The Official Journal of the International Society on Infant Studies
Infancy

Infancy, ***19***(4), 409–425, 2014

ISSN: 1525-0008 print / 1532-7078 online
DOI: 10.1111/infa.12052

Hierarchical Processing in Seven-Month-Old Infants

Ágnes M. Kovács
Cognitive Development Centre
Central European University

Ansgar D. Endress
Department of Technology
Universitat Pompeu Fabra
Department of Psychology
City University London

Hierarchical structures are crucial to many aspects of cognitive processing and especially for language. However, there still is little experimental support for the ability of infants to learn such structures. Here, we show that, with structures simple enough to be processed by various animals, seven-month-old infants seem to learn hierarchical relations. Infants were presented with an artificial language composed of "sentences" made of three-syllable "words." The syllables within words conformed to repetition patterns based on syllable tokens involving either adjacent repetitions (e.g., *dubaba*) or nonadjacent repetitions (e.g., ***du**ba**du***). Importantly, the sequence of word structures in each sentence conformed to repetition patterns based on word types (e.g., *aba-**abb-abb***). Infants learned this repetition pattern of repetition patterns and thus likely a hierarchical pattern based on repetitions, but only when the repeated word structure was based on adjacent repetitions. While our results leave open the question of which exact sentence-level pattern infants learned, they suggest that infants embedded the word-level patterns into a higher-level pattern and thus seemed to acquire a hierarchically embedded pattern.

Correspondence should be sent to Ágnes M. Kovács, Cognitive Development Centre, CEU, Hattyu u 14, Budapest H-1015, Hungary. E-mail: kovacsag@ceu.hu

From "high-level" processes such as planning complex action sequences to the "low-level" processes of visual processing or motor control, cognitive processes are hierarchically organized (Byrne & Russon, 1998; Jackendoff, 2007; Lashley, 1951; Marr, 1982; Norman & Shallice, 1986; Poggio & Bizzi, 2004). Hierarchical organization is especially important for language. For example, sentences contain words, which often contain syllables, which in turn contain phonemes. Or, in the case of syntax, sentences can contain different kinds of phrases (e.g., verb phrases) that have their own constituents, and the meaning of a sentence is determined by the specific combination of its constituents as well as the hierarchical relationships between them (Chomsky, 1957; Fodor, 1975; Hauser, Chomsky, & Fitch, 2002). For example, the sentence "I see the man with the telescope" famously can refer either to the speaker seeing a man who is equipped with a telescope or to the speaker seeing a man *through* a telescope. The intended meaning depends entirely on the hierarchical relations between "with the telescope" and the other constituents of the sentence.

Interest in hierarchical processing has increased following Hauser et al's (2002) proposal that many capacities used by the language faculty have homologues or analogues in nonlinguistic domains and nonhuman animals, but what sets humans apart from other animals, and what makes language possible in humans, might be the availability of computational machinery that can, in principle, keep track of an unbounded number of hierarchical levels, at least when abstracting away from memory constraints (but see Pinker & Jackendoff, 2005).

However, despite its importance, it has been surprisingly hard to demonstrate experimentally that humans can learn hierarchical structures in the domain of language. Some hierarchical processes are certainly available to young infants. For example, using visual stimuli, Rosenberg and Feigenson (2013) showed that 14-month-olds use hierarchical processes for memory retrieval, by hierarchically chunking the items they had to remember.

However, there is little experimental evidence for hierarchical structure building in infancy, especially regarding linguistic stimuli. For example, some of this evidence built on the proposal that infants can learn words from the speech they encounter by tracking which syllables tend to co-occur and led to the conclusion that such word-learning mechanisms can be deployed hierarchically. Specifically, if a syllable like "ba" has a high "transition" probability of being followed by a syllable like "by," infants might conclude that "baby" forms a word and commit it to memory (Aslin, Saffran, & Newport, 1998; Saffran, Newport, & Aslin, 1996a; Saffran, Aslin, & Newport, 1996b; but see Endress & Mehler, 2009; Endress & Hauser, 2010; Yang, 2004). Saffran and Wilson (2003) proposed that infants can deploy these statistical processes hierarchically to first discover syllables that go together to form words and then, at a higher level, to discover words (i.e., the syllable combinations discovered on the previous level) that go together to form sentence-like structures. However, learners are not only sensitive to transition probabilities between immediately adjacent syllables, but also between syllables across intervening syllables (i.e., second-order transition probabilities, Peña, Bonatti, Nespor, & Mehler, 2002; Endress & Bonatti, 2007). In Saffran and Wilson's (2003) experiments, the legal test "sentences" differed from the illegal ones not only with respect to the (hierarchically defined) transition probabilities among words, but also with respect to the transition probabilities among nonadjacent syllables in the sentences. This makes it difficult to decide whether infants engaged in hierarchical processing, or whether they relied on second-order statistics.

