

Gender differences in non-adjacent dependency learning

Investigating the small female advantage in language
development

Lianne van Setten

30/06/2014

30-06-2014

Utrecht University

Student nr: 3698637

F. N. K. Wijnen & I. C. Grama

KI3V12012 - 15 ECTS

Gender differences in non-adjacent dependency learning: investigating the small female advantage in language development

Lianne van Setten

Abstract

Theories about the advantage for girls in language development are well known, but not always sufficiently backed up by empirical evidence. The purpose of this study is to determine whether there is a difference between the learning and language abilities of boys and girls at a young age, focusing on pattern learning and the generalization of these word patterns. This study investigated whether 18-month-old infants are able to generalize non-adjacent dependency patterns to novel contexts, and whether we can identify gender differences in the performance on this task. The looking behavior of 21 infants was measured in an Artificial Grammar Learning experiment using the Headdturn Preference Procedure. The Dutch version of the MacArthur-Bates Communicative Development Inventory (N-CDI) was used to assess children's productive and receptive vocabulary. In contrast to the predictions there were no significant differences between the N-CDI scores of boys and girls. In the Artificial Grammar Learning experiment, infants were overall not able to distinguish correct from incorrect dependencies, but we did find a significant effect of gender on the ability to discriminate correct from incorrect dependencies when looking only at the first four trials. Although the N-CDI data does not support a gender distinction, the behavioral data suggest an advantage for girls in distributional learning, which collaborates with the theory that girls have an overall advantage in language development.

Introduction

Although interest in differences between male and female humans has existed for centuries, in the last few years there has been a proliferation in scientific studies investigating gender differences. Questionnaires and other standard research methods are still in use, but there are now also new research techniques that make it possible to investigate subtle differences between the male and female brain, cognition and behavior. An example is the use of functional Magnetic Resonance Imaging (fMRI), which led to an increasing amount of studies investigating these gender differences in different tasks at different ages (Kaiser, Haller, Schmitz & Nitsch, 2009). The language development of boys and girls is an interesting subject and generally known to be different between the two sexes. However, there are a lot of contradictions in the outcomes of several studies that investigated gender differences in language development. The ages of the participants and varied methods used in these studies, made it difficult to compare the results (Andreano & Cahill, 2009) and to make clear statements about the existence of gender differences in language acquisition. This study focuses on one specific skill that might play a role in grammar learning, which is non-adjacent dependency learning. The purpose of this study is to investigate whether there are differences between the performance of 18-month-old boys and girls in the ability to learn non-adjacent dependencies and generalize them to novel contexts.

One of the studies that provided evidence for the existence of gender differences in language development, was done by Kramer, Kaplan, Delis, O'Donnell and Prifitera (1997). This study used questionnaires and a word-recall task with 811 children between the ages of five and sixteen years old that were tested. The Wechsler Preschool and Primary Scale of Intelligence was the questionnaire used for 5-year-old children and the Wechsler Intelligence Scale for Children-Revised was used for children aged six to sixteen. They were all administered the California Verbal Learning Test- Children's Version, which contains various trials of 15-word lists with words that were semantically categorized and which they had to recall in various ways and orders afterwards. The results of this large study showed that girls performed better than boys on all of the immediate and

delayed word-recall trials and on the delayed word-recognition trials. Results showed that girls clustered these words in semantic categories (clustering strategy), which was a more effective long-term memory mechanisms, which boys didn't use as often as girls. Boys however, scored higher on the Wechsler Intelligence Scale for Children, so the observed female superiority in verbal learning could not be attributed to sex differences in overall intelligence (Kramer et al., 1997).

There is a difference between language development of boys and girls at younger ages as well; Galsworthy, Dionne, Dale and Plomin (2000) tested 2-year-old twin pairs using the MacArthur-Bates Communicative Development Inventory and the Parental Report of Children's Ability. The results showed significantly higher scores for girls on verbal measures like vocabulary production and comprehension, and non-verbal measures like mimicking, block-building and specific behaviors.

Questionnaires like these make it possible to compare language development at a very young age. The MacArthur-Bates Communicative Development Inventory (CDI) is a parental report about the general language skills of infants, measuring the understanding and production of several hundreds of words (Zink & Lejaegere, 2002). Bouchard, Trudeau, Sutton, Boudreault and Denault, (2009) found gender differences after testing 8-to-30-month-old infants using the French version of the CDI and analyzed samples of spontaneous speech of each infant. The results confirmed the classical idea of a linguistic advantage of girls over boys; the amount of words girls produced was significantly higher and their utterances were more complex and contained more different grammatical constructions than the boys' utterances. At the age of 18 months, the mean number of different words (verbs, nouns, etc.) produced by boys was significantly lower than the number of words girls produced, while the syntactic complexity did not differ at this age (Bouchard et al., 2009). Berglund, Eriksson and Westerlund (2005) tested 18-month-old-infants as well, using the Swedish version of the CDI. These results confirm the higher productive vocabulary scores of girls over boys at this age (Berglund et al., 2005).

