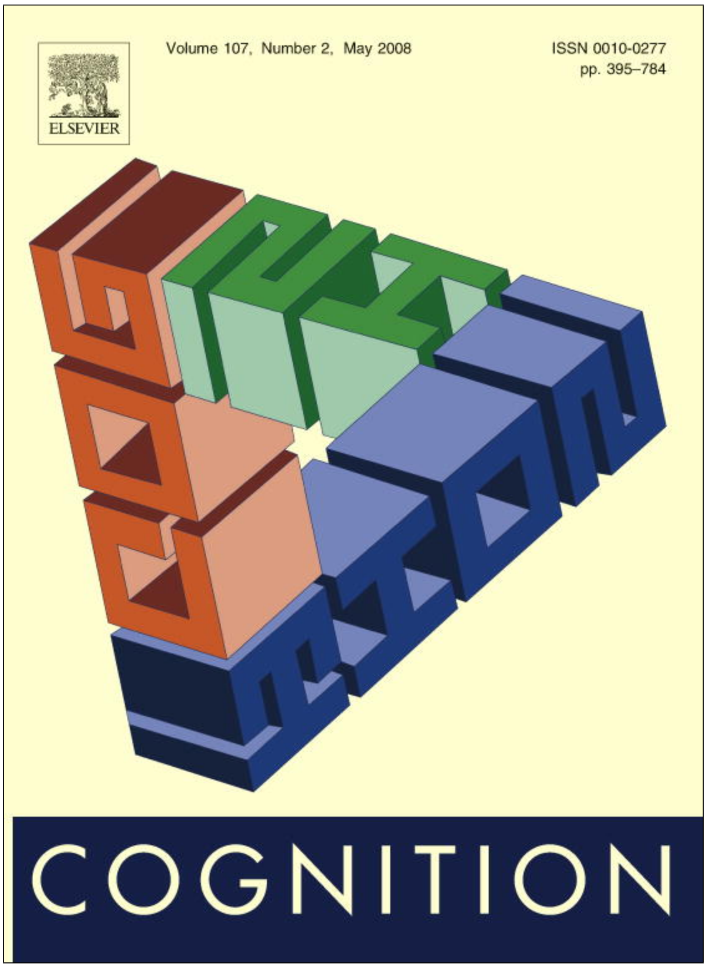
Volume 107, Number 2, May 2008
ISSN 0010-0277
pp. 395–784
ELSEVIER
COGNITION

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Cognition 107 (2008) 763–774
COGNITION
www.elsevier.com/locate/COGNIT

Brief article

Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures

Meinou H. de Vries [a,*], Padraic Monaghan [b], Stefan Knecht [a], Pienie Zwitserlood [c]

[a] *Department of Neurology, University of Münster, A. Schweitzerstrasse 33, D-48129 Münster, Germany*
[b] *Department of Psychology, University of York, UK*
[c] *Department of Psychology, University of Münster, Germany*

Received 20 September 2006; revised 3 September 2007; accepted 14 September 2007

Abstract

Embedded hierarchical structures, such as "the rat the cat ate was brown", constitute a core generative property of a natural language theory. Several recent studies have reported learning of hierarchical embeddings in artificial grammar learning (AGL) tasks, and described the functional specificity of Broca's area for processing such structures. In two experiments, we investigated whether alternative strategies can explain the learning success in these studies. We trained participants on hierarchical sequences, and found no evidence for the learning of hierarchical embeddings in test situations identical to those from other studies in the literature. Instead, participants appeared to solve the task by exploiting surface distinctions between legal and illegal sequences, and applying strategies such as counting or repetition detection. We suggest alternative interpretations for the observed activation of Broca's area, in terms of the application of calculation rules or of a differential role of working memory. We claim that the learnability of hierarchical embeddings in AGL tasks remains to be demonstrated.

---

[*] Corresponding author. Tel.: +49 2518345695; fax: +49 2518348181.
E-mail address: mdevries@uni-muenster.de (M.H. de Vries).