Likewise, Fitch and Hauser (2004) proposed that humans, but not other primates, can learn "grammars" involving hierarchical processing. Specifically, they proposed that human adults, but not cotton-top tamarin monkeys (*Saguinus oedipus*), can learn A^nB^n grammars, where a number of A items have to be followed by an equal number of B items (e.g., *pupupulilili*). However, subsequent studies suggest that Fitch and Hauser's (2004) results can be explained by alternative strategies not relying on hierarchical processing (van Heijningen, de Visser, Zuidema, & ten Cate, 2009; Hochmann, Azadpour, & Mehler, 2008; Perruchet & Rey, 2005; de Vries, Monaghan, Knecht, & Zwitserlood, 2008).

It thus appears that, despite a substantial interest in hierarchical learning, there is still little evidence that humans (and particularly young infants) can indeed perform such computations in an experimental setting.

Here, we ask whether infants as young as seven months can learn hierarchical, embedded patterns based on identity relations. We chose repetition-based patterns because such patterns are a widely used test case for rule learning in infancy (Gómez & Gerken, 1999; Marcus, Vijayan, Rao, & Vishton, 1999; Saffran, Pollak, Seibel, & Shkolnik, 2007). For example, Marcus et al. (1999) showed that 7-month-old infants can efficiently extract repetition patterns among syllables; infants notice that, in items such as *wo-wo-fe*, the first two syllables are identical, while in *wo-fe-wo*, the first and the last syllables are identical. While repetition-based patterns might be processed more easily with linguistic stimuli (Marcus, Fernandes, & Johnson, 2007), infants process such patterns in other domains as well, including music and vision (Dawson & Gerken, 2009; Saffran et al., 2007). Furthermore, rhesus monkeys, rats, and even bees are sensitive to such repetition patterns (Giurfa, Zhang, Jenett, Menzel, & Srinivasan, 2001;

Hauser & Glynn, 2009; Murphy, Mondragon, & Murphy, 2008). Hence, while repetition-based patterns seem to be particularly easy to process (Endress, Dehaene-Lambertz, & Mehler, 2007), they do not seem to rely on specifically linguistic processes but rather on a "repetition detector" (Endress et al., 2007; Endress, Nespor, & Mehler, 2009; see also Brooks & Vokey, 1989; Gómez & Gerken, 2000; Gómez, Gerken, & Schvaneveldt, 2000; Tunney & Altmann, 2001, for similar conclusions in the artificial grammar learning literature).

In the experiments below, we will ask whether infants can process hierarchical patterns based on repetitions in two situations with graded difficulty (see below). In the familiarization phase of both studies, infants were presented with artificial "sentences" composed of three words. The sentence patterns are shown in Figure 1. Each of these words conformed to one of two patterns based on repetitions; words followed either an *abb* pattern (where the second syllable was the same as the third one as in *du**baba***) or an *aba* pattern (where the first syllable was the same as the last one, as in ***du**ba**du***).

On top of the repetition-based word-level patterns, the sentences also followed one of two repetition-based patterns. For half of the infants, the first two *words* had the same pattern while the last word had a different

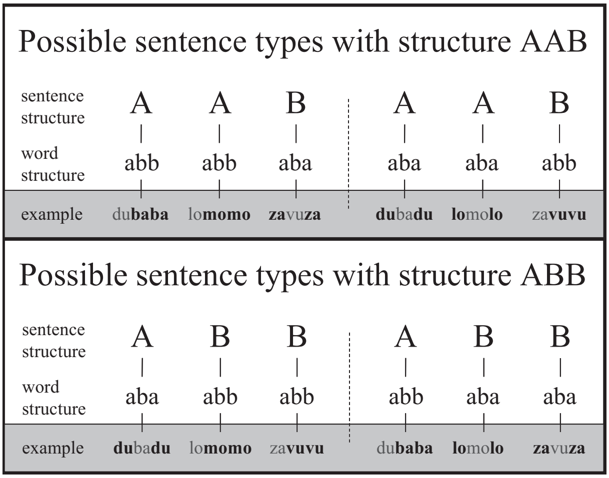

Figure 1 Structures used in the familiarization phase of Experiments 1 and 2. Both words and sentences conformed to certain structures. Words followed either an *abb* pattern (e.g., *du**baba***) or an *aba* pattern (e.g., ***du**ba**du***). (Top) For half of the participants, sentences had the structure *AAB*. (Bottom) For the remaining participants, sentences had the sentence-level structure *ABB*.