Complementary on these studies (Berglund et al., 2005; Bouchard et al., 2009), it is likely that that results of a Dutch version of the MacArthur-Bates Communicative Development Inventory will show that girls produce and comprehend a larger amount of words than boys do at this age, but do not show an advantage in syntactic sophistication of their productions compared to 18-month-old boys.

Gender differences in language development can also be supported by neuroimaging studies that investigated differences between male and female brain structures. Andreano and Cahill (2009) concentrated on both neurological and hormonal differences and found that the brain areas that are activated during several language-specific tasks differ between adult males and females. An example is a task at which adult males and females had to learn words of an unknown language. The activation pattern of the fMRI showed more bilateral activation of the fusiform gyrus for the females, while man showed more left lateralized activation in the fusiform gyrus (Andreano & Cahill, 2009). This might suggest that the word recognition process differed between these two groups. Andreano and Cahill (2009) suggested that the outstanding performance of females on verbal tasks extends to numerous other non-verbal tasks, including tests of spatial and autobiographical abilities and a small significant advantage for episodic memory. Differences in the activated neurons during the processing of phonological information were found in a neuroimaging study of Clements et al. (2006). Their results provided evidence for a difference between males and females in the processing of visuospatial information as well; the neural activation in males seemed to be more left

lateralized during phonological tasks, while females showed bilateral activity during these tasks (Clements et al., 2006).

These gender differences in adult brain structures do not necessarily mean that these differences also exist in children and newborns. Children between the ages of nine and fifteen were tested by Burman, Bitan and Booth (2008) using fMRI. Their brain activity was measured while performing two tasks: an orthographic judgment task, in which the children needed to judge whether two words contained the same letters, and a phonological judgment task, in which they needed to judge whether two words rhymed. The results of these tasks showed that girls had more activation in several brain areas than boys, even at this young age. During all tasks girls showed greater activation in the inferior frontal gyrus and left fusiform gyrus, and they showed bilaterally greater activation in the superior temporal gyrus during auditory tasks. The accuracy scores in these tasks did not differ between boys and girls, although the overall reaction time of the girls was lower (Burman et al., 2008). This might suggest a difference between boys and girls in the way they process orthographic and phonological information.

The gender differences found in the behavioral and neurological studies cited above do not necessarily indicate genetic differences between the sexes that evolve in observable behavioral differences between males and females. Kaiser et al. (2009) state that gender differences might develop in reaction to social influences that differentiate males and females in society. This hypothesis is supported by Sung, Fauto-Serling, Coll and Seifer (2013), who investigated the communication between mothers and their newborns and how this affected the language development of their child. Results of this study showed that vocal communication differed for the mother-daughter and mother-son interactions at the ages of 6-8 and 9-11 months. The mother-daughter interactions showed a mutual increase in direct vocal responses, comparing the recordings from 3-5 to 9-11 months and from 6-8 to 9-11 months (Sung et al., 2013). Although social influence might cause an increasing differentiation between the sexes, neurological differences can also cause these variations in the behavior of males and females (Kaiser et al., 2009), which makes it difficult to explain these differences and attribute them to socialization or genetic factors.

Although many studies provided evidence for the differences between the language development of young boys and girls, results of neuroimaging studies might suggest an underlying cause. The differences in activated brain areas during linguistic tasks can indicate that boys and girls employ different learning mechanisms when learning complex words and sentences.

Reanalyzed data of Marcus et al. (1992) by Hartshorne and Ullman (2006) showed how many times 10 boys and 15 girls of on to four years over-regularized verbs in recorded spontaneous speech. The recordings of these children were not statistically different in length, amount of words, verbs or sentences and there were no overall differences between the boys and girls in age, social class or linguistic background. According to their data girls over-regularized three times more than boys did, and they tended to retrieve language by remembering words in a holistic way, while boys compose complex words from remembered stored morphological units. Their hypothesis is that girls remember regular forms better and use these forms to generalize them to similar-sounding verbs while boys use rules to remember verbs. They suggest there is a difference in the learning mechanism that causes the differences between boys and girls in language production (Hartshorne & Ullman, 2006).

This differences in the mechanisms boys and girls use to produce and learn complex words, is a hypothesis supported by reviewed data of Wallein (2009). Among other things he compared two studies (Kramer et al., 1997; Kramer, Delis, and Daniel, 1988) that used the California Verbal Learning Test (CVLT) to investigate differences between the language development of males and females or boys and girls. The study of Kramer et al. (1988) showed that females used a clustering strategy to organize words during the task. This method was also found in 5-to-16-year-old girls in the mentioned research of Kramer et al. (1997). Wallein (2009) stated that the difference between language development of boys and girls existed due to a difference in learning mechanisms; girls were more likely to use a semantic clustering strategy at memory tasks. This learning mechanism extends beyond the language domain, since the differences between boys and girls are not only found in verbal information, but also in the way they categorize and organize non-verbal information.