0010-0277/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2007.09.002

Keywords: Artificial grammar learning; Syntax; Context free grammar; Finite-state grammar; Centre embeddings; Hierarchical structure learning

1. Introduction

A fundamental issue in language acquisition research concerns which rules children develop as a part of their grammatical knowledge and how these rules may be discovered (e.g., Chomsky, 1957; Reali & Christiansen, 2005). Artificial grammar learning (AGL) is a potentially valuable paradigm for determining processes of rule learning, both in terms of what structures are learnable (e.g., Fitch & Hauser, 2004; Gentner, Fenn, Margoliash, & Nusbaum, 2006; Newport, Hauser, Spaepen, & Aslin, 2004), and which properties of the language facilitate learning of these structures (e.g., Gomez & Gerken, 2000; Newport & Aslin, 2004; Onnis, Monaghan, Richmond, & Chater, 2005).

A natural language structure that has attracted interest in recent AGL studies is hierarchical centre embeddings (Fitch & Hauser, 2004, henceforth F&H; Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006; Gentner et al., 2006; Perruchet & Rey, 2005). In English, structures exemplified by *The rat [the cat ate] was brown* illustrate such centre embeddings, with additional embeddings possible, e.g., *The rat [the cat [the boy chased] ate] was brown.* Critically, these centre-embedded structures establish dependencies between constituents. Thus, such sentences have the structure $A_3A_2A_1B_1B_2B_3$, where the index values indicate the dependency between A_i- and B_i-elements. Such hierarchical embeddings are notoriously difficult to process in natural language (Bach, Brown, & Marslen-Wilson, 1986; Blaubergs & Braine, 1974; Foss & Cairns, 1970). Thus, demonstrating their learnability in AGL-experiments is a notable success.

Hierarchical embeddings have also been claimed to be of theoretical importance, as they require a context free grammar[1] to generate them and have been the focus of studies of human-unique structures in artificial language learning (Fitch, Hauser, & Chomsky, 2005; Hauser, Chomsky, & Fitch, 2002; Premack, 2004). In this respect they have been classified as different from the structures generated by finite-state grammars, for which local transitional dependencies can generate the sequence. F&H observed that humans could discriminate AAABBB-syllable sequences from ABABAB-syllable sequences (finite-state grammar), where A-syllables were spoken by a male human voice and B-syllables were spoken by a female. In contrast, cotton-top tamarins were insensitive to this distinction (though see Perruchet & Rey, 2005 for an explanation in terms of biological relevance, rather than structural dis-

---

[1] Context free grammars also enable hierarchically embedded patterns that do not entail dependencies among constituent elements, however the relevance of these patterns to human language is questionable. The studies by Fitch and Hauser (2004) and Gentner et al. (2006), for instance, focused on sequences without such dependencies. In this paper, however, we will focus on sequences that do entail dependencies.

tinctions between species). F&H thus claimed that humans were sensitive to the distinction between context free and finite-state grammars,[2] whereas nonhuman primates were not.

Friederici et al. (2006) and Bahlmann and Friederici (2006) also contrasted learning of hierarchical ($A_3A_2A_1B_1B_2B_3$)[3] and finite-state grammar ($A_1B_1A_2B_2A_3B_3$) sequences. A_i- and B_i-syllables were distinguished in terms of phonological properties (see Section 3.1). They observed that processing of hierarchical embeddings selectively activated Broca's area (BA44/45) – typically involved in syntactic processing (see Kaan & Swaab, 2002) – whereas processing finite-state grammars selectively engaged the left frontal operculum. Broca's area is thought to be phylogenetically younger (Friederici, 2004) and, in these studies, was claimed to be functionally specific to processing hierarchical embeddings.

We argue here that the data from the studies reported above can be explained by alternative learning strategies which do not imply hierarchical embeddings, but, instead, involve counting and matching the number of A- and B-elements. The relevance of AGL to human language becomes obscure without explicitly testing learning of hierarchical embeddings, as otherwise these sequences may not probe linguistically relevant processing. Our arguments critically hinge on the materials used in the testing phases of AGL tasks. The illegal sequences during testing should differ only in terms of their hierarchical structure if this is the property being tested. We will show, however, that such violating sequences differ also in terms of surface features enabling alternative, non-linguistic strategies to be applied during learning. We present data indicating that participants do indeed use alternative strategies instead of learning the rules of hierarchical embeddings in AGL-tasks. As such strategies depend on information not present in natural language centre-embedding structures, we challenge the evidence provided for such processing using current AGL tasks.