pattern. These sentences thus followed an AAB pattern (where each letter stands for the *pattern* of a word). In the following, we will use capitals for the sentence-level patterns and small letters for the word-level patterns. For the second group of infants, the sentences followed an ABB pattern; the last two words thus had the same pattern, while the first word had a different pattern. Importantly, all three words in a sentence were different; the sentence-level pattern thus was implemented through repetitions of word-level *patterns*, and not simply through repetitions of words. Hence, infants had not only to detect repeated syllables as in earlier studies (Gómez & Gerken, 1999; Marcus et al., 1999), but rather repetitions of *patterns* that were carried by different words with different syllables. For instance, an AAB sentence could be composed of the words *dubaba lomomo zavuza*; that is, while all three words were different, the first two shared a pattern and the third had a different pattern.

Infants were first familiarized with a sequence of sentences with one of the sentence-level patterns mentioned above (ABB or AAB). For example, half of the infants in Experiments 1 and 2 were familiarized with the sentence pattern AAB (see Table 1, Participant Group 1). In terms of word-level patterns, these sentences had either the pattern ***aba–aba***–*abb* or ***abb–abb***–*aba* (both conforming to the sentence pattern AAB). This design ensured that adjacent and nonadjacent syllable repetitions occurred equally often in each position within a sentence. The remaining infants in Experiments 1 and 2 (Participant Group 2) were familiarized with ABB

TABLE 1
Familiarization Sentences of Experiments 1 and 2. All Items are Given in SAMPA Transcription

Sentence structure	*Word structure*	*Sequence*
Participant group 1		
AAB	aba-aba-abb	dubadu lomolo zavuvu
AAB	abb-abb-aba	dubaba lomomo zavuza
AAB	aba-aba-abb	lovulo zabaza dumomo
AAB	abb-abb-aba	lovuvu zababa dumodu
AAB	aba-aba-abb	zamoza duvudu lobaba
AAB	abb-abb-aba	zamomo duvuvu lobalo
Participant group 2		
ABB	aba-abb-abb	dubadu lomomo zavuvu
ABB	abb-aba-aba	dubaba lomolo zavuza
ABB	aba-abb-abb	lovulo zababa dumomo
ABB	abb-aba-aba	lovuvu zabaza dumodu
ABB	aba-abb-abb	zamoza duvuvu lobaba
ABB	abb-aba-aba	zamomo duvudu lobalo

sentences, which contained the same words following an ABB sentence pattern.

Following this familiarization, infants were tested on new sentences with new syllables and new words that either agreed with or violated the sentence pattern (ABB and AAB sentences, see Table 2), and their looking times were recorded. If infants can learn the hierarchical pattern they were familiarized with, they should discriminate novel sentences conforming to these patterns from novel sentences violating them and thus look longer while listening to violations than to "legal" sentences.

The difference between Experiments 1 and 2 lies in the saliency of the word-level patterns. Specifically, repetitions of adjacent syllables are easier to process than repetitions of nonadjacent syllables (Gervain, Macagno, Cogoi, Peña, & Mehler, 2008; Kovács & Mehler, 2009). In Experiment 1, we thus present infants with a stronger test of hierarchical processing, using embedded repeated patterns based on nonadjacent identity relations. In Experiment 2, we present infants with a somewhat easier task and ask whether they would generalize embedded repeated patterns based on adjacent identity relations.

Specifically, during the test phase of Experiment 1, the word pattern occurring twice was *aba* (as in ***tipeti***), leading to test sentences of the form ***aba-aba-****abb*. During the test phase of Experiment 2, in contrast, the word pattern occurring twice was *abb* (as in *ti**pepe***), leading to sentences of the form ***abb-abb-****aba*. Importantly, during the familiarization phases of both experiments, both *abb* and *aba* could occur twice. Based on previous results suggesting that patterns involving repetitions between nonadjacent syllables (as in *aba*) are harder to process than patterns with repetitions between adjacent items (as in *abb*; Kovács & Mehler, 2009; Gervain et al.,

TABLE 2
Test Sentences of Experiments 1 and 2. All Items are Given in SAMPA Transcription

