing, as one possibility, suggested by, for example, Levelt (1974). The push-down stack model can be seen as a simple way to translate generative theories into performance models and its predictions for sequential processing was more recently discussed in Dienes and Longuet-Higgins (2004). In the following section, we will highlight the differences between the so-called external and internal memory devices. The push-down stack model is an example of an external memory device.

Efficient computations in information processing systems sometimes involve prediction, for instance, prediction of expected elements far downstream in a sequence, which is possible when there are non-adjacent dependencies. There are at least two possible views (see Fig. 1). In the general case, each element pushes the system into a new internal state. The state transitions instantiate an internal memory (internal to the computation as it is not meaningful to segregate processing and memory from this point of view). The different possible states and the way they are connected represent the grammar that can generate or parse the sequence. This is Turing's original perspective (Turing, 1936; Wells, 2005). From this internal memory perspective, there is no need to describe processing as ever restoring elements upstream in the sequence.

The presence of a certain upstream element is instead encoded in the current state. However, in some cases,¹ it might be relevant to describe the mechanism as involving an external memory structure allowing upstream elements to be restored. For instance, when thinking about dynamical variables as implemented in the brain, where different biological mechanisms live on different time scales, and these mechanisms interact, the slower mechanisms might be relevantly described as an external memory structure with respect to the faster mechanism. Typically, there are different forms of constraints on the access of stored

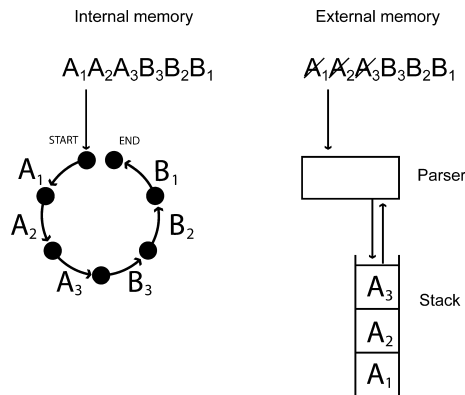


Fig. 1. In the internal memory case, sequences are parsed by traversing the transition graph from the start node to the end node along the directions indicated by the arrows and ticking of the letters written on the traversed arrows. Each element pushes the system into a new internal state and the state transitions instantiate the internal memory. In the external memory case, here exemplified by the push-down stack model, a separate parser pushes elements on a stack as they are parsed.

elements and those constraints correspond to the major memory architectures. Push-down stacks (first-in-last-out memories) represent one possible constraint on access to information, queues another (Davis et al., 1994). In the push-down stack, elements are *pushed* down on a stack as they are parsed and only the top element is accessible in one step. In order to access an element beneath the top element, a series of *pops* (erasing the top element to reveal a new top element) needs to be taken before some piece of information can be accessed (Taylor, 1998). Queues (first-in-first-out memories) work similarly, but now only the bottom element can be accessed in one step. When no constraints are present, the memory is of random access type (Savage, 1998), which means that all stored elements can be accessed in one step. The drawback of such an architecture is that addressability becomes an issue. Here, we are investigating the predictions of the access aspects. If the brain implements, for example, a stack, this could be regarded as a specialized memory structure for processing multiple *nested* non-adjacent dependencies (see Fig. 4).²

It is sometimes argued that non-adjacent dependencies are extremely difficult to acquire (Gómez & Maye, 2005; Newport, Hauser, Spaepen, & Aslin, 2004). Interestingly, results show that cotton-top tamarins are capable of acquiring simple non-adjacent dependencies (Newport et al., 2004). Moreover, Gómez and Maye (2005) investigated the acquisition of simple non-adjacent dependencies of the AXB type in infants, where X varies. The results showed that 15-month-old children acquired sensitivity to this type of non-adjacent dependency. In the present study, we tested whether the difference between acquiring and processing multiple non-adjacent and adjacent dependencies is qualitative or quantitative in nature. A qualitative difference would support the idea that different memory architectures are involved in acquiring or processing non-adjacent and adjacent dependencies. We used a complex artificial language with multiple non-adjacent dependencies in an implicit AGL paradigm. The types of dependencies we used occur in natural language, for example, in number or gender agreement between multiple nouns and verbs at non-adjacent sentence positions. One of the grammars we used included a crossed syntactic structure of multiple non-adjacent dependencies, as reflected in the indices of the sequence $A_1A_2A_3B_1B_2B_3$. This organization is famous in linguistics for being perhaps the only naturally occurring context-sensitive construction and versions exist in Dutch (Bach, Brown, & Marslen-Wilson, 1986) and Swiss German (Shieber, 1985). In particular, the crossed organization of multiple non-adjacent dependencies poses a problem for theories of non-adjacent dependency processing that rely on the following principles: (a) the gestalt principle and (b) the principle of tiers (Newport & Aslin, 2004). Both describe how to effectively remove the intervening material between the two dependent non-adjacent elements. This results in an abstract representation (the tier or the gestalt) in which a non-adjacent dependency is reduced to an adjacent dependency. In the case of crossed non-adjacent dependencies, however, the intervening material contains elements of the same type as the original dependent elements, and the intervening material constitutes parts of other crossing non-adjacent dependencies. Thus, these two processing principles cannot account for the acquisition of crossed non-adjacent dependencies. If we can show successful acquisition of crossed non-adjacent dependencies, we have thus shown that there must exist processing of non-adjacent dependencies as non-adjacent, before a possible reduction of these dependencies to adjacent dependencies can take place.

Our implicit AGL paradigm makes it possible to systematically investigate implicit acquisition of new syntactic structures from grammatical examples without performance feedback. In this context, implicit acquisition is a process whereby complex, rule-governed knowledge is acquired largely independent of awareness of both the process and product of acquisition (Reber, Walkenfeld, & Hernstadt, 1991). By implicit knowledge, we mean that the knowledge was acquired incidentally, without the use of hypothesis testing strategies (see Forkstam & Petersson, 2005, for a more detailed discussion). Our notion of implicit learning also entails the condition of the absence of access to (a) the acquired knowledge and (b) the acquisition mechanisms. The relevance of our use of the term lies in the analogy with natural language processing. Speakers/listeners generate/parse sentences without explicit access to the syntactic knowledge that allows them to do so. Similarly, they learn their native language without being able to describe the acquisition process or the result of the acquisition process explicitly.

In AGL, one separates the acquisition and testing phases. In the acquisition phase, participants are typically engaged in a short-term memory task using an acquisition sample of sequences generated from a formal grammar. In the standard version, subjects are informed after acquisition that the sequences were generated according to a complex system of rules and asked to classify novel items as grammatical or not, typically with the instruction to base their classification decisions on their immediate intuitive impression (i.e., guessing based on “gut feeling”). It is a robust finding on grammars with adjacent dependencies³ that subjects perform well above chance and more so after several days of acquisition (Folia et al., 2008; Forkstam, Hagoort, Fernandez, Ingvar, & Petersson, 2006; Uddén et al., 2008).