2. Counting vs. hierarchical processing

In the AGL studies of context free grammars reported above, knowledge of the precise hierarchical connections between elements was not explicitly tested. In F&H, participants had to distinguish alternating male/female voices from male sequences followed by female sequences. Perruchet and Rey (2005) replicated this study, and found that participants were unable to distinguish $A_3A_2A_1B_1B_2B_3$ from

---

[2] In their paper, Fitch and Hauser (2004) refer to the hierarchical sequences as phrase structure grammars. This is formally correct as phrase structure grammars can generate such sequences, however a context free grammar is sufficient for generating such sequences, and is lower in the Chomsky hierarchy of languages (Chomsky & Schützenberger, 1963).

[3] The example here contains paired indices, which are the *sine qua non* condition for embedding hierarchical structure. Fitch and Hauser (2004) did not pair the indices in their "hierarchical" structures. It is unclear from their method section whether indices were paired in the study by Friederici et al. (2006). Bahlmann and Friederici (2006) did use paired indices.

$A_3A_2A_1B_1B_3B_2$ sequences, if in the latter the dependencies between hierarchical elements were broken but were not marked by pitch distinctions.

Friederici et al. (2006) tested participants' ability to distinguish $A_3A_2A_1B_1B_2B_3$ sequences from sequences where an A-syllable replaced a B-syllable, or vice versa. Participants learned to reject, for example, $A_1A_2A_3A_4B_2B_1$- and $A_1A_2A_3B_3B_2A_4$-sequences, where A- and B-syllables contained different vowels. So matching the number of A-syllables to B-syllables was perhaps sufficient to solve the task, without needing to encode dependencies between A_i- and B_i-syllables.

In Experiment 1, we tested which strategies were used to distinguish hierarchical sequences from violations using similar materials to the above studies. Our hypothesis (cf. Coleman, Kochanski, Rosner, & Grabe, 2004; Perruchet & Rey, 2005) was that people use strategies such as counting instead of learning the dependencies of the hierarchical sequences if available. We also tested whether hierarchical sequences *could* be learned when no alternative strategies are available to solve the task.

3. Experiment 1

To tease apart different strategies, we tested participants' learning of hierarchical sequences using an AGL, but varied the testing conditions to compare learning a counting strategy to hierarchical dependency learning. In this experiment, we replicated Friederici et al.'s (2006) study comparing hierarchical sequences to number-violating sequences. We tested whether the learning effect in this study was due to counting by removing the hierarchical dependencies in sequences. We also tested whether learning could occur if sequences were distinguished only by hierarchical dependencies.

3.1. Methods

3.1.1. Participants

Thirty students (18 female), aged 19–27, from the University of Münster, participated in the experiment. They received payment or course credit. All were native German speakers, right handed, and had normal or corrected-to-normal vision.

3.1.2. Materials

The same syllables were used as in Friederici et al. (2006), comprising a set of A-syllables {de, gi, le, ri, se, ne, ti, mi} and B-syllables {bo, fo, ku, mo, pu, wo, tu, gu}, distinguished by their vowels. The pairing was: de-bo, gi-fo, le-ku, ri-mo, se-pu, ne-wo, ti-tu, mi-gu. Probability of occurrence of syllables was balanced. Sequences consisted of 4, 6, or 8 syllables with hierarchical dependencies between syllables.

As in Friederici et al. (2006), counting-violating sequences in the test phase were formed by replacing an A-syllable with a B-syllable or vice versa (e.g., $A_3A_2A_1B_1B_2A_4$), occurring at different positions: In 4-syllable sequences at any position, in 6-syllable sequences at position 1, 3, 4, or 6, and in 8-syllable sequences at positions 1, 4, 5, or 8. Syllables were not repeated within a sequence.

3.1.3. Procedure

Participants were randomly assigned to one of three groups, 10 per group. All groups were trained with the same hierarchical sequences but crucially, the testing phase differed between groups. The "hierarchical-violations" (Hier-Viol) group was tested with hierarchical vs. counting-violating sequences, the "scrambled-violations" (Scram-Viol) group with ordered sequences of A and B syllables vs. counting-violations, and the "hierarchical-scrambled" (Hier-Scram) group with hierarchical vs. scrambled sequences (see Table 1). The Hier-Viol condition replicated Friederici et al.'s (2006) study.