Experiment	*Sentence structure*	*Word structure*	*Sequence*
1	AAB	aba-aba-abb	tipeti RejERe fEsisi
	AAB	aba-aba-abb	ResiRe fEpefE tijEjE
	ABB	abb-aba-aba	tipepe RejERe fEsifE
	ABB	abb-aba-aba	Resisi fEpefE tijEti
2	AAB	abb-abb-aba	tipepe RejEjE fEsifE
	AAB	abb-abb-aba	Resisi fEpepe tijEti
	ABB	aba-abb-abb	tipeti RejEjE fEsisi
	ABB	aba-abb-abb	ResiRe fEpepe tijEjE

2008), infants might show a better performance in Experiment 2 than in Experiment 1.[1]

Our main prediction for the experiments below is that infants listen differentially to sentences consistent with the familiarization pattern and to sentences inconsistent with it. While we would expect them to listen longer to items violating the hierarchical patterns than to items conforming to them, it is sometimes suggested that infants should have a preference for familiar over novel items when the learning task is complex. Hence, one might expect infants to listen longer to items that conform to the hierarchical pattern. However, familiarity preference is quite rare in general, and in other experiments with fairly complex stimuli, infants showed longer looking times for violations as well (Gómez, 2002). Hence, our default prediction is that infants might listen longer to inconsistent items.

EXPERIMENT 1

Materials and method

Participants

Fourteen new healthy full-term 7-month-olds (mean age: 7.21, range 7.06–7.29, seven females), randomly assigned to one of two familiarization streams (see below), were retained for the analysis. An additional six infants did not complete the experiment due to fussiness.

Apparatus

Infants were tested using a modified head-turn preference procedure (Kemler-Nelson et al., 1995). Infants were seated on a caretaker's lap in a dimly lit, sound-attenuated room with three monitors, one facing the infants and two on their sides. Auditory stimuli were presented from a loudspeaker behind each monitor. Caretakers listened to masking music.

[1]We manipulated the repeated lower-level structure only in the test phase, but not during familiarization, because using only one of them during familiarization would have allowed infants to distinguish legal sentences from illegal ones without computing any hierarchical patterns: They could have succeeded by just monitoring the position of repetitions among syllables in the first words of the sentences. For example, if the repeated structure had always been *abb* in familiarization, all AAB sentences would have started with a word with adjacent repetitions, while all ABB sentences would have started with a word with nonadjacent repetitions. In contrast, using both *abb* and *aba* tokens as repeated lower-level structures in the familiarization streams allowed us to have *abb* and *aba* tokens with the same frequency in each possible position.

The experiment was controlled by PsyScope X (http://psy.ck.sissa.it/). Infants' looking behavior was recorded by a camera hidden behind the central monitor.

Materials

Stimuli were synthesized using the soft de7 voice of MBROLA (Dutoit, 1997). Phoneme duration was 200 ms (syllable duration 400 ms) with a fundamental frequency of 200 Hz. Syllables within words were separated by 100 ms silences, words within sentences by 300 ms silences, and sentences by 1,200 ms silences. Word duration was 1,400 ms and sentence duration 4,800 ms. Tables 1 and 2 show the familiarization and test sequences, respectively. Sample files can be found at http://www.endress.org/demos/rep_embed/.

Familiarization

Infants listened to a familiarization stream of 4.47 min. For half of the infants, all sentences had the pattern AAB (composed of words with the patterns *aba–aba–abb* and *abb–abb–aba*); for the remaining infants, all sentences had the pattern ABB. The streams contained eight repetitions of six different sentences with the same sentence-level pattern in random order.

While the stream was continuously played from the central loudspeaker, infants watched a blinking light on the central monitor or on one of the side monitors, with no contingency between the lights and the stream. The blinking light always appeared first on the central screen. When infants attended to it, the light was shown on one of the side monitors (randomly selected); when infants looked away from this side monitor for more than 2 sec, the light was displayed again on the central monitor.

Test

All infants were presented with two new AAB and two new ABB streams. For each infant, two of the streams were thus consistent with the familiarization stream, while the other two violated the familiarization pattern. Each stream consisted of repetitions of the same test sentence (see below). These four streams were presented three times in pseudo-random order for a total of twelve trials.

In Experiment 1, the repeated word pattern was always *aba* (in contrast to the familiarization phase, where both *aba* and *abb* patterns could be repeated). That is, test sentences had the form *aba–aba–abb* (for the sentence pattern AAB) or *abb–aba–aba* (for the sentence pattern ABB).