In this study, the combination of three crucial design features constitutes a methodological development to investigate implicit learning of non-adjacent dependencies: (a) during the acquisition phase, we use enough time (~2 weeks) for abstraction and consolidation processes to take place; (b) during the classification phase, the non-grammatical sequences are constructed so that explicit strategies, for example, counting, are unhelpful; and (c) the critical measure is the participants’ preference for grammatical and relative aversion to non-grammatical sequences. Because of the latter aspect, participants only need to indicate whether they like or dislike a sequence, and therefore, it is not necessary to inform them about the presence of a complex rule system before classification in order for the classification instruction to make sense. Informing participants prior to classification is the standard version of the AGL paradigm. Moreover, from the subject’s point of view, there is no correct or incorrect response and we instruct them to make their choice according to their gut feeling. The motivation to use explicit (e.g., problem solving) strategies is thus minimized. A preference for grammaticality has been found repeatedly for grammars with adjacent dependencies (Folia et al., 2008; Forkstam, Elwér, Ingvar, & Petersson, 2008; Zajonc, 1968). The most important observation in these studies is that the pattern of performance in the preference task is highly similar to the pattern of performance during the standard AGL conditions (i.e., when subjects subsequently get informed about the presence of a complex rule system). Generally, however, the performance levels under standard AGL conditions are somewhat higher than in preference AGL (Folia et al., 2008; Forkstam et al., 2008). This might be due to greater motivation and a reliance/trust in the subjective gut feeling.

In the present study, this is encouraged by informing the participants that their gut feelings correspond to real differences in the stimuli. A general finding, under the standard AGL conditions, is that the response times are longer compared to preference AGL. This suggests a more elaborate decision-making process, potentially enhanced by motivational factors. As an additional methodological improvement, we created a large set of acquisition sequences with a lot of variation in the surface structure so that explicit memorization would be practically impossible.

In the first experiment, we used the preference AGL paradigm to test whether implicit acquisition of multiple non-adjacent dependencies is experimentally achievable to the same performance levels observed with adjacent dependencies. More generally, we wanted to test whether the acquisition of non-adjacent dependencies is qualitatively different from that of adjacent dependencies. There are several ways in which performance could be qualitatively different: (a) there could be significant learning of adjacent but not non-adjacent dependencies; (b) the performance levels could be lower for non-adjacent compared with adjacent processing; (c) there could be a evidence of acquisition in the preference but not in the grammaticality task, or the other way around; (d) relative performance differences from the preference to the grammaticality task could interact with whether non-adjacent dependencies were present; or (e) explicit knowledge of the grammar could co-evolve with successful acquisition of non-adjacent dependencies. No explicit knowledge has been observed after the acquisition of adjacent dependencies in similar paradigms (Folia et al., 2008; Forkstam et al., 2006, 2008; Uddén et al., 2008). In this study, the comparison between adjacent and non-adjacent dependency processing is not made in a within-subject design. This would require acquisition of multiple grammars by the same subject, which is methodologically challenging because of potential interference between grammars. Rather, we compare the pattern of results for non-adjacent dependency processing to robust results of processing of adjacent dependencies, which has been demonstrated in a number of previous studies (Folia et al., 2008; Forkstam et al., 2006, 2008; Uddén et al., 2008) in a between-subject comparison.

To summarize, the goals of the study are to (1) test whether multiple non-adjacent dependencies can be acquired to robust performance levels matching those seen for the acquisition of adjacent dependencies, in an implicit learning paradigm; (2) test whether the acquisition of adjacent and non-adjacent dependencies is qualitatively different more generally. If they are not, there is no need to postulate separate mechanisms for the acquisition of adjacent and non-adjacent dependencies, such as separate memory architectures. Finally, we want to (3) test whether there is evidence for the push-down stack model when comparing implicit acquisition of nested and crossed non-adjacent dependencies. The prime alternative model is the random access memory model. We test (1) and (2) in Experiment 1, by exposing subjects to a large grammar with crossed non-adjacent dependencies and by comparing this to results on adjacent dependencies from the literature. In Experiment 2, we test (3) in a between-group design. We expose one group to nested non-adjacent dependencies and one group to crossed non-adjacent dependencies. The push-down stack model predicts that the nested group should have an acquisition advantage, whereas the random access model does not predict such an advantage.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Nineteen right-handed healthy university students volunteered to participate in the study (15 females, 4 males, mean age \pm SD = 24 \pm 3 years). They were all pre-screened for medication use, history of drug abuse, head trauma, neurological or psychiatric illness, and family history of neurological or psychiatric illness. All participants gave written informed consent and the study was run under the Donders Center for Cognitive Neuroimaging Experimental Approval from the local medical ethics committee at the UMC St. Radboud.

2.1.2. Stimulus material

We generated grammatical sequences from a formal grammar including a crossed dependency part (e.g., $A_1A_2A_3B_1B_2B_3$). This subsequence consisted of letters from the alphabet [F, D, X, L, P, K] and was pre- and post-fixed with adjacent dependency parts from the alphabet [M, T, V, W, S, R]. The first half of the crossed part was always taken from the set [F, D, X] and the last half from [L, P, K] (see Fig. 2). This was to dissociate the surface structure from the underlying grammar and to minimize the likelihood of explicit memorizing or similar strategies by adding variance to the generated sequences. This also prevents initial and terminal position effects. The dependency structure was then created by pairing F with L, D with P, and X with K (e.g., $F_1X_2D_3L_1K_2P_3$). Thus, the first half of the non-adjacent dependency subsequence syntactically predicted the second half (containing L's, P's, and K's). One (FFFLLP*), two (FFFLKP*), or three (FFFPKP*) switches, for example,

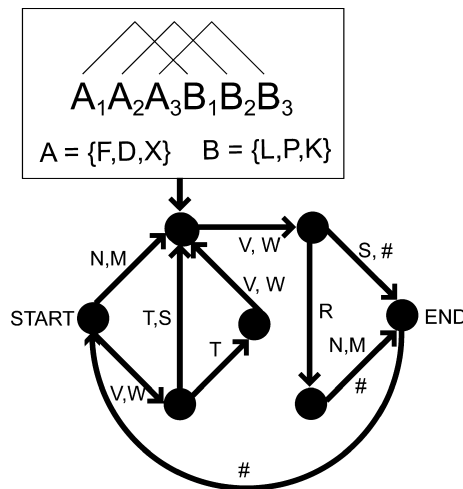


Fig. 2. The transition graph representation of the grammar, indicating where crossed fragments are inserted. Grammatical sequences are generated by traversing the transition graph from the start node to the end node along the directions indicated by the arrows and concatenating the letters written on the traversed arrows.

from L to P or K, created violations of the dependency structure. There could be two or three non-adjacent dependencies in a sequence. See Table 1 for example sequences. Non-grammatical sequences were selected to match the grammatical sequences in terms of complete sequence associative chunk strength (ACS; i.e., collapsed over order information within sequences; for further technical details, see Forkstam et al., 2006). The complete sequence ACS is a measure of the familiarity of the surface form of a sequence, operationalized as the average familiarity across all possible subsequences of two or three letters (i.e., bigrams and trigrams). The familiarity of a subsequence is determined by its frequency in the acquisition sequences.

For all conditions and all classification sets, irrespective of grammaticality status, sequences did not differ significantly in terms of ACS, as tested with two-sample *t*-tests (mean ACS \pm SD; grammatical sequences: 12.24 ± 0.26 , non-grammatical sequences: 12.25 ± 0.11) or length (length range = 5–12; mean length grammatical sequences = 10.95 ± 0.72 , mean length non-grammatical sequences = 10.77 ± 0.77). There were three different classification sets. Each of these three sets could be used in the baseline test, the preference test, or the grammaticality classification test. Thus, any differences between performance in the baseline, preference, and grammaticality tests could not be explained by the particular stimuli used in that test. The presentation order was randomized over subjects.