The experiment consisted of twelve learning blocks of 10 hierarchical sequences, each followed by a testing block of 10 sequences, of which 5 were correct and 5 incorrect. For learning blocks, participants were instructed to extract the rule underlying the syllable sequences. Sequences were presented visually, with syllables presented successively. In testing blocks, participants had to decide whether sequences conformed to the rule, and responded by button press. Feedback was given. For the Scram-Viol-group, a correct response was recognizing that the numbers of As and Bs match, regardless of the particular order of As and Bs. Positive feedback was given if participants accepted these sequences and rejected sequences with different numbers of As and Bs. Afterwards, participants had to write down what rule they had learned from the language.

Sequences started with a fixation cross (500 ms), then syllables were presented successively for 300 ms, with a 200 ms inter-stimulus-interval. After the last syllable of a testing sequence, a response screen appeared, and a decision had to be made. Feedback was given for 500 ms. The experiment lasted approximately 50 min.

3.2. Results and discussion

We performed an ANOVA on the d'-values of the responses of each subject for the final testing block.[4] There was a main effect of Group (Hier-Viol, Scram-Viol, Hier-Scram), $F(2, 27) = 8.73$, $p = .001$, $\eta^2 = .39$ (see Fig. 1). Performance was significantly more accurate than chance for both the Hier-Viol-group ($d' = 2.24$, 73% correct) and Scram-Viol-group ($d' = 1.78$, 67% correct), $t(9) = 3.89$, $p < .005$, $d = 1.74$ and $t(9) = 4.17$, $p < .005$, $d = 1.87$, respectively. Responses for the Hier-Scram group ($d' = -.21$, 48% correct) were not significantly different from chance, $t(9) = -.79$, $p = .45$, $d = -.36$.

T-tests (justified by the Kolmogorov–Smirnov test, $p > .05$ for all groups) on d'-values revealed a significant difference between the Hier-Viol- and Hier-Scram-group, $t(18) = 6.08$, $p < .001$, $d = 1.73$, and between the Scram-Viol- and Hier-Scram-group, $t(18) = 1.99$, $p = <.001$, $d = 3.04$, but not between the Hier-Viol- and Scram-Viol-group, $t(18) = 1.12$, $p = .55$, $d = -.33$, suggesting that a strategy

---

[4] The distribution of d' values met the assumption of normality, $p > .05$ for Kolmogorov–Smirnov tests for each group.

Table 1
Conditions in Experiment 1

Group	Training phase	Test condition
Hier-Viol	Hierarchical structure (e.g., $A_3A_2A_1B_1B_2B_3$)	Hierarchical structure (e.g., $A_3A_2A_1B_1B_2B_3$) vs. violations (e.g., $A_3A_2A_1A_4B_2B_3$)
Scram-Viol	Hierarchical structure (e.g., $A_3A_2A_1B_1B_2B_3$)	Scrambled structure (e.g., $A_1A_2A_3B_1B_3B_2$) vs. violations (e.g., $A_3A_2A_1A_4B_2B_3$)
Hier-Scram	Hierarchical structure (e.g., $A_3A_2A_1B_1B_2B_3$)	Hierarchical structure (e.g., $A_3A_2A_1B_1B_2B_3$) vs. scrambled structure (e.g., $A_1A_2A_3B_1B_3B_2$)