In each test trial, the blinking light first appeared on the central screen. After infants fixated to it, it disappeared and reappeared on one of the side monitors. When infants oriented toward the blinking light, the test item was played repeatedly from the loudspeaker behind the monitor to which they attended. The test items were presented until 30-sec cumulative looking had accumulated or until infants looked away for 2 sec from the monitor to which they attended.[2] When one of these criteria was met, both the sound and the blinking light stopped and a new test trial started. Pairings between test items and sides were counter-balanced for each infant. Infants' looking times toward the side from where the sound was played were monitored and controlled online through Psyscope X and coded off-line for data analysis.[3] Trials with a cumulative looking time of <2.9 sec (corresponding to the offset of the first phoneme of the third syllable in the second word) were excluded from analysis, because the test patterns could not be distinguished before that time.[4]

Results

Figure 2 (left) shows the results of Experiment 1. We analyzed the raw looking times from Experiment 1 using an ANOVA with congruence (consistent test sentence versus inconsistent test sentence) as the within-subject predictor, familiarization pattern as the between-subject predictor as well as with the interaction. We observed no main effects or interactions (all F's < 1). Hence, infants' looking times to sentences inconsistent with the pattern heard during familiarization (M = 7.72 sec, SD = 3.56 sec) did not differ statistically from the looking times to sentences consistent with the familiarization pattern (M = 7.06 sec, SD=3.28 sec), $W = 57$, $p = 0.807$, ns. Eight of 14 infants looked longer to inconsistent sentences

[2]In principle, we implemented a third termination criterion, namely reaching the maximum test duration of 58 sec. In practice, however, no infant met this criterion.

[3]By coding the looking times offline, we made sure that trials were terminated when infants looked away for 2 sec or after 30 sec of cumulative looking had accumulated. While trials that are terminated prematurely (due to erroneous online coding) are usually excluded, no such case happened in the present dataset.

[4]This was the case for 33% of trials in Experiment 1 and 25% in Experiment 2. The criterion for including an infant into the study was to have a minimum of six total valid trials and a minimum of two trials of each type.

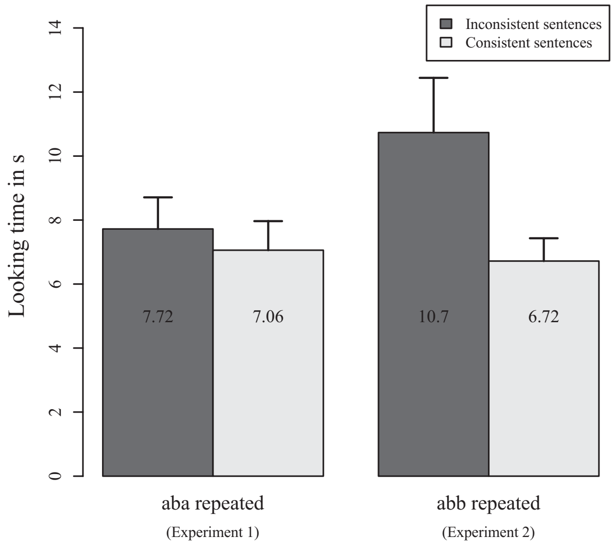

Figure 2 Results of Experiments 1 and 2. Bars show the sample averages and error bars the sample standard errors. Infants learned a hierarchical structure of repetitions when the repeated word structure during test was *abb* (Experiment 2) but not when it was *aba* (Experiment 1).

than to consistent sentences, $p(N = 14) = 0.388$ (binomial test; all statistical tests are two-tailed).

Discussion

The results of Experiment 1 did not reveal any discrimination between novel sentences consistent with the hierarchical pattern and sentences inconsistent with it. However, as mentioned above, the lower-level patterns involved nonadjacent repetitions (i.e., *aba*), and past research has shown that word-level patterns with adjacent repetitions (i.e., *abb*) are easier to process than word-level patterns with nonadjacent repetitions (i.e., *aba*; Kovács & Mehler, 2009). Hence, it is possible that infants might be more likely to show evidence for hierarchical processing with the easier lower-level patterns involving adjacent repetitions. We addressed this question in Experiment 2.

While in the test phase of Experiment 1, the repeated word-level pattern was *aba*, during the test phase of Experiment 2, in contrast, the repeated word-level pattern was *abb*.