2.1.3. *Experimental procedure*

The experiment spanned 9 days spread over 2 weeks, with one implicit acquisition session each day. On Day 1, a baseline preference classification test was administered before the first acquisition session commenced. On the seventh day, subjects performed a preference classification test. Acquisition then continued for 2 days, and on the ninth day, the subjects were given a grammaticality classification test. The experimental procedure is described below (for more detailed information, see Folia et al., 2008). Three of the subjects chose to end the experiment on the seventh day and the remaining 16 subjects completed the experiment.

2.1.4. *Implicit acquisition*

The acquisition task (~30 min, 100 sequences) was presented as a short-term memory recall task to the subjects. Each sequence was presented for 4 s (whole sequence presentation). After the sequence disappeared, subjects typed the sequence from memory on a keyboard in a self-paced manner. No feedback was provided.

Table 1
Example sequences

Grammatical	Non-Grammatical
VTVDXPKVRM	WSDXKPVRM
WTVDDPPWRN	VSXFPPVRM
WSDXDPKPV RN	VSXFFLPLVR
VTDFFP LLWR	WTWDXFKPLV

2.1.5. Preference and grammaticality classification

Subjects were instructed to indicate, as rapidly as possible after sequence onset, whether they liked a sequence, based on their immediate intuitive impression (i.e., guessing based on “gut feeling”), by pressing the corresponding key with their left or right index finger. Each sequence was presented for 3.5 s followed by an inter-stimulus interval of 2.5 s. On the last day, just before the grammaticality test, the subjects were informed about the existence of a complex system of rules (but they were not informed about the actual rules). They were then instructed to classify novel sequences as grammatical or not, based on their immediate intuitive impression, in the same manner as in the preference test.

2.1.6. Post-experimental questionnaire

The post-experimental questionnaire was distributed after the last grammaticality test. The questionnaire started with open questions, and subjects were instructed to write down their spontaneous thoughts about what they had noticed in the test, what the grammar might have consisted of, and similar questions. Then the subjects were requested to produce examples of the grammar. In the forced-choice post-experimental questionnaire part, there were 10 statements (presented one-by-one) about the grammar and subjects were informed that some of them were true and some false. The subjects had to indicate both whether they thought the statement was true or false, and whether they had used the principle mentioned in the statement during the classification.

2.1.7. Data analysis

Repeated-measures ANOVAS and *t*-tests were used to analyze the data with a significance level of $p < .05$. We analyzed the classification performance (endorsement rates) with the factors TEST (baseline vs. preference classification) and GRAMMATICALITY (grammatical vs. non-grammatical sequences). The endorsement rate was defined as the number of sequences classified as grammatical independent of their actual status, divided by the total number of recorded answers for each factor level (Meulemans & Van der Linden, 1997). Mean values are reported with standard errors of the mean, if not otherwise stated. For analyzing forced-choice post-experimental questionnaires, we used the Binomial test.

2.2. Results

2.2.1. Classification performance

We analyzed the acquisition effect in an ANOVA by comparing the preference classification on the seventh day with the baseline preference classification with endorsement rates as the dependent variable. The main effect of TEST was significant, $F(1,18) = 5.9$, $p < .05$, suggesting an effect of TEST on the response bias. The mean endorsement rates \pm SEM were 0.47 ± 0.02 in the first test and in the last test 0.52 ± 0.02 . The main effect of GRAMMATICALITY was also significant, $F(1,18) = 18.1$, $p < .001$, showing higher endorsements of grammatical sequences. The interaction between these factors was significant, showing

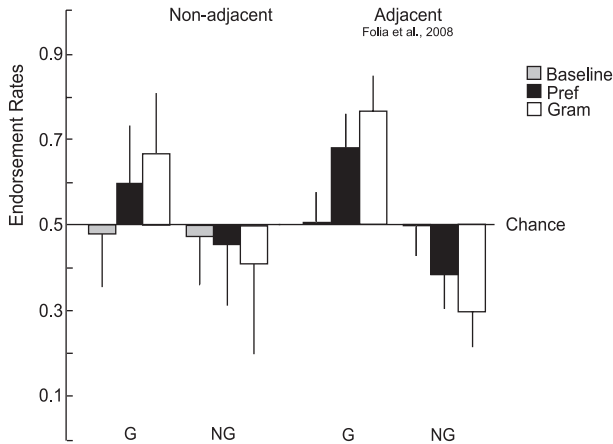


Fig. 3. Classification performance in endorsement rates of G = Grammatical and NG = Non-grammatical sequences. Baseline, baseline preference; Pref, Day 7 preference classification; Gram, Day 9 grammaticality classification. To the left, the results show significant successful acquisition of the crossed non-adjacent dependencies. To the right, for comparison, we show similar results adopted from one (Folia et al., 2008) of several (Forkstam et al., 2006, 2008; Uddén et al., 2008) similar 5-day studies with adjacent dependencies. Qualitatively, the results are similar. Error bars indicate the standard deviation.

successful implicit acquisition, $F(1,18) = 7.5$, $p = .01$ (Fig. 3). We then analyzed the last grammaticality classification session, where grammatical sequences were endorsed more often than non-grammatical sequences, $t(15) = 4.51$, $p < .001$. For the baseline preference, the corresponding test was non-significant ($p = .34$), and for the seventh day preference test, this was significant, $t(18) = 4.50$, $p < .001$.

2.2.2. Post-experimental questionnaire

The participant reports were, in our experience, typical for AGL experiments. As we informed the participants about the presence of a grammar before the grammaticality classification test, it was not surprising that most reported that there were some sort of dependencies between letters. Subjects were then given a list of correct and incorrect statements about the grammar and were asked to indicate whether they thought these were true and whether they used this knowledge during classification. Only two (out of 16) subjects thought that the full correct rule was true and three claimed that they had used some of this insight during classification. However, these subjects did not perform differently compared with those who reported no rule use. In other words, the self-reported insight or rule use did not translate into a performance benefit. Rather, the low recognition rate of the correct rules suggests that the participants had little or no demonstrable explicit knowledge of the underlying grammar. In addition, their reports suggest that they made little or no effective use of any potential knowledge. In fact, the participants classified the correctness of all other rules in the questionnaire at chance level (there were <79% correct answers at all other questions, $p > .06$).

2.3. Discussion

In Experiment 1, we used the preference AGL paradigm to test whether implicit acquisition of non-adjacent dependencies is experimentally achievable to the same performance levels observed with adjacent dependencies. The results of the first experiment suggest that acquisition and processing of non-adjacent dependencies does not differ qualitatively from that of adjacent dependencies. Moreover, the acquisition and processing of non-adjacent dependencies are not exceedingly difficult. This argues against the possibility that an additional, or external, memory architecture is needed to process non-adjacent dependencies.