As stated before, results are inconsistent and although Wallein (2009) found significant differences between the receptive and productive vocabulary scores of boys and girls at a very young age, he also found that these differences gradually decreased with increasing age, which contradicts some of the previous claims (Andreano & Cahill, 2009; Burman et al., 2008; Kramer et al., 1997; Wallein, 2009). Bornstein, Hahn and Haynes (2004) investigated the language development of 329 infants using a variety of questionnaires, reports and transcripts. They only found a significant difference in the language development of boys and girls at the age of five, but no differences whatsoever before and after that age. This suggests that not all studies necessarily found evidence for the existence of gender differences in learning and language development. It is important to explore the different methodologies and populations employed in these studies, but it is even more important to note that studies failing to find gender differences are hard to find due to the publication bias: significant gender differences are interesting to publish, whereas similarities and non-significant differences are not. In fact, the only published studies that revealed no differences between the sexes, are studies that did not specifically focus on these differences, but included gender only as a secondary factor in their analysis (Kaiser et al., 2009).

Although a lot of studies are performed with children and adults, young infants are less influenced by socialization than these groups are. This makes it interesting to investigate whether there are gender differences in the performance of infants at linguistic tasks. The linguistic abilities of infants can be tested using tasks that do not only test the prior knowledge of the infants about words and syntax of their own language, but also their ability to adjust to unknown words and patterns in word streams. Non-adjacent dependency learning is used to test infants on their ability to learn patterns in a stream of unknown words; sensitivity to non-adjacent dependencies relies on a learning mechanism that might aid with the acquisition of natural as well as artificial languages (Gómez, 2002), since such dependencies abound in natural languages, and mark important syntactic relationships. An example of a non-adjacent dependency in natural language is a verb and infinite verb that always elaborate in the same pattern (1).

1. a) *The Woman is always laughing.*
b) *Ik heb het eten gekookt. (English: I have cooked dinner)*

The ability to learn non-adjacent dependencies can be tested with young children using the Headturn Preference Procedure (Kemler-Nelson et al., 1995). This procedure consists of two phases and relies on orientation to visual stimuli to assess attention to an auditory stimulus. During the familiarization phase, the infant is exposed to an artificial language, with strings of nonsense words, which exhibited certain structural regularities. After the child is familiarized with this artificial language, the test phase begins. During this phase several trials are presented, containing each either strings that are consistent with the regularities of the language heard at the familiarization phase, or strings that are inconsistent with these regularities (Kemler-Nelson et al., 1995). The Headturn Preference Procedure is an infant preference procedure that is used by researchers to find out whether infants are able to distinguish different conditions or different stimuli. Hunter and Ames (1988) expound that infants can show a preference for familiar or novel stimuli, meaning they show either a familiarity effect, a novelty effect or no effect at all. This preference depends on the age of the child and the difficulty of the task; at first infants show a familiarity effect, and in time this effect will develop into a novelty effect. Infants are expected to show a novelty effect at tasks that are easy for them, and a familiarity effect at difficult tasks, since familiarity effect precedes a novelty effect (Hunter & Ames, 1988).

The Headturn Preference Procedure was used by Gómez (2002) to investigate whether adults and 18-month-old infants were able to learn non-adjacent dependencies out of three-element strings of nonsense words. The infants were exposed to strings of three nonsense words that took the form aXc and bXd (Language 1) or aXd and bXc (Language 2). The X-elements were different bisyllabic nonsense words and the infants were exposed to variable conditions that contained 3, 12 or 24 different X-elements that were combined with the dependencies to form word strings. The dependencies were four different monosyllabic nonsense words. The word strings that were used in the familiarization phase were all consistent to one of the languages. The word strings that were played during the test phase, varied between word strings that were consistent with the ones of the familiarization phase, and word strings that were inconsistent with the learned grammar. The results of this procedure presented that the infants showed no effect at the conditions of 3 and 12 different X-elements. Results of the last condition differed: fifteen out of the sixteen 18-month-old infants showed a novelty effect when exposed to a set-size of 24 X-elements, and there was a significant interaction between grammaticality and the set-size of the X-elements.

Gómez and Maye (2005) also used this method to investigate whether 12-month-old, 15-month-old and 17/18-month-old infants were able to distinguish learned non-adjacent dependencies from novel ones, with set sizes of 12, 18 or 24 recorded X-elements. Results showed that 12-month-old infants were not able to learn these patterns, while the 15-month-old infants showed a commencing ability to distinguish between the two. This group of infants listened longer to the pattern that they had heard in the familiarization phase, as opposed to the 17/18-month-old infants, who listened longer to the novel pattern. This confirmed the predictions of Hunter and Ames (1988) that familiarity effect precede novelty effects.

Kerkhoff, De Bree, De Klerck and Wijnen (2013) also tested 18-month-old infants on non-adjacent dependency learning, but compared infants at familial risk for dyslexia with typically developing controls. Following Gómez and Maye's (2005) design closely, they also used languages that combined two a_c and b_d dependencies with 24 different X-elements, into a total of 48 aXb strings. While at-risk infants did not show a significant preference for either novel or familiar items

at test, the control group showed a significant novelty effect, which proved that they were able to distinguish learned patterns from novel ones.