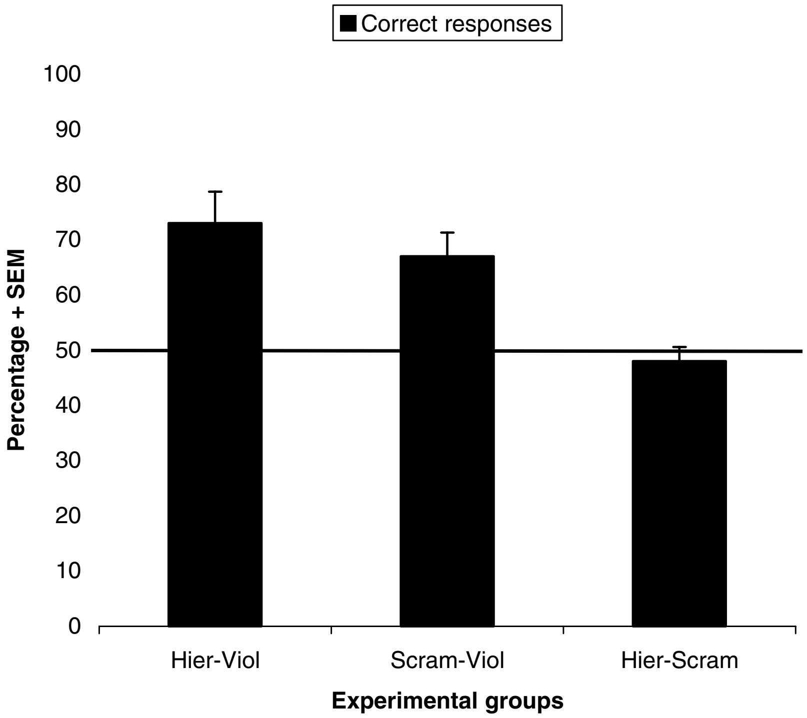

Fig. 1. Experiment 1: Mean correct responses by Group. Error bars indicate standard error of the mean.

used to solve Hier-Viol and Hier-Scram sequences did not apply in the case of making Scram-Viol distinctions.

The data from the Hier-Viol-group replicated Friederici et al.'s results (2006). But the data from the Scram-Viol-group and Hier-Scram-group suggested that no hierarchical sequence learning took place. The similarity of performance for the Hier-Viol-group and the Scram-Viol-group suggests a similar strategy for solving both these tasks. When this strategy fails, as in the Hier-Scram-group, performance was close to chance. This interpretation is consistent with participants' written comments: 12 of 20 participants in the Hier-Viol- and Scram-Viol-group explicitly reported that e/i-syllables only occurred at the beginning and o/u-syllables at the end. Ten reported the number of e/i-syllables had to match the number of o/u-syllables. None reported detecting hierarchical dependencies.

The results above were for the final test block, for ease of comparison with the results of Friederici et al. (2006). The results did not differ significantly across the testing blocks. Just as for the final block analyses, for the Hier-Viol-group and Scram-Viol-group, the summed correct responses across all testing blocks did not differ significantly from each other, $t(22) = .79$, $p = .44$, $d = .32$, and both were again above chance, $t(11) = 6.02$, $p < .001$, $d = 2.46$ and $t(11) = 7.12$, $p < .001$, $d = 2.92$, respectively. The summed responses in the Hier-Scram-group did not significantly differ from chance, $t(11) = -.51$, $p = .62$, $d = -.21$. An ANOVA on d' values of all responses per subject with trial position (1–10) as a within-subjects factor revealed no significant main effect, $F(9, 190) = 1.81$, $p = .12$, $\eta^2 = .08$, indicating that there was no effect of feedback within each learning phase. Furthermore, there were no significant differences between accuracy for 4-, 6-, and 8-syllable sequences: an ANOVA with within-subjects factor sequence length revealed no significant main effect, $F(2, 87) = .39$, $p = .68$, $\eta^2 = .01$), which is compatible with our prediction that participants used strategies immune to increasing hierarchical complexity.

Counting is not the only possible strategy here. Of the two kinds of violations, $A_3A_2A_1B_1B_2A_4$ and $A_3A_2A_1A_4B_2B_3$, the former could have been rejected just by monitoring a transition from B to A – another strategy irrelevant to learning the hierarchical structure of the language. Indeed, incorrect responses to violations, throughout the experiment, in both Hier-Viol- and Scram-Viol-groups, show that violations like $A_3A_2A_1B_1B_2A_4$ are significantly easier to detect than $A_3A_2A_1A_4B_2B_3$, $t(38) = 5.52$, $p < .001$, $d = 1.75$.

Bahlmann and Friederici (2006, henceforth B&F) added phonological information to support the dependency relation between syllables, which may be a critical support for learning dependencies. Experiment 2 trained participants on A_i–B_i pairs which shared phonological properties.

4. Experiment 2

4.1. Methods

4.1.1. Participants

Ten students (5 female), aged 19–27, from the University of Münster participated for payment or course credit. All were native German speakers, right handed, and had normal or corrected to normal vision. None had participated in Experiment 1.