3. Experiment 2

In the second experiment, we wanted to test whether syntactic computation of non-adjacent dependencies rely on a push-down stack-like memory structure. The push-down memory organization predicts a processing advantage for a hierarchically nested organization (e.g., $A_1A_2A_3B_3B_2B_1$) over the crossed dependencies (e.g., $A_1A_2A_3B_1B_2B_3$). Elements are pushed down on a stack as they are parsed from left to right; for example, when $A_1A_2A_3$ has been parsed, only the top element A_3 is available for processing when the next element arrives (B_3 for nested and B_1 for crossed, cf., Fig. 4). For the nested dependencies, the parsing process can proceed directly, whereas, in the crossed case, the information stored on the first stack has to be rearranged on a second stack, or the process will fail.

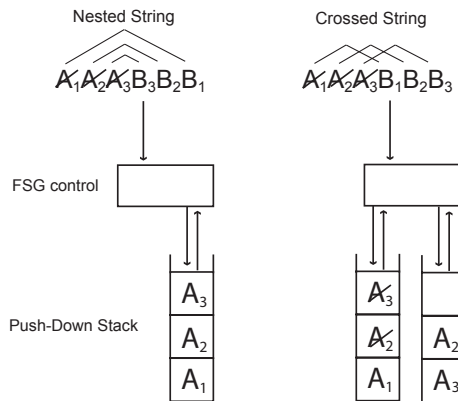


Fig. 4. Multiple non-adjacent dependencies can be organized in a hierarchically nested or non-hierarchical crossed way, as subscripts and lines indicate. In the push-down stack model, a finite-state grammar (FSG) is a control device pushing elements on a stack as they are parsed. When $A_1A_2A_3$ is parsed, only the top element A_3 is available for comparison with the next element. B_3 – B_1 can easily be compared in the nested structure. To reach A_1 in the crossed case, A_2 – A_3 first have to be popped and pushed on to another stack. This extra step will create a processing disadvantage for the crossed structure.

The comprehension of crossed and nested dependency types was investigated in a seminal cross-linguistic study by Bach et al. (1986). Dutch and German participants received natural speech with similar sentences organized in a crossed or nested structure, respectively, and rated how easy the sentence was to understand immediately afterward. Crossed sentences received higher comprehension scores than the nested type. This result was unexpected when considering that nested structures are more common than crossed structures cross-linguistically and was interpreted as evidence against a push-down stack. This argument was reiterated by Christiansen and Chater (1999). Semantic cross-linguistic variance in German and Dutch may, however, influence these results. From the pure sequence-processing point of view, or a syntactic perspective, this result would be more relevant if established in a paradigm where semantic factors are controlled for, for example, in an AGL paradigm. Thus, in the second experiment, we used the preference structural mere-exposure AGL paradigm to study whether there is a processing bias for nested compared with crossed dependency structures in the absence of lexical and sentence-level semantics.

3.1. Methods

3.1.1. Participants

Thirty-nine right-handed healthy university students, who did not participate in Experiment 1, volunteered to participate in the study (28 females, 11 males, mean age \pm SD = 21 \pm 2 years). Nineteen of the participants were exposed to an artificial grammar generating multiple crossed dependencies and 20 participants to a grammar generating multiple nested dependencies. Subject management procedures were the same as in Experiment 1.

3.1.2. Stimulus material

The stimulus generation procedure was identical to Experiment 1 with some modifications. We added grammatical sequences with a nested dependency part ($A_1A_2A_3B_3B_2B_1$). See Table 2 for example sequences. We had to remove the X-K letter pair in order to match for ACS across the different grammar types. This is explained by the fact that the middle chunks A_3B_3 are high-frequency chunks in nested structures and the corresponding A_3B_1 chunks in the crossed designs receive progressively lower frequency when more letter pairs are introduced. Each test set consisted of 64 sequences (32 grammatical, 32 non-grammatical). For all conditions in all classification sets, irrespective of nested or crossed

Table 2
Example sequences

Nested Grammatical	Nested Non-Grammatical	Crossed Grammatical	Crossed Non-Grammatical
MDDFLPPVRN	MFDPPLWS	MDFDPLWRN	NFFDPPLWR
NFDDPPLWRM	VXWDFFLPPWRM	NFFDLLPVRN	VXVDFDPLLVS
VSDFFLLPVRM	NFFFLPPWRM	VXFFFLLLS	VXWDDFLPPVR
VSDFDPLPW	VXVFFDPPLWR	VXWFDLPWS	NDDDLLPWRM

Table 3
Mean ACS and sequence lengths with SD

	Nested Grammatical	Nested Non-Grammatical	Crossed Grammatical	Crossed Non-Grammatical
ACS	19.94 ± 0.83	19.58 ± 1.5	19.84 ± 0.97	19.78 ± 0.93
Length	10.18 ± 0.91	10.03 ± 1.37	10.01 ± 0.80	10.16 ± 0.92

dependencies and irrespective of the grammaticality status, the sequences did not differ significantly in terms of ACS or sequence length, as tested with two-sample *t*-tests (length range = 5–12; for mean lengths, see Table 3).

3.1.3. Experimental procedure and data analysis

The experimental procedure was the same as in Experiment 1, except that the last preference test was performed on the last day and the grammaticality test followed directly afterwards. The data analysis was the same as in Experiment 1. To recap, we analyzed the classification performance (endorsement rates) with the factors TEST (baseline vs. preference classification) and GRAMMATICALITY (grammatical vs. non-grammatical sequences). In additional testing, we tested the difference between grammars and we analyzed the effect of violation positions between grammars. Between-grammar effects were predicted to be the largest for sequences with three non-adjacent dependencies, where the length of the non-adjacent part of the sequence is six letters, for the following reason. In the case of crossed dependencies, all have the same length of the intervening material, whereas, for the nested dependencies, the length of the intervening elements varies. The number of intervening elements grows faster with the number of dependencies for nested compared with crossed grammars. The difference between crossed and nested dependencies is thus largest for sequences with more non-adjacent dependencies. For example, for two crossed non-adjacent dependencies, the intervening material for crossed dependencies is one letter, whereas the corresponding maximum for nested dependencies is two intervening letters. For three crossed dependencies, the intervening material is two letters, whereas for the nested dependencies, the maximum is four intervening letters. Thus, by definition, nested and crossed sequences are more different (and we can predict larger processing and memory differences) for three compared with two non-adjacent dependencies. In addition, this prediction has been empirically validated by Bach et al. (1986) as comprehension scores of crossed and nested differed only for sentences with three non-adjacent dependencies and not for sentences with two non-adjacent dependencies, in natural language. In addition, as explained in Fig. 4, if the brain implements a push-down stack, the number of additional computations needed for crossed compared with nested structures increases with the number of elements initially pushed on the stack. According to this reasoning, when comparing the crossed against nested grammar, we looked specifically at sequences with three non-adjacent dependencies. As there were too few sequences with two non-adjacent dependencies in the stimulus material, we did not analyze them separately. When effect sizes (Cohen's *d*) are needed to support our argument, they are reported.

3.2. Results

3.2.1. Classification performance

3.2.1.1. *Crossed non-adjacent dependencies*: There was no significant effect of TEST on response bias ($p = .14$), but there was a significant main effect of GRAMMATICALITY, $F(1,18) = 31.5, p < .001$, and the interaction between TEST and GRAMMATICALITY was significant, $F(1,18) = 31.3, p < .001$ (see Fig. 5). Grammatical sequences were endorsed significantly more often than non-grammatical sequences in the preference, $t(18) = 5.95, p < .001$, and in the grammaticality test, $t(18) = 9.07, p < .001$, but not in the baseline test ($p = .87$).