A slightly different method was used by Lany and Gómez (2008) to investigate whether 12-month-old infants were able to generalize experience with adjacent dependencies to non-adjacent dependencies. The method used was the Habituation-Dishabituation Paradigm, with an additional habituation phase between the familiarization phase and the test phase. During familiarization infants were exposed to adjacent dependencies between classes of words (e.g. aX, bY), whereas during habituation phase the infants were exposed to the same type of dependencies, but with an intervening word that rendered them non-adjacent (e.g. acX, bcY). Results showed that 12-month-old infants were able to generalize adjacent dependencies to more difficult non-adjacent dependencies. Females however, showed greater discrimination than males, suggesting a superior ability to generalize the rules they learned, perhaps due to an advantage in associative memory. Generalization to novel contexts is an important part of learning patterns in language. The female advantage on language-related tasks is consistent with most research discussed previously that suggested a female advantage in language development and verbal skills (Kramer et al., 1997; Berglund et al., 2005; Bouchard et al., 2009), but also a difference in verbal learning mechanisms which could also be one of the reasons for the gender differences found in data of Lany and Gómez (Hartshorne & Ullman, 2006; Lany & Gómez, 2008; Wallein, 2009).

The study presented here was performed to extend the findings of Lany and Gómez (2008) and Kerkhoff et al. (2013). The procedure Lany and Gómez used differed from the procedure Kerkhoff et al. used, although both studies are based on non-adjacent dependency learning. Kerkhoff et al. did not report any gender differences, and their task did not require the infants to generalize their knowledge of dependencies to novel contexts, but merely recognize them in already familiar context, unlike the study of Lany and Gómez.

In this study we wanted to investigate (i) whether 18-month-old infants are able to not only learn but also generalize non-adjacent dependencies to novel contexts, and (ii) whether there is a gender difference at the performance of this task. Answers to these questions will provide more insight in the linguistic abilities of 18-month-old infants and the differences between boys and girls in their ability to detect and generalize patterns. If gender differences were evidenced in this learning task, this would confirm the hypothesis that boys and girls employ different learning mechanisms in dealing with unfamiliar input. Gender differences in receptive and productive vocabulary will also be investigated in order to get a more complete overview of the variation in linguistic skills of 18-month-old boys and girls.

This exact age (18 months) is interesting because of the previous findings using different versions of the MacArthur-Bates Communicative Development Inventory suggested a significant advantage in productive vocabulary for girls at this age (Bouchard et al., 2009; Berglund et al., 2005); we wish to extend these findings to Dutch infants, and identify a potential correlation between size of (productive) vocabulary and pattern-learning abilities in both sexes.

As stated before, results of Lany and Gómez (2008) showed a gender difference between the pattern-learning abilities and ability to generalize this pattern to novel items in artificial language at an age of 12 months. Both boys and girls are able to distinguish familiar from novel dependencies in an artificial language at the age of 18 months (Gómez & Maye, 2005; Gómez, 2002; Kerkhoff et al., 2013), but their ability to generalize these patterns to new strings has not been tested so far. The

present study bridges this gap, focusing on the difference between the sexes at the age of 18 months. The previous mentioned findings (Gómez & Maye, 2005; Gómez, 2002; Kerkhoff et al., 2013; Lany and Gómez, 2008) suggest that we can expect that 18-month-old infants are able to learn non-adjacent dependencies and favor novel patterns, and that girls show better performance at non-adjacent dependency learning, when this learning entails generalizing non-adjacent patterns to novel contexts than the ones heard during familiarization.

Method

Participants

A total of 33 infants participated in this experiment. The data of 12 of these infants was excluded due to excessive crying, lack of attention or fuzziness ($n = 6$), not completing enough test trials ($n = 2$), failing to meet the selection criteria ($n = 1$) or insufficient looking times per condition ($n = 3$). The infants of the included data were 10 males and 11 females with an average age of 18;22 months (range = 18;10 - 18;29). These infants had normal weight at birth, were not post- or premature and were reported by their caregiver to have normal hearing, normal vision and no neurological problems or risk for dyslexia.

The infants were recruited by written requests to the parents of newborns in the municipality of Utrecht. The caregivers were asked to fill in background information into a standard questionnaire developed by the BabyLab of Utrecht University. This background information supplied knowledge about the general health, education levels and history of the families and also provided information about the overall development and the motor skills of the infants. The caregivers also filled in the Dutch version of the MacArthur-Bates Communicative Development Inventory (N-CDI) about the general language skills of their infants (Zink & Lejaegere, 2002).

Stimuli

The stimuli were divided into two artificial languages. The languages both contained the same nonsense words, combined in three-word phrases. All nonsense words were taken from Kerkhoff et al. (2013, based on Gómez, 2002) and were phonotactically legal in Dutch. The exact combinations of these nonsense words differed between the two languages: while Language 1(L1) consisted of phrases with the combinations aXc and bXd, Language 2(L2) consisted of the combinations aXd and

Table 1.