4.1.2. Materials

Participants were trained on hierarchical sequences. A-syllables began with voiced plosives and ended with –e/–i, and B-syllables began with voiceless plosives and ended with –o/–u. The plosives were paired according to their place of articulation: b-p, g-k, d-t. This yielded A-syllables {de, di, be, bi, ge, gi} and B-syllables {tu, to, pu, po, ku, ko}, which were paired as follows: d(e/i)-t(o/u), g(e/i)-k(o/u), b(e/i)-p(o/u), exactly as in B&F. The order of the syllable sequences was varied across participants. B&F tested learning of hierarchical sequences against two violation-types: First, ''consonant–vowel'' violations, where the occurrence of consonant–vowel pairs was disordered, e.g., $A_2A_1B_1A_2$, for short, and $A_3A_2A_1B_1B_2A_3$ for long sequences, which could be distinguished by a counting strategy. Second, ''plosive–concatenation'' violations, where the match between particular plosive types was disordered, e.g., $A_2A_1B_1B_3$ for short, and $A_3A_1A_2B_2B_1B_2$ for long sequences. Note that these longer sequences contained repeated syllables in B&F's materials.

There were two testing stages. The first compared hierarchical to scrambled sequences (Hier-Scram) where counting and monitoring repetitions were not possible strategies. In the second, we included repeated syllables in violating sequences, as in the violating indices sequences from B&F (Hier-Scram+Rep), e.g., $A_1A_2A_3B_1B_2B_1$ (Table 2). Although the materials were similar, the learning procedure differed from B&F's study in that their participants were trained on shorter sequences first. Our procedure was identical to Experiment 1.

4.1.3. Results and discussion

Analyses were again restricted to the final test block. Performance on Hier-Scram sequences was not significantly different from chance ($d' = -.16$, 57% correct), for d' values, $t(9) = -.21$, $p = .84$, $d = -.93$. Performance was not significantly different to the Hier-Scram-group of Experiment 1, $t(18) = 3.79$, $p = .95$, $d = -.03$. Participants did not learn to distinguish hierarchical from scrambled sequences even in the presence of phonological cues to pairings, providing no evidence for the learning of hierarchical embeddings.

For the testing phase that included syllable repetitions (Hier-Scram+Rep), performance was above chance ($d' = 2.27$, 78% correct, see Fig. 2), $t(9) = 6.16$, $p < .001$,

Table 2
Conditions in Experiment 2

Group	Training phase	Test condition
Hier-Scram	Hierarchical structure (e.g., $A_3A_2A_1B_1B_2B_3$)	Hierarchical structure (e.g., $A_3A_2A_1B_1B_2B_3$) vs. scrambled structure (e.g., $A_1A_2A_3B_1B_3B_2$)
Hier-Scram+Rep	Hierarchical structure (e.g., $A_3A_2A_1B_1B_2B_3$)	Hierarchical structure (e.g., $A_3A_2A_1B_1B_2B_3$) vs. scrambled structure with repetitions (e.g., $A_1A_2A_3B_1B_3B_1$)