3.2.1.2. *Nested non-adjacent dependencies*: There was a main effect of TEST on response bias, $F(1,19) = 23.9, p < .001$, and a significant main effect of GRAMMATICALITY, $F(1,19) = 44.0, p < .001$. We also found a significant interaction between TEST and GRAMMATICALITY, $F(1,19) = 18.8, p < .001$. The grammatical sequences were endorsed significantly more often than non-grammatical sequences in the last preference test, $t(19) = 6.01, P < .001$, and the grammaticality test, $t(19) = 8.42, p < .001$, but not in the preference baseline test ($p = .06$).

3.2.1.3. *Between-grammar-type effects*: We predicted the largest differences between grammar types on the sequences with three non-adjacent dependencies. We tested these sequences in a three-way ANOVA with the factors TEST, GRAMMATICALITY, and the between-subjects factor GRAMMAR. The main effect of TEST type was significant, $F(1,37) = 25.70, p < .001$, as well as the main effect of GRAMMATICALITY, $F(1,37) = 52.56, p < .001$. The interaction between TEST, GRAMMATICALITY (levels:

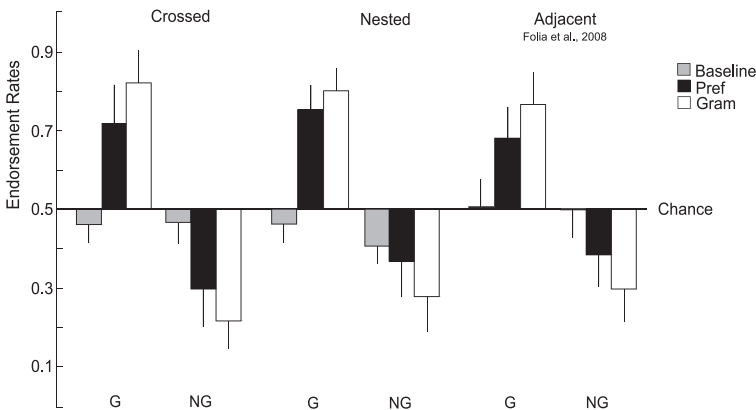


Fig. 5. Classification performance in endorsement rates of Grammatical (G) and Non-grammatical (NG) sequences. Pref, preference classification; Gram, grammaticality classification. To the right, for comparison, we show similar results adopted from one (Folia et al., 2008) of several (Forkstam et al., 2006, 2008; Uddén et al., 2008) similar 5-day studies with adjacent dependencies. Qualitatively, the results are similar. Error bars indicate standard deviations.

grammatical vs. non-grammatical), and GRAMMAR (levels: nested vs. crossed) was significant, $F(1,37) = 9.22$, $p < .005$, $d = 0.34$. The crossed group showed a greater interaction between TEST and GRAMMATICALITY, $F(1,18) = 34.45$, $p < .001$, $d = 1.80$ (see Fig. 6), compared with the nested group, $F(1,18) = 8.53$, $p < .01$, $d = 1.24$.⁴ There was no interaction between GRAMMATICALITY and GRAMMAR in the grammaticality classification ($p = .37$).

When only analyzing the non-grammatical sequences with three non-adjacent dependencies, there was a highly significant interaction between GRAMMAR (nested vs. crossed) and TEST, $F(1,37) = 31.8$, $p < .001$. There was no main effect of bias between the two grammars in the baseline test or last-day preference test ($p > .15$). The interaction between GRAMMAR and TEST was not present for the grammatical sequences ($p > .73$). This interaction approached significance when testing all non-grammatical sequences, that is, when also including the ones two non-adjacent dependencies ($p = .06$). This can be explained by the initial sensitivity to the grammaticality factor in the nested group, which approached significance ($p = .06$). The initial sensitivity was stronger in the second half of the baseline test, compared to the first half ($t(19) = 2.68$, $p = .02$), which points to early acquisition of sensitivity to at least some aspects of the nested grammar. Alternative explanations, such as some surface feature (e.g., the number of instances of the letter V) causing the initial sensitivity, does not account to the same extent for the change of preference during the baseline test. Thus, the fact that the acquisition effect from the first baseline test to the last preference test is larger for crossed compared with nested grammars can be explained by crossed grammars having fewer aspects that are immediately acquired when exposed to a set of both grammatical and non-grammatical sequences. Such early acquisition has been found in other studies as well (Forkstam et al., 2008; Stadler & Frensch,

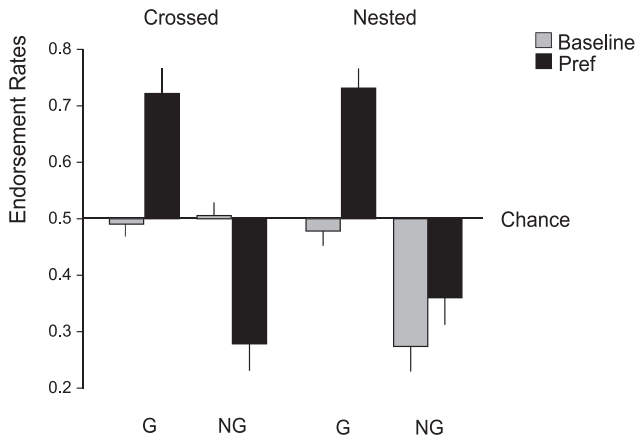


Fig. 6. Classification performance in endorsement rates of Grammatical (G) and Non-grammatical (NG) sequences. Pref, preference classification; Gram, grammaticality classification. This figure displays sequences with three non-adjacent dependencies only, whereas Fig. 5 displays sequences with two non-adjacent dependencies, in addition. The crossed group showed a significantly greater interaction between TEST and GRAMMATICALITY. Error bars indicate standard error of mean.

1998) and is possible although subjects are exposed to as many grammatical and non-grammatical sequences, as grammatical sequences display more structure than the non-grammatical sequences. Note that non-grammatical sequences also display grammatical subsequences but are interrupted with violations.

When testing only sequences with three non-adjacent dependencies, there was no interaction between GRAMMATICALITY and GRAMMAR in the last preference test ($p = .49$) or in the grammaticality classification task ($p = .26$).