EXPERIMENT 2

In Experiment 2, we asked whether infants would show evidence for hierarchical processing with lower-level patterns that are easier to process than in Experiment 1. Experiment 2 was identical to Experiment 1, except that the repeated word pattern *during test* was always *abb* (while the repeated word pattern during familiarization could be both *abb* and *aba*). That is, the familiarization phase was identical to that of Experiment 1. In contrast, test sentences with the pattern AAB had the form *abb–abb–aba*, and test sentences with the pattern ABB had the form *aba–abb–abb* (Table 2). If infants find it more difficult to extract the pattern *aba*, they might find it easier to generalize the hierarchical patterns in the test of phase Experiment 2, where the repeated word-level pattern was *abb*.

Materials and method

Participants

Fourteen healthy full-term 7-month-olds (mean age: 7.19, range 7.05–7.29, seven females), randomly assigned to one of two familiarization streams (ABB or AAB), completed the study. An additional seven infants did not complete the experiment due to fussiness.

Familiarization

The familiarization phase was identical to that in Experiment 1.

Test

The test phase was similar to that in Experiment 1, with one crucial exception. While in the test phase of Experiment 1, the repeated word-level pattern was *aba*, during the test phase of Experiment 2, the repeated word-level pattern was *abb*.

Results and discussion

Figure 2 (right) shows the results of Experiment 2. The crucial dependent variable is the difference between looking times to test sentences inconsistent with the familiarization pattern and looking times to test sentences consistent with that pattern. We analyzed the data by submitting them to an ANOVA with familiarization pattern as a between-subject predictor and congruence (consistent test sentence versus inconsistent test sentence) as a within-subject predictor as well as with their interaction. This analysis yielded a main effect of congruence, $F(1,12) = 5.05$, $p = 0.044$, $\eta_p^2 = 0.292$. Neither the main effect of familiarization pattern, $F(1,12) = 2.73$, $p = 0.124$, $\eta_p^2 = 0.185$, nor the interaction, $F(1,12) = 0.23$, $p = 0.643$, $\eta_p^2 = 0.0131$, reached significance.[5] Hence, although all sentences during test were new (and involved new syllables), infants looked significantly longer to sentences inconsistent with the familiarization pattern ($M = 10.7$ sec, $SD = 6.16$ sec) than to sentences consistent with it ($M = 6.72$ sec, $SD = 2.57$ sec). Furthermore, 12 of 14 infants looked longer to inconsistent sentences than to consistent sentences, $p(N = 14) = 0.013$ (binomial).

We compared Experiments 1 and 2 in an ANOVA, with the familiarization sentence pattern (AAB or ABB) and the repeated word pattern during test (*aba* or *abb*; that is, the difference between Experiments 1 and 2) as between-subject factors and congruence (consistent or inconsistent sentence) as within-subject factor. We observed a marginal main effect of congruence, $F(1,24) = 4.25$, $p = 0.0502$, $\eta_p^2 = 0.138$, but no other main effects or interactions.

When analyzing Experiments 1 and 2 jointly, infants showed overall discrimination between consistent and inconsistent sentences. However, the results of Experiment 1 alone do not show a differential looking pattern toward consistent and inconsistent sentences when the hierarchical patterns are implemented using nonadjacent word-level repetitions during the test phase. These results contrast with those of Experiment 2, where infants successfully detected hierarchical patterns when the patterns were implemented using adjacent word-level repetitions during the test phase.

This finding is in line with earlier studies suggest that patterns involving repetitions between nonadjacent syllables (as in *aba*) are harder to process than patterns with repetitions between adjacent items (as in *abb*; Kovács & Mehler, 2009). However, as the results of Experiments 1 and 2 did not differ significantly, and as we observed a main effect of congruence, we cannot reject the hypothesis that infants might have a residual ability to learn hierarchical patterns with nonadjacent word-level repetitions as well.

[5]While our crucial result—the main effect of congruence—is significant at an α level of .05, the relatively high p value is due to one participants who looked to the test items with the novel structure for the maximum duration of 30 sec. While this participant also looked longer to the inconsistent structure than to the consistent one, her longer looking time deviated by 4.2 standard deviations from the overall average looking time. Removing this participant from the analysis yielded a main effect of congruence, $F(1,11) = 17.35$, $p = 0.0016$, $\eta_p^2 = 0.567$, while neither the main effect of familiarization structure, $F(1,11) = 1.50$, $p = 0.247$, $\eta_p^2 = 0.12$, nor the interaction, $F(1,11) = 2.25$, $p = 0.162$, $\eta_p^2 = 0.074$, reached significance.