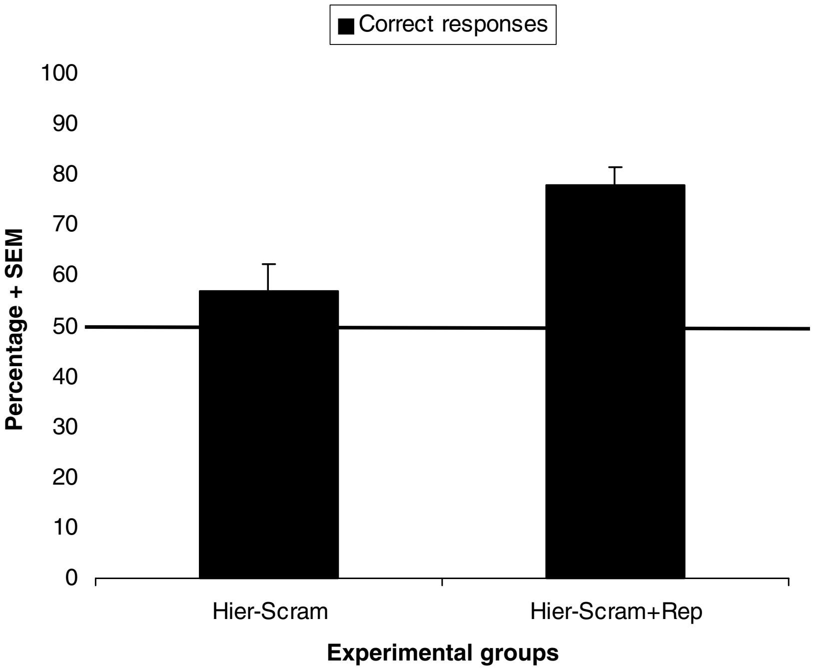

Fig. 2. Experiment 2: Mean correct responses by Group. Error bars indicate standard error of the mean.

$d = 2.75$. The d'-values of the Hier-Scram- and the Hier-Scram+Rep-groups were significantly different from each other, $t(18) = 1.75$, $p < .05$, $d = -1.30$. Participants in B&F's study trained their participants until they had reached 90% correct responses, and began training on short sequences first, possibly explaining the difference with our results. However, note that for solving all "consonant–vowel" violations, counting suffices, and that for solving the six-syllable "plosive–concatenation" violations, a repetition detection strategy could have been applied, and would be sufficient for 87.5% correct performance. Hence, learning the dependencies in the sequences was not necessary for each item to decide whether it was grammatical or not.

5. General discussion

What do participants learn from AGL-tasks when trained with hierarchical sequences? We found no evidence for the learning of hierarchical embeddings in Experiments 1 or 2. This is clear from the data for the Hier-Scram-group of Experiment 1: Participants could not discriminate structures without hierarchical dependencies (scrambled structure) from those requiring dependency-learning (hierarchical structure). Moreover, the same strategy appeared to be used to distinguish both hierarchical and scrambled sequences from violations: performance in the Hier-Viol- and Scram-Viol-group was similar. This suggests the prevalence of a strategy such as counting as a response to the AGL-task. We also found no evidence of learning hierarchical embeddings when participants were trained on structures with more salient links between A_i and B_i syllables in Experiment 2. Thus, instead of learning hierarchical embeddings, participants rather switch to alternative strategies, such as counting. Such a suggestion is not new (Coleman et al., 2004; Liberman, 2004), but, to our knowledge, we provide the first direct evidence that participants indeed use these alternative strategies.

Of course, we cannot know what participants in other studies really learned, but the pattern of behaviour in our experiments suggests that previous studies of hierarchical embeddings learning and their neural correlates may have instead examined participants' performance on distinguishing sequences based on properties other than their linguistic structure. Although Broca's area is implicated in syntactic processing, it is possible that the Broca's area activation reported by Friederici et al. (2006) and B&F for hierarchical embeddings is instead due to differential memory requirements for applying a strategy such as counting or repetition-monitoring, which contrasts with processing the local dependencies in finite-state sequences. This argument applies equally when brain activation is measured only for correct sequences, as in B&F. Broca's area has, indeed, been found to be activated for tasks involving meaningful symbolic operations and application of calculation rules (Gruber, Indefrey, Steinmetz, & Kleinschmidt, 2001; Hinton, Harrington, Binder, Durgerian, & Rao, 2004).

Broca's area has also been implicated in tasks involving working memory load in both linguistic and non-linguistic tasks (cf. Fiebach, Schlesewsky, Lohmann, Von Cramon, & Friederici, 2005; Paulesu, Frith, & Frackowiak, 1993), and working memory is crucial in simple arithmetic operations like addition (Logie, Gilhooly, & Wynn, 1994). Moreover, a linear relationship has been found between working memory load and activity in Broca's area, with a pure working-memory task that did not involve syntactic learning (Braver et al., 1997). These findings could provide an alternative explanation as to why Broca's area activation was found when processing hierarchical embeddings and not when processing finite-state grammars (Bahlmann & Friederici, 2006; Friederici et al., 2006): For the latter sequences, keeping elements available in memory is not required. However, one must keep in mind that Broca's area is not a unified area and that it is part of a larger network in the cited studies.