3.2.1.4. Violation position effects: We then analyzed how endorsement rates depended on the position of the first violation (POSITION) in the sequences within six-letter non-adjacent fragments. We used post hoc Tukey's HSD tests, corrected for three comparisons: first (B_1) against second (B_2), first against third (B_3), and second against third position. Position had a similar effect on the data in the last preference test and the grammaticality test and thus we pooled the results. The only significant difference in difficulty between positions was the difference between the first and the second violation position ($p < .01$) for the crossed grammar (for the rest of the comparisons in the nested and crossed groups, $p > .24$). The first violation position was more difficult than the second for the crossed grammar. The interaction between position and grammar type for the first against the second position approached significance, $F(1,37) = 3.71$, $p = .06$. As the push-down stack model predicts a unique role for the first position against the last two positions for the crossed grammar but not for the nested grammar, we also tested the first violation position against an average of

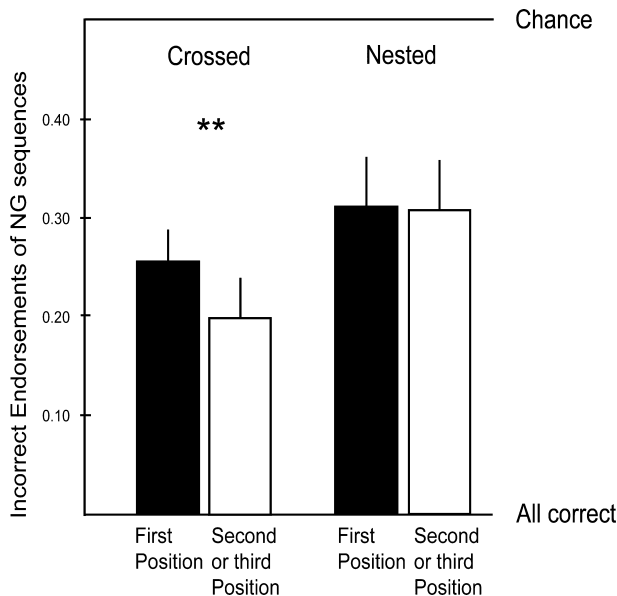


Fig. 7. Classification performance dependent on violation position, in endorsement rates, of crossed and nested non-grammatical (NG) sequences. For the crossed grammar, violations in the first position (e.g., FFFDLL*) were significantly harder to spot than violations in the second (e.g., FFFLDL*) or third positions (e.g., FFFLLD*). This was not the case for the nested grammar. Error bars indicate standard error of the mean.

the last two positions. Using a paired, two-tailed *t*-test, this comparison was significant, $t(18) = 2.99$, $p < .01$, whereas this was not the case for the nested grammar, $t(19) = 0.25$, $p = .81$ (see Fig. 7).

3.2.2. Post-experimental questionnaires

The post-experimental questionnaire was distributed after the grammaticality test. The participant reports were typical; subjects spontaneously reported some high-frequency chunks, mostly from the prefix and suffix fragments (with adjacent dependencies only). As in Experiment 1, subjects reported that there was some sort of dependencies between letters, which is not surprising as we had informed them about the presence of a grammar before the last grammaticality classification. However, when the participants were provided with the correct dependency constraints, they performed at chance level when deciding whether these were correct in a forced-choice task (Crossed: 53% “Yes” 47% “No,” $p = 1.00$; Nested: 45% “Yes” 55% “No,” $p = .82$). Thus, there was little evidence for any explicit knowledge, consistent with the suggestion that the structural knowledge was implicitly acquired, as in Experiment 1.

4. General discussion

This study demonstrates the feasibility of successful acquisition of multiple non-adjacent dependencies in an implicit AGL paradigm without performance feedback. The implicit acquisition of non-adjacent dependencies showed no qualitative difference to the acquisition of adjacent dependencies. Specifically, at the end of acquisition, we observed similar overall proficiency and a similar boost in performance for the grammaticality compared with the preference instruction (see Figs. 3 and 5). Finally, significant explicit knowledge was absent in both cases. There was, however, one quantitative difference: Non-adjacent dependencies took some days longer to acquire. Thus, it seems that the acquisition of non-adjacent dependencies is a matter of time. This conclusion is consistent with children mastering non-adjacent dependencies later in the course of language acquisition compared with adjacent dependencies (Gómez & Maye, 2005). The quantitative difference might be related to the necessity for various abstraction and consolidation processes to occur. In a recent study, the effects of offline wake time and sleep on implicit AGL were disentangled by varying these parameters in a between-subjects comparison. The results suggest that implicit AGL depends on sleep specifically (Nieuwenhuis, Folia, Forkstam, Jensen, & Petersson, unpublished data). Overall, these results are consistent with the idea that common mechanisms support the acquisition of syntactic dependencies, whether adjacent or non-adjacent.

The same computational development can be viewed from a representational perspective. To acquire representations of adjacent dependencies, an intact memory trace will do, but to acquire a non-adjacent dependency, there is an apparent need to erase (i.e., abstract away) the intervening material from the representation. This is consistent with the fact that introducing variable intervening material improves the acquisition of non-adjacent dependencies (Gomez, 2002). However, we observed little evidence for a qualitative difference between

non-adjacent and adjacent acquisition behaviorally, which argues against large architectural differences in memory and processing of adjacent versus non-adjacent dependencies. Rather, we speculate that forming a representation of a non-adjacent dependency and thus abstracting away the intervening positions from the representation can be achieved, for example, by a single but bidirectional acquisition mechanism, at the biological level. There are several examples of bidirectional plasticity mechanisms, including spike-timing-dependent plasticity, at the neuronal synapses. For instance, a couple of synapse strengths in a neuronal population, modulated by the bidirectional mechanism of spike-timing-dependent plasticity, might at the same time represent frequently co-occurring non-adjacent elements and be flexible with respect to the intervening material by encoding the relatively infrequent co-occurrence of the elements constituting the varying intervening material with respect to the elements in the non-adjacent dependency. Another example mechanism is AMPA receptor modulation of synapse strength. For example, an increasing or decreasing number of AMPA receptors results in strengthening or weakening of the synapse (Kessels & Malinow, 2009; for a review, see Uddén, Folia, & Petersson, 2010).

Whatever the mechanisms might be, we have shown that subjects need more exposure to settle on a correct representation for non-adjacent compared with adjacent dependencies. Thus, a possible reason for reported acquisition failures for complex non-adjacent dependencies (Hochmann, Azadpour, & Mehler, 2008; de Vries et al., 2008) might be that subjects were not provided with sufficient time or exposure to grammatical items during acquisition. In other implicit AGL paradigms, when subjects were considerably less exposed to the grammar than in the present experiments, the general observation was that adjacent dependencies could be acquired, but that non-adjacent dependencies failed to be acquired (Johnstone & Shanks, 2001; Mathews et al., 1989). In one study using a so-called *inversion rule* with crossed dependencies in tone sequences, above chance preference classification after implicit learning of non-adjacent dependencies was demonstrated, also after brief exposure (Kuhn & Dienes, 2005). These results were replicated in one study (Dienes, in press) but failed to replicate in another follow-up study using the same rules but with slightly different materials (Desmet, Poulin-Charronnat, Lalitte, & Perruchet, 2009; Kuhn & Dienes, 2006; for comments, see Desmet et al., 2009). Irrespective of the robustness of these results, our findings extend the demonstration of implicit learning of multiple non-adjacent dependencies to robust performance levels similar to those observed with adjacent dependencies. Thus, we are capturing a later part of the acquisition phase compared with Kuhn and Dienes (2005). de Vries et al. (2008) did not observe any sensitivity to grammaticality when the dependency structure was identical to the nested dependency constraints used in the present study. In the study of de Vries et al. (2008), exposure was limited to one brief acquisition session, and several similar studies have yielded similar outcomes (Hochmann et al., 2008). This might also explain why European starlings acquired the A^nB^n grammar (Gentner, Fenn, Margoliash, & Nusbaum, 2006), whereas cotton top tamarins failed to do so (Fitch & Hauser, 2004). The European starlings received extensive training ($\sim 30,000$ trials on average), whereas the tamarins were only given 20 min of exposure on the day preceding testing.