All stimuli used is this experiment

X's familiarization phase:			X's test phase:	L1:	L2:
Banip	Kengel	Rajee	Klepin	tep X lut	tep X jik
Densim	Loga	Rogges	Lotup	sot X jik	sot X lut
Dieta	Movig	Seeta	Tarzin		
Domo	Naspu	Snigger			
Fidang	Nilbo	Sulep			
Gopem	Noeba	Vami			
Hiftam	Plizet	Wadim			
Kasi	Poemer	Wiffel			

bXc (table 1). The dependencies in L1 were inconsistent with L2, and vice-versa. The intervening X's were bisyllabic words with a strong-weak pattern and the words forming the dependencies (e.g. a_c, b_d / a_d, b_c) were monosyllabic. The dependencies were combined with the intervening X's; during the familiarization phase twenty-four X's were combined with the dependencies for a total of 48 strings, whereas in the test phase three novel X's were employed (table 1), to ensure that infants did not just remember aXb strings from the familiarization, but were required to generalize the nonadjacent dependencies to novel contexts.

All words were recorded by a Dutch female native speaker in a child-friendly voice. The X-items were recorded as a direct object noun in a Dutch carrier sentence, while the a- and b-elements were recorded emphasizing the monosyllabic dependent elements. The stimuli were spliced from the recordings and concatenated into aXb strings, with 250 milliseconds pauses between the words in a string, and 750 milliseconds pauses between strings. The twenty-seven recorded X's were alike in pitch and duration. The dependent elements were also matched in pitch and duration. Table 2 shows some examples of word strings.

Table 2.

Examples of strings for the two languages

<u>Language 1</u>	<u>Language 2</u>
tep wadim lut	tep wadim jik
sot wadim jik	sot wadim lut
tep snigger lut	tep snigger jik
sot snigger jik	sot snigger lut
tep kengel lut	tep kengel jik
sot kengel jik	sot kengel lut

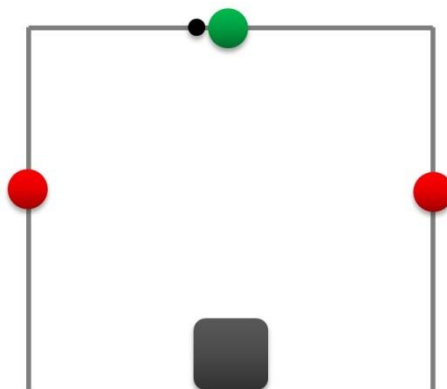


Figure 1. Schematic of the test cabin with the red and green lights. The caregiver and infant are situated at the square and they are recorded by the camera above the green light.

Procedure

The Head-turn Preference Procedure is a commonly used method to test perceptual preferences of infants (Kemler-Nelson et al., 1995) and was used in this experiment. The infant was placed on the

caregiver's lap in a test cabin equipped with one green light in front of the infant and two red ones on the infant's left and right. The lights consisted of three circles LED lights that sequentially lit up. A camera was situated in front of the infant, while two speakers were situated behind the two red lights (figure 1).

In total the experiment lasted about eight minutes. The caregivers were wearing headphones throughout, to prevent them from unintentionally influencing their infant. All caregivers were instructed not to talk, point or move during the experiment. All participants were participating voluntarily and knew they were allowed to stop at any time during the experiment and they all gave explicit written permission that allowed us to use the data and recordings of their visit in research and publications. The looking behavior of the infants was recorded by the camera and followed by the experimenter outside the cabin, who used a button box to signal the child's orientation towards or away from the blinking lights.

All trials started with the green light blinking in front of the infant to get their attention. When the infant was looking at the green light, this light was switched off while simultaneously one of the red lights to the side lit up. The sound started as soon as the infant oriented towards the red light. When the infant looked away for two seconds or more, the trial ended and the green light was switched on again to reorient the child's gaze to the front before the start of a new trial. The left and right lights were used in a pseudo-randomized order (per participant), with a maximum of two consecutive uses of the same side. The experiment consisted of three consecutive phases.

Phase 1: Familiarization phase

The first phase was the training phase infants were assigned randomly to L1 or L2 and exposed to the forty-eight strings of the respective language. In this phase twenty-four out of the twenty-seven X's are used, all twice in random order. The sound started when the infant looked at the blinking red light for the first time, and played continuously until the forty-eight strings were completed. In this phase the sound came from both speakers on the sides.

Phase 2: Contingency phase

In this phase the infants were familiarized with the contingency between the blinking lights and the sounds. When the infants looked at the blinking red light, a neutral tone started playing from the speaker on the same side, and the sound stopped when infants looked away for more than two seconds. The contingency phase consisted of two trials, and used the lights on both sides.

Phase 3: Test phase

The test phase contained eight trials of fifteen three-word-strings per trial. The three X's used in this phase, are different from the twenty-four X's that are used during the training phase (table 1). For each infant half of the trials presented word strings consistent with the language they heard during familiarization phase, and the other half of word strings that were inconsistent; there were four trials consistent with language L1 and four trials consistent with language L2.

Each trial started with lighting the green light in front of the cabin, until the child oriented to it, whereupon this light was extinguished and one of the side lights began to blink. When the infant looked at the red blinking light, the sound started playing at the same side until the infant looked away for two seconds. The order and side of the trials containing word strings of L1 and L2 was randomized per participant, with a maximum of two times the same side or language in a row.