GENERAL DISCUSSION

Taken together, the results of the two experiments suggest that infants successfully generalized the hierarchical patterns, although the results of Experiment 2 (where the repeated word pattern during test was *abb*) were numerically stronger than in Experiment 1 (where the repeated word pattern during test was *aba*). This might be related to the fact that it is much harder to learn patterns involving repetitions of nonadjacent syllables than patterns involving repetitions of adjacent syllables. At minimum, however, the results of the present study suggest that, once infants can detect the word pattern, they seem to be able to also embed it into a hierarchical sentence pattern.

How did infants process the repetition pattern? There are at least three related possibilities that all seem to involve hierarchical processing. First, after infants detect the repeated syllables at the word level, they might notice that two word patterns are always identical, either at the beginning of a sentence or at the end.

Second, infants might notice that sentences start or end with exactly *two* words with the same pattern, while the third word has a different pattern. The difference between the first possibility and this alternative is mainly one of terminology: either infants detect that, in the sentence pattern AAB, the first two word patterns are identical, and, therefore, a hierarchical repetition pattern. Or they count the number of occurrences of consecutive identical patterns in a specific position. However, to count the number of occurrences of identical patterns (on the sentence level), infants have to notice that two patterns are identical in the first place (on the word level), which would be, in our view, just the hierarchical repetition pattern described above.

Third, infants might focus on words with adjacent syllable repetitions.[6] For example, if + stands for a word with an adjacent repetition and − for a word without, AAB familiarization sentence had either the form + + − or − − +. If infants focus on words with adjacent syllable repetitions but do not learn well from words with nonadjacent syllable repetitions, they should generalize in Experiment 2, because legal AAB test sentences have the pattern + + − and legal ABB have the pattern − + +. In contrast, they should perform less well in Experiment 1, because they would need to discriminate between the sentence patterns − − + and + − −.[7] Crucially, this explanation

[6]We are grateful to an anonymous reviewer for pointing out this possibility.

[7]In contrast, it is not sufficient for infants to monitor whether sentences start or end with "more" words with adjacent syllable repetitions. Specifically, when the repeated word pattern comprises adjacent repetitions, AAB sentences have the form + + − and ABB sentence − + +; however, when the repeated word pattern has nonadjacent repetitions, AAB sentences have the form − − +, and ABB sentences + − −. Hence, if infants learned that AAB sentences have more initial adjacent syllable repetitions, they should significantly prefer illegal sentences in Experiment 1.

also involves hierarchical processing just like the second one, but takes into account that repetition patterns with adjacent repetitions are easier to process than repetition patterns with nonadjacent repetitions.

Irrespective of which of these interpretations is correct, it is also possible that infants might generalize the structural relations they acquired to a broader set of patterns. For instance, after a familiarization as in Experiment 1, infants might discriminate *aab*-*aab*-*abb* items from *abb*-*aab*-*aab* items, although they have not between familiarized with *aab* as a word-level pattern. If so, they might generalize the familiar sentence-level pattern to novel lower-level patterns. However, we can exclude the possibility that infants did not encode the word-level patterns during familiarization; after all, if they had not encoded the word-level patterns in the first place, they could not have noticed that two word-level patterns were identical.

Our results leave open which exact sentence-level patterns infants learned: a more specific one (e.g., ABB, where the last two patterns are identical and preceded by a different one) or a more inclusive one (e.g., XBB, with $X \rightarrow \{\alpha, X\alpha\}$ where α is a syllable sequence, and X a symbol for the kind of constituent that can occur sentence initially; in other words, the last two patterns are identical and can be preceded by any number of words conforming to any pattern). In either case, the data show that infants embedded the word-level patterns into a higher-level pattern, be it another repetition pattern, a pattern based on the cardinality of patterns (which, as mentioned above, requires detecting the identity of two patterns as well), or another related pattern. Hence, the current results suggest that infants seem to deploy hierarchical computations when learning repetition-based patterns.

Importantly, although hierarchical processing is crucial for language, the (repetition-based) patterns we employed are not specific to language and can be processed outside the language faculty, for example with visual, auditory, or olfactory stimuli by human infants and nonlinguistic animals (Dawson & Gerken, 2009; Giurfa et al., 2001; Hauser & Glynn, 2009; Murphy et al., 2008; Saffran et al., 2007); moreover, syntax-like regularities based on repetitions cannot even be learned by human adults (Endress & Hauser, 2009).

Given that infants learned a hierarchically embedded pattern based on repetitions nonetheless, our results raise the question of whether such hierarchically embedded patterns are unique to the language faculty, or whether they might be available in other cognitive domains as well.