There was a larger acquisition effect for crossed compared with nested non-adjacent dependencies. In this respect, the results speak against the push-down stack memory model.

However, the greater acquisition effect in the crossed group compared to the nested group can be explained by an initial sensitivity to the grammaticality factor in the nested group. The initial sensitivity was stronger in the second half of the baseline test, which points to early acquisition of sensitivity to at least some aspects of the nested grammar. Thus, the apparent advantage of acquisition of crossed compared with nested grammars can be explained by crossed grammars having fewer aspects that are immediately acquired when exposed to a set of both grammatical and non-grammatical sequences and thus an, at least initial, acquisition advantage for nested compared with crossed structures. Thus, from this point of view, the push-down stack model is supported.

An alternative to the push-down stack model is that human online memory (working memory) is characterized by (weighted) random access in combination with certain forgetting characteristics (e.g., primacy/recency effects). The pattern of serial position curves in the memory literature suggests that there are strong primacy and recency effects (Glanzer & Cunitz, 1966). Based on serial position effects, it can be argued that noun-verb pairs in the middle of the nested or crossed sentence (i.e., A_2 - B_2) should be more difficult to parse. However, our results from the crossed sequences showed that the first violation position was less salient than the middle one. The pattern of violation position results can be predicted by the push-down stack model, as the processing cost is higher at the first compared with the second and third positions (see Fig. 7), as well as the absence of an effect of violation position for nested sequences. The analysis of violation positions thus provides some evidence for the use of the push-down stack for processing crossed grammars and is consistent with its use in nested grammars. As a general caution, we note that it is possible that the results reported here might change if sequences were drawn from a significantly larger alphabet.

It has been proposed that the underlying neurophysiology of memory might have stack-like properties (Siegelmann, 1999). The brain represents information in terms of dynamical variables, for example, a set of membrane potentials, which evolve over time and therefore can be viewed as a set of dynamic registers (i.e., information is represented in the decimal expansion of the membrane potentials). Recurrent neural processing in combination with multiplication/division of the membrane potentials by synaptic weights can be shown to implement memory representations with stack-like properties in artificial neural networks (see e.g., Siegelmann, 1999). Thus, it might be fruitful to specify possible neural mechanisms for multiplication and division, which represent each other's inverse transformations. Remember that the operation "pop" is the removal of the top element (in this case, the top element is the first digit of the decimal expansion) from the stack and "push" is the storage of an element at the top of the stack. To implement "pop," we need (a) multiplication and (b) a specification how the popped element can be unmixed and separated from the rest of the decimal expansion. In recurrent neural networks models, the stack operations "push" and "pop" are implemented by division and multiplication with synaptic weights, respectively. Alternatively, multiplication/division and thus "pop"/"push" can be implemented at the level of dendrites (Koch & Poggio, 1985), at the single neuron level, or in networks of neurons. An approximate multiplicative synaptic mechanism is exemplified by *shunting inhibition* (Koch, 1999). In networks of neurons, multiplication can be

implemented in the linear excitatory and inhibitory connections between neurons that implement a logarithmic transfer function. If Neuron A receives input x_1 and outputs $\log(x_1)$ and Neuron B receives input x_2 and outputs $\log(x_2)$, then Neuron C, receiving both of their outputs (again assuming linear summation), will receive the input $\log(x_1x_2)$, by the logarithmic laws (Koch & Poggio, 1992). Together with our results, which lend some support for a memory model with stack-like properties, this sketch suggests that it is at least conceivable that online neural processing memory has stack-like neural memory properties. However, it is plausible that both memory architectures coexist. For instance, each biological entity that might implement a stack, for example, each membrane potential, could be used as a register in a random access memory.

Newport and Aslin (2004) discuss non-adjacent dependency processing in relation to the gestalt principle of similarity and the principle of phonological tiers. Both of these abstract mechanistic suggestions are ways of describing how the intervening material effectively can be separated, or erased, in the abstract representation (the tier or the gestalt). This results in the non-adjacent dependencies becoming adjacent dependencies. Support for these ideas comes from the fact that non-adjacent dependencies between vowels are easier to learn when the intervening material is composed of consonants and vice versa, compared to when the non-adjacent dependency and the intervening material is instantiated in the same type of elements (Newport & Aslin, 2004). The fact that the crossed non-adjacent dependencies in our study are successfully processed although the intervening material contains elements of the same type, and although the intervening material are parts of other crossing non-adjacent dependencies, poses a problem for this kind of explanation. There are no pre-established categories according to which the distant elements can be processed together as a gestalt/tier. Thus, if processing of non-adjacent dependencies is achieved by abstract representations of gestalts/tiers, these representations still have to be acquired. Postulating the existence of tiers might be a good solution when explaining proficient processing of non-adjacent dependencies, but this will not solve the question of how representations of new non-adjacent dependencies are acquired. To solve the acquisition problem, a mechanism for how the gestalts/tiers are created is needed, in addition. From this point of view, each letter pair would constitute a gestalt/tier, or more simply put: a representation of each non-adjacent dependency.

The stimulus materials in the current study do not allow us to conclusively exclude that the observed performance is related to the acquisition of so-called repetition structures, defined by Brooks and Vokey (1991). For example, “FFDLLP” and “DDFPPL” can be represented as “112334,” where the later, abstract representation, captures the structure of repetition of certain elements. The repetition structure is one arbitrary way to represent non-adjacent dependencies. However, this possibility is one of an infinite number of plausible alternatives. We are not making any claims concerning the exact nature of the representations that subjects acquired in our study. Moreover, in the light of subsequent experiments, where we have explored the generalization capacity of participants, it is clear that all information is not captured by repetition structures as defined by Brooks and Vokey (1991). Instead we conclude that (a) implicit acquisition can produce similar performance levels in preference and grammaticality classification of multiple non-adjacent and adjacent

dependencies, respectively; (b) there is no reason to postulate separate mechanisms for the acquisition of adjacent and non-adjacent dependencies; and (c) the evidence for the push-down stack model when comparing implicit acquisition of nested and crossed non-adjacent dependencies is ambiguous.

5. Summary and conclusions

We have presented the results from a complex implicit AGL paradigm that demonstrates robust implicit learning of multiple non-adjacent dependencies organized in nested as well as crossed dependency structures. The results showed little qualitative difference between the acquisition of these non-adjacent dependencies and adjacent dependencies acquired under similar conditions. The ecological validity of our implicit structural mere-exposure AGL paradigm in relation to natural syntax acquisition was improved by using longer acquisition periods and by minimizing the meaningfulness, motivation, and use of explicit strategies. The results of the present study extend earlier results from natural language processing, which suggest that there is lack of evidence for a push-down stack-type memory organization after extensive acquisition, although it is possible to interpret the results as supportive of a push-down stack-type memory during the early acquisition phase. When analyzing how the saliency of grammar violations depends on the violation position, the push-down stack model predicted the data well. The push-down stack model is thus partly supported by the results and we suggest stack-like properties are one natural property, among others, characterizing the underlying neurophysiological mechanisms that implement the online memory resources used in language and structured sequence processing.