After completing the experiment, the data was recoded with PsyCode by a different researcher, who did not know which condition (L1 or L2) the infants were assigned to. If the total looking time of a trial was below two seconds, it was excluded. Trials that were terminated by the experimenter although the looking-away time was less than two seconds were excluded after recoding. If fewer than two valid trials per condition (consistent/inconsistent) were left after recoding the data of an infant, all data for that infant was excluded as well. We analyzed the total looking time in the test phase per trial as a measure of the child's attention to the consistent and inconsistent stimuli. All data is analyzed using IBM SPSS Statistics 20.

Results

The difference between the two sexes was measured by the difference in looking times for the consistent and inconsistent word strings during the eight trials of the test phase. Half of the infants learned L1 as the correct language during the familiarization phase, while the other half of the infants learned L2 as the correct language. The dependent measure was the total amount of time the infants listened to the trials, which were either consistent or inconsistent with the learned language.

We analyzed the data using a Linear Mixed Model in IBM SPSS Statistics 20 to determine the influence of different factors on the total looking times per trial. Factors of interest were condition (familiar or novel stimuli), trial number and gender. We were also interested in possible covariates (total looking time as a measurement for infants' attention in the familiarization phase, and N-CDI scores for productive and receptive vocabulary), as well as two- and three-way-interactions between these factors. The subject number is added as a random factor in the Linear Mixed Model. We added the main effect for each factor and covariate one by one into the baseline model (with only subjects as random fact) and checked whether it significantly improved the model, using a Chi-Square test based on the the -2LogLikelihood difference and the difference in degrees of freedom between the baseline model and the model with the factor introduced. The only significant effects found were the N-CDI scores for both receptive and productive vocabulary, but these factors are highly correlated ($r = .486, n = 121, p < .01$). Because NCDI scores for receptive vocabulary showed a slightly better contribution to the model, we used it as the only factor reflecting the infants' linguistic development. The only interaction effect that was significant, was the interaction between gender and trial number.

Analysis of this Linear Mixed Model revealed the significant effects of receptive vocabulary ($F(1,121) = 4.413, p = .036$) and the interaction between trial and gender on the total looking time per trial ($F(15,121) = 1.953, p = .024$). We did not find an effect of condition on total looking time, or an interaction between gender and condition. Although there is no interaction found between gender and condition on the total looking time, there is a visible difference between the looking times of the male and female infants per condition (Table 3); while boys seems to have listened slightly longer to familiar stimuli, girls listen much longer to novel stimuli.

There is no significant interaction effect between gender, condition and the total looking time, but as stated before, there is a significant correlation between gender and trial on looking times ($F(15,121) = 1.953, p = .024$). When looking at the total looking times per trial per gender (Table 4), we found that girls seemed quite consistent during the eight trials, while the boys looked longer in the first four trials than they did in the last four trials.

Table 3.

Mean looking time and Standard Deviation for each condition per gender.

	Familiar stimuli	SD	Novel stimuli	SD
Male	10963,74	3346,57	9797,94	5633,46
Female	10099,05	4721,33	15092,65	9751,24

Table 4.

Mean looking time and Standard Deviation for each trial per gender.

trial	M LT Male	SD Male	M LT Female	SD Female
1	12328.20	11664.10	11761.11	9144.88
2	10604.11	9468.51	17449.89	13169.19
3	14080.00	9008.60	12010.33	12474.86
4	11838.88	11335.01	9566.40	10970.98
5	8341.25	5611.37	12010.91	13851.07
6	9475.75	5264.94	9671.00	10021.64
7	5644.00	3462.33	21917.25	13800.89
8	9416.22	8662.62	5199.56	3902.72

We decided to look into the first four trials of the test phase separately. Since predominately boys seemed to have lost their attention in the last four trials of the test phase, we expected to see sharper differences and clearer effects in the first four trials, where infants were expected to be more focused at the task. The results of a Linear Mixed Model analysis executed in the same way as before showed again a significant interaction between receptive vocabulary and the total looking time ($F(1,59) = 6.990, p = .010$). In addition to that, and different from our analysis of the entire dataset, there was also a significant interaction effect of the condition and gender on the total looking time ($F(3,59) = 3.057, p = .035$). In these first four trials there was a significant difference between boys and girls in looking behavior for each condition. Looking at the data per condition per gender, there was a very small difference between the looking times of boys in each condition, and a larger difference between the looking times of girls in each condition (Figure 2). Namely, the girls' looking times to the inconsistent stimuli are longer than the looking times to the consistent stimuli. An Analysis of Variance (Univariate) with only the looking times of the first four trials of the girls showed that there is a nearly-significant novelty effect ($F(1,32) = 3.599, p = .067$). This nearly-significant effect was not found across the whole dataset: analysis of all eight trials pointed out that girls did not show an overall novelty effect ($F(1,69) = 2.073, p = .154$).