Our study indicates that there is as yet no firm evidence for the learning of these structures in AGL-studies. Training participants on complex natural language structures is no guarantee that they learn these structures. To ensure that such learning is assessed, participants must be tested with violating sequences that can only be distinguished from rule-conforming stimuli on the basis of the structural property in question. We have shown that using alternative strategies produced results which are interpreted elsewhere as evidence for learning of hierarchical embeddings. These alternative strategies are not directly relevant to language processing, and, consequently, AGL-tasks to which such strategies can be applied provide little or no insight into language processing: For AGL-tasks to be relevant to language processing, we need to ensure that similar mechanisms are engaged by the artificial sequence and by natural language structure. It is important to note that claims about relevance of AGL to natural language are hypothetical, as it is not yet clear how natural language centre embeddings are processed. However, our results suggest that counting and repetition are not candidate strategies, since natural language does not provide repetitions or countable sequences in order for such strategies to be applied to centre embedded sentences. To conclude, we believe it remains a challenge to the field to demonstrate the learnability of hierarchical embeddings using the AGL paradigm, and also a further challenge to determine the computations involved in *natural language* processing of such structures.

Acknowledgements

This research was supported by an EU Sixth Framework Marie Curie Research Training Network Program on Language and Brain: http://www.hull.ac.uk/RTN-LAB/. Meinou de Vries, Padraic Monaghan, and Stefan Knecht are members of this training network. Many thanks to Pierre Perruchet for helpful discussions and suggestions with respect to Experiment 2, to Isabel Ellerbrock, Christin Döpke, Anna-Victoria Schmidt, Julia Elen Beumler, and Anne Schürmann for their valuable assistance with the experiments, and to two anonymous reviewers for their helpful comments.

References

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

Bahlmann, J., & Friederici, A.D. (2006). FMRI investigation of the processing of simple linear and embedded hierarchical structures: An artificial grammar task. *Journal of Cognitive Neuroscience*, Annual Meeting Supplement, p.126.

Blaubergs, M. S., & Braine, M. D. S. (1974). Short-term memory limitations on decoding self-embedded sentences. *Journal of Experimental Psychology, 102*, 745–748.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage, 5*, 49–62.

Chomsky, N. (1957). *Syntactic structures*. The Hague, The Netherlands: Mouton.

Chomsky, N., & Schützenberger, M. P. (1963). The algebraic theory of context free languages. In P. Braffort & D. Hirschberg (Eds.), *Computer programming and formal languages* (pp. 118–161). Amsterdam: North Holland.

Coleman, J., Kochanski, G., Rosner, B., & Grabe, E. (2004). Available from http://kochanski.org/gpk/papers/2004/FitchHauser/FitchHauserScienceLetter.pdf.

Fiebach, C. J., Schlesewsky, M., Lohmann, G., Von Cramon, D. Y., & Friederici, A. D. (2005). Revisiting the role of Broca's area in sentence processing: Syntactic integration versus syntactic working memory. *Human Brain Mapping, 24*, 79–91.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science, 303*, 377–380.

Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition, 97*, 179–210.

Foss, D. J., & Cairns, H. S. (1970). Some effects of memory limitations upon sentence comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 9*, 541–547.

Friederici, A. D. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends in Cognitive Sciences, 8*, 245–247.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences, 103*, 2458–2463.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature, 440*, 1204–1207.

Gomez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*(5), 178–186.

Gruber, O., Indefrey, P., Steinmetz, H., & Kleinschmidt, A. (2001). Dissociating neural correlates of cognitive components in mental calculation. *Cerebral Cortex, 11*, 350–359.

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.

Hinton, S. C., Harrington, D. L., Binder, J. R., Durgerian, S., & Rao, S. M. (2004). Neural systems supporting timing and chronometric counting: An FMRI study. *Cognitive Brain Research, 21*, 183–192.

Kaan, E., & Swaab, T. Y. (2002). The brain circuitry of syntactic comprehension. *Trends in Cognitive Sciences, 6*(8), 350–356.

Liberman, M. (2004). Available from http://itre.cis.upenn.edu/~myl/languagelog/archives/000355.html.

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory and Cognition, 22*(4), 395–410.

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

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in speech processing. *Journal of Memory and Language, 53*, 225–237.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature, 362*, 342–345.

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

Premack, D. (2004). Is language the key to human intelligence? *Science, 303*, 318–320.

Reali, F., & Christiansen, M. H. (2005). Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science, 29*, 1007–1028.