To understand what is truly unique about hierarchical processing in language, one thus needs to investigate the extent and limits of hierarchical processing not only in language, but also in other domains and in nonhuman animals. A systematic comparison with the kinds of hierarchical patterns found in language might reveal what aspects of hierarchical processing are specific to language, and how those aspects have evolved.

ACKNOWLEDGMENTS

This research was partly supported by the European Research Council under the European Union's Seventh Framework Programme (*FP7/2007-2013*)/ERC starting grant to AMK (n° 284236-REPCOLLAB), and a Marie Curie Incoming Fellowship (303163-COMINTENT) and a Spanish Ministerio de Economía y Competitividad grant (PSI2012-3253) to ADE. We thank participants and their parents for participation, and J. Mehler and E. Téglás for helpful comments on a previous version of this manuscript.

REFERENCES

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*, 321–324.

Brooks, L. R., & Vokey, J. R. (1989). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General*, *120*, 316–323.

Byrne, R. W., & Russon, A. E. (1998). Learning by imitation: A hierarchical approach. *Behavioral and Brain Sciences*, *21*, 667–709.

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

Dawson, C., & Gerken, L. (2009). From domain-generality to domain-sensitivity: 4-month-olds learn an abstract repetition rule in music that 7-month-olds do not. *Cognition*, *111*, 378–382.

Dutoit, T. (1997). *An introduction to text-to-speech synthesis*. Dordrecht: Kluwer.

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*, 247–299.

Endress, A. D., Dehaene-Lambertz, G., & Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition*, *105*, 577–614.

Endress, A. D., & Hauser, M. D. (2009). Syntax-induced pattern deafness. *Proceedings of the National Academy of Sciences of the USA*, *106*, 21001–21006.

Endress, A. D., & Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology*, *61*, 177–199.

Endress, A. D., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language*, *60*, 351–367.

Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences*, *13*, 348–353.

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

Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences of the USA*, *105*, 14222–14227.

Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. V. (2001). The concepts of 'sameness' and 'difference' in an insect. *Nature*, *410*, 930–933.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436.

Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109–135.

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

Gómez, R. L., Gerken, L., & Schvaneveldt, R. (2000). The basis of transfer in artificial grammar learning. *Memory & Cognition*, *28*, 253–263.

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.

Hauser, M. D., & Glynn, D. (2009). Can free-ranging rhesus monkeys (*Macaca mulatta*) extract artificially created rules comprised of natural vocalizations? *Journal of Comparative Psychology*, *123*, 161–167.

van Heijningen, C. A. A., de Visser, J., Zuidema, W., & ten Cate, C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Sciences of the USA*, *106*, 20538–20543.

Hochmann, J.-R., Azadpour, M., & Mehler, J. (2008). Do humans really learn AnBn artificial grammars from exemplars? *Cognitive Science*, *32*, 1021–1036.

Jackendoff, R. (2007). *Language, consciousness, culture: Essays on mental structure (Jean Nicod Lectures)*. Cambridge, MA: MIT Press.

Kemler-Nelson, D., Jusczyk, P., Mandel, D., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*, *18*, 111–116.

Kovács, A. M., & Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. *Science*, *325*, 611–612.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley.

Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science*, *18*, 387–391.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. (1999). Rule learning by seven-month-old infants. *Science*, *283*, 77–80.

Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.

Murphy, R. A., Mondragon, E., & Murphy, V. A. (2008). Rule learning by rats. *Science*, *319*, 1849–1851.

Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. Davidson, G. Schwartz, & D. Shapiro (Eds.), *Consciousness and self regulation: Advances in research and theory* (Vol. *4*, pp. 1–18). New York, NY: Plenum Press.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*, 604–607.

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

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, *95*, 201–236.

Poggio, T., & Bizzi, E. (2004). Generalization in vision and motor control. *Nature*, *431*, 768–774.

Rosenberg, R. D., & Feigenson, L. (2013). Infants hierarchically organize memory representations. *Developmental Science*, *16*, 610–621.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996b). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996a). Word segmentation: The role of distributional cues. *Journal of Memory Language*, *35*, 606–621.

Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: infant rule learning is not specific to language. *Cognition*, *105*, 669–680.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multi-level statistical learning by 12-month-old infants. *Infancy*, *4*, 273–284.

Tunney, R., & Altmann, G. T. (2001). Two modes of transfer in artificial grammar learning. *Journal Experimental Psychology: Learning, Memory & Cognition*, *27*, 614–639.

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.

Yang, C. D. (2004). Universal Grammar, statistics or both? *Trends in Cognitive Sciences*, *8*, 451–456.