Note that not all parents filled in the N-CDI, causing missing data for the questionnaires (we had N-CDI data for 18 out of 21 infants). We also investigated gender differences in the N-CDI scores for productive and receptive vocabulary, and in attention at familiarization quantified as the total looking time in this phase. Analyses of Variance conducted with gender as a fixed factor, and receptive vocabulary, productive vocabulary score and total looking time at familiarization as dependent variables revealed no difference between boys and girls with respect to the total looking times during familiarization ($F(1,16) = .945, p = .945$), or productive vocabulary ($F(1,16) = .378, p = .547$) or receptive vocabulary ($F(1,16) = .005, p = .944$).

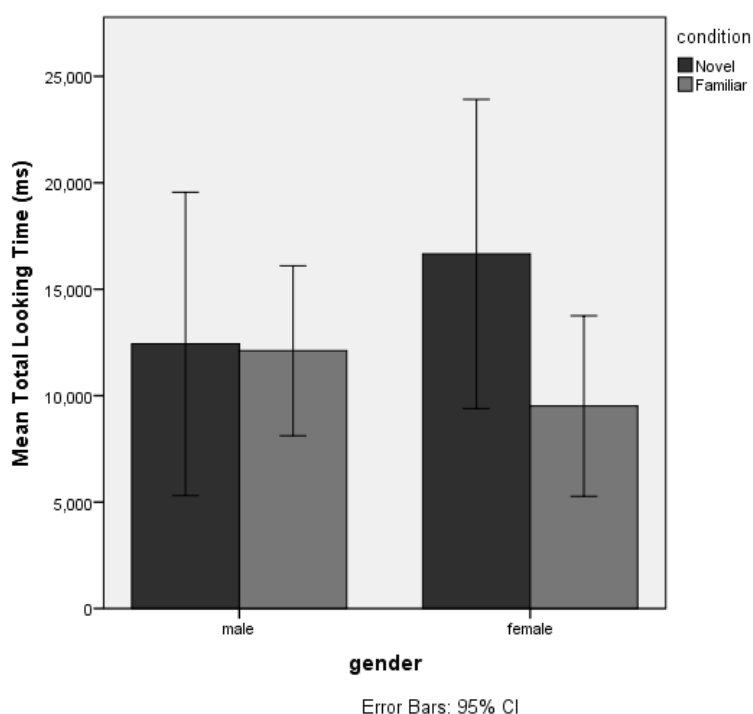


Figure 2. Chart of the mean looking times of the first four trials for each condition separated by gender

Discussion

We tested whether 18-month-old boys and girls differ in their ability to learn non-adjacent dependencies and generalize the learned pattern to novel nonsense words. The ability to learn patterns from spoken input may well play an important role in the acquisition of natural languages, and research has shown that infants have this ability at 18 months (Gómez and Maye, 2005; Kerkhoff et al., 2013). Although Lany and Gómez (2008) found a difference between the performance of male and female infants at a similar task with the generalization of adjacent patterns to non-adjacent ones, this findings had not been tested at 18-month-old infants.

Since infants show a rapid increase in productive and receptive vocabulary around the age of 18 months (the so-called ‘vocabulary spurt’), and previous research showed that girls have an advantage at language development over boys at this age (Bouchard et al., 2009), our data is quite surprising. We had expected a significant difference between the NCDI scores for productive and receptive vocabulary for boys and girls, which we did not found in our data. Furthermore, in addition

to previous mentioned artificial language learning studies (Gómez & Maye, 2005; Gómez, 2002; Kerkhoff et al., 2013; Lany and Gómez, 2008), we had expected our results to show differences between (i) the overall looking times for the familiar and novel stimuli and (ii) the looking times of boys and girls to these familiar and novel stimuli. Results did not show an overall discrimination of consistent and inconsistent test trials, and no significant interaction between gender and condition on the total looking time when all trials were analyzed. Analysis of the first four trials revealed significant interaction between gender and condition on the looking time by trial, suggesting a difference between boys and girls in the way they discriminate between consistent and inconsistent test items.

During the first four trials, girls alone showed an almost significant novelty effect. Gómez and Maye (2005) and Kerkhoff et al. (2013) found novelty effects in their studies with 18-month-old infants, but their tasks were less complicated since it did not include the generalization of X-elements. The novelty effect of girls alone contributed to these results; at a task that is slightly more complex, girls were able to distinguish word patterns, whereas boys were not. For boys this task might have been more difficult, since they were not able to distinguish the learned pattern from the novel one with the new X-items. The girls however tend to have a preference for novel stimuli, especially in the first four trials. The first four trials of the test phase could be a better reflection of the infants' learning than the entire test phase, since mainly the boys were more focused during the first four trials and the entire experiment is quite long for 18-month-olds. Furthermore, the last few trials might show effects of learning during the test phase, that is, infants starting to internalize strings from the inconsistent test trials, which would weaken their representations of the consistent word strings and the dependencies they had learned during the familiarization phase.

Surprisingly, the significant effect of receptive vocabulary on total looking time is a negative effect; infants with high N-CDI scores on their receptive vocabulary, looked less at the lights. This negative effect could be explained by the task itself; infants that understand a lot of words are ahead in language development compared to other infants in the same age group. The infants with high scores on the N-CDI in word-comprehension might have found the task boring and did not have that much overall attention because of it.

Contrary to our predictions, there were no significant differences between boys and girls for either receptive or productive vocabulary. The CDI-data of this study differs from the data used by Berglund et al. (2005) and the data of Bouchard et al. (2009) in several ways. The amount of filled in questionnaires differed a lot; Berglund et al. (2005) had 989 filled-in questionnaires, from which a small sample of 18-month-old infants ($n=36$) while all others were aged 8-to-30-months-old. Bouchard et al. (2009) had a total of 1,018 questionnaires for 18-month-old infants. The data of this last group was collected during a regular check-up at a health care center, which might explain the differences between the average N-CDI scores of the latter study and our own. Table 5 shows the mean results of all male and female infants in our study, the percentile-score according to the official N-CDI guide, and the average score 18-month-old infants have had on the N-CDI (50th percentile). This table reveals that the infants tested in this study have average scores that are much higher than regular average scores of the population 18-month-old, since the average percentile scores are respectively 70/75 for word comprehension and 60/80 for word production. Differences as shown in this table could possibly be attributed to the fact that most of the parents that decided to sign-up after the BabyLab recruitment, are very high educated. The different education levels of parents could be an influence, if it would mean that infants of high-educated parents are talked to a

Table 5.

N-CDI scores (raw scores and relative scores)

	Word comprehension	Percentile	Word production	Percentile
Mean male infants (n=9)	269	75	98	80
Mean female infants (n=10)	267	70	66	60
50 th percentile male infants	187	50	30	50
50 th percentile female infants	213	50	55	50

lot and get a lot of varied input, whereas infants of low-educated parents get less varied input, which could have made the difference between the language abilities of boys and girls bigger. This could affect the experiment, since these infants do not represent the whole population of 18-month-old infants.

A more representative sample of Dutch 18-month-olds would have been ideal, but this is probably very difficult to achieve given the sample bias inherent in voluntary recruitment. It might be that we do not have enough data and the difference is quite small, which means that a larger group of infants could make the data more clear. It could also be the case that there is no gender difference at all, but because of the publication bias we do not get all the studies that did not find any differences.

Although mean looking times of the familiarization phase showed a small difference between boys and girls, there was no significant correlation. It might be interesting to distinguish other groups that would explain the behavior of the infants during the whole experiment better. Measurements that start during the familiarization phase could be in an indication of the infants' attention throughout the experiment. The comparison of the looking times of infants that are high-attentive and low-attentive during familiarization phase, could affect the influence of condition on the total looking time during all trials of the test phase. The regular reported behavior and temperament of these infants could also be used together with the looking times of the familiarization phase to create different groups of infants that are high-attentive and low-attentive in their overall daily behavior. These groups could behave differently during the test phase and there might be differences between the two groups and their preference for consistent or inconsistent strings.

The findings of this study are remarkable and give a lot to reflect on. Surprisingly we did not find any significant gender differences in the N-CDI results of both receptive- and productive vocabulary. We did find a tendency for girls to show a novelty effect, whereas boys showed hardly any effect. The generalization of word patterns to novel nonsense words was probably too difficult for boys, since they were unable to distinguish familiar stimuli from novel stimuli. This task did not seem that difficult for girls, since they tend to a novelty effect and were able to generalize the familiar word patterns to novel stimuli.

Literature

- Andreano, J. M., & Cahill, L. (2009). Sex influences on the neurobiology of learning and memory. *Learning Memory, 16*, 248-266.
- Bouchard, C., Trudeau, N., Sutton, A., Boudreault, M.-C., & Denault, J. (2009). Gender differences in language development in French Canadian children between 8 and 30 months of age. *Applied Psycholinguistics, 30*(4), 685-707.
- Bornstein, M. H., Hahn, C.-S., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language, 24*(3), 267-304.
- Berglund, E., Eriksson, M., & Westerlund, M. (2005). Communicative skills in relation to gender, birth order, childcare and socioeconomic status in 18-month-old children. *Scandinavian Journal of Psychology, 46*, 485-491.
- Burman, D. D., Bitan, T., & Booth, J. R. (2008). Sex differences in neural processing of language among children. *Neuropsychologia, 46*(5), 1349-1362.
- Clements, A. M., Rimrod, S. L., Abel, J. R., Blankner, J. G., Mostofsky, S. H., Pekar, J. J., Denckla, M. B., & Cutting, L. E. (2006). Sex differences in cerebral laterality of language and visuospatial processing. *Brain and Language, 98*, 150-158.
- Galsworthy, M. J., Dionne, G., Dale, P. S., & Plomin, R. (2000). Sex differences in early verbal and non-verbal cognitive development. *Developmental Science, 3*(2), 206-215.
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431-436.
- Gómez, R. L., & Maye, J. (2005). The development trajectory of nonadjacent dependency learning. *Infancy, 7*(2), 183-206.
- Hartshorne, J. K., & Ullman, M. T. (2006). Why girls say 'holded' more than boys. *Developmental Science, 9*(1), 21-32.
- Hunter, M., & Ames, E