Notes

1. There is, however, no deep distinction between an internal and external memory architecture, because (bounded) external memory architectures can always be implemented as an internal memory in a larger system.
2. Note that the question of whether there are specialized memory structures is independent of the details of how sequences are represented and whether the brain implements finite-state, context-free, and/or context-sensitive grammars.
3. More precisely mainly adjacent non-deterministic dependencies.
4. We note that *preferences increased for the non-grammatical items* with three non-adjacent dependencies *as an affect of acquisition*. This is an effect of the general trend that participants like sequences irrespective of grammatical status more after acquisition, which is present in both experiments, both for nested and crossed grammars. In this case, the general familiarity preference is stronger than the preference for grammatical compared with non-grammatical sequences.

Acknowledgments

This work was supported by Max Planck Institute for Psycholinguistics; Donders Institute for Brain, Cognition, and Behaviour; Stockholm Brain Institute; Fundação para a Ciência e Tecnologia (PTDC/PSI-PCO/110734/2009; IBB/CBME, LA, FEDER/POCI 2010); Vetenskapsrådet; Hedlunds Stiftelse; and Stockholm County Council (ALF, FoUU). We would like to thank Susana Arujo for her help with the data collection.

References

- Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes* 1, 249–262.
- Brooks, L. R., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: A comment on Reber and Mathews et al. *Journal of Experimental Psychology: General* 120, 316–323.
- Chomsky, N. (1963). Formal properties of grammars. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 323–418). New York: John Wiley.
- Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23, 157–205.
- Cutland, N. J. (1980). *Computability: An introduction to recursive function theory*. Cambridge, UK: Cambridge University Press.
- Davis, M., Sigal, R., & Weyuker, E. J. (1994). *Computability, complexity, and languages: Fundamentals of theoretical computer science*. San Diego, CA: Academic Press.
- Desmet, C., Poulin-Charronnat, B., Lalitte, P., & Perruchet, P. (2009). Implicit learning of nonlocal musical rules: A comment on Kuhn and Dienes (2005). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 299–305.
- Dienes, Z. (in press). Conscious versus unconscious learning of structure. In P. R. J. Williams (Ed.), *Statistical learning and language acquisition*. Amsterdam, The Netherlands: Mouton de Gruyter Publishers.
- Dienes, Z., & Longuet-Higgins, C. (2004). Can musical transformations be implicitly learned? *Cognitive Science*, 28, 531–558.
- Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303, 377–380.
- Folia, V., Forkstam, C., Ingvar, M., Hagoort, P., & Petersson, K. M. (2011). Implicit artificial syntax processing: Genes, preference, and bounded recursion. *Biolinguistics*, 5, 105–132.
- 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, 132–150.
- Forkstam, C., Elwér, Å., Ingvar, M., & Petersson, K. M. (2008). Instruction effects in implicit artificial grammar learning: A preference for grammaticality. *Brain Research*, 1221, 80–92.
- Forkstam, C., Hagoort, P., Fernandez, G., Ingvar, M., & Petersson, K. M. (2006). Neural correlates of artificial syntactic structure classification. *NeuroImage*, 32, 956–967.
- Forkstam, C., & Petersson, K. M. (2005). Towards an explicit account of implicit learning. *Current Opinion in Neurology*, 18, 435–441.
- Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, 440, 1204–1207.
- Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 5, 351–360.
- Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13, 431–436.

- Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, 7, 183–206.
- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, 9, 416–423.
- Hochmann, J.-R., Azadpour, M., & Mehler, J. (2008). Do humans really learn AnBn artificial grammars from exemplars? *Cognitive Science: A Multidisciplinary Journal*, 32, 1021–1036.
- Johnstone, T., & Shanks, D. R. (2001). Abstractionist and processing accounts of implicit learning. *Cognitive Psychology*, 42, 61–112.
- Kessels, H. W., & Malinow, R. (2009). Synaptic AMPA receptor plasticity and behavior. *Neuron*, 61, 340–350.
- Koch, C. (1999). *Biophysics of computation: Information processing in single neurons*. New York: Oxford University Press.
- Koch, C., & Poggio, T. (1985). A simple algorithm for solving the cable equation in dendritic trees of arbitrary geometry. *Journal of Neuroscience Methods*, 12, 303–315.
- Koch, C., & Poggio, T. (1992). Multiplying with synapses and neurons. In T. McKenna, J. Davis, & S. F. Zornetzer (Eds.), *Single neuron computation* (pp. 315–345). San Diego, CA: Academic Press Professional, Inc.
- Kuhn, G., & Dienes, Z. (2005). Implicit learning of nonlocal musical rules: Implicitly learning more than chunks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1417–1432.
- Kuhn, G., & Dienes, Z. (2006). Differences in the types of musical regularities learnt in incidental and intentional learning conditions. *Quarterly Journal of Experimental Psychology*, 59, 1725–1744.
- Levelt, W. J. M. (1974). *Formal grammars in linguistics and psycholinguistics*. (Vol. III, Psycholinguistic applications). The Hague, The Netherlands: Mouton.
- Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 1083–1100.
- Meulemans, T., & Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 1007–1028.
- Minsky, M. L. (1967). *Computation: Finite and infinite machines*. Englewood Cliffs, NJ: Prentice Hall.
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127–162.
- Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, 49, 85–117.
- Petersson, K. M. (2005). On the relevance of the neurobiological analogue of the finite-state architecture. *Neurocomputing*, 65–66, 825–832.
- Petersson, K. M. (2008). On cognition, structured sequence processing, and adaptive dynamical systems. *Proceedings of the American Institute of Physics*, 1060, 195–199.
- Petersson, K. M., Folia, V., & Hagoort, P. (2010). What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language*, 170, 155–164. doi:10.1016/j.bandl.2010.08.003
- Pullum, G. K., & Scholz, B. C. (2009). For universals (but not finite-state learning) visit the zoo. *Behavioral and Brain Sciences*, 32, 466–467.
- Pullum, G. K., & Scholz, B. C. (2010). Recursion and the infinitude claim. In H. v. d. Hulst (Ed.), *Recursion in human language* (pp. 1–9). Berlin: Mouton de Gruyter.
- Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 888–896.
- Savage, J. E. (1998). *Models of computation*. Reading, MA: Addison-Wesley.
- Shieber, S. M. (1985). Evidence against the context-freeness of natural language. *Linguistics and Philosophy*, 8, 333–343.
- Siegelmann, H. T. (1999). *Neural networks and analog computation: Beyond the Turing limit*. Basel, Switzerland: Birkhäuser.
- Stadler, M. A., & Frensch, P. A. E. (1998). *Handbook of implicit learning*. Thousand Oaks, CA: Sage Publications.

- Taylor, R. G. (1998). *Models of computation and formal languages*. Oxford, England: Oxford University Press.
- Turing, A. (1936). On computable numbers with an application to the Entscheidungs problem (part 1). *Proceedings of the London Mathematical Society*, 42, 230–240.
- Uddén, 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 Research*, 1224, 69–78.
- Uddén, J., Folia, V., & Petersson, K. M. (2010). Neuropharmacology of implicit learning. *Current Neuropharmacology*, 8, 367–381.
- Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75, 105–143.
- de Vries, M. H., Christiansen, M. H., & Petersson, K. M. (2011). Learning recursion: Multiple nested and crossed dependencies. *Biolinguistics*, 5, 10–35.
- 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, 763–774.
- Wells, A. (2005). *Rethinking cognitive computation: Turing and the science of the mind*. Hampshire, UK: Palgrave Macmillan.
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9, 1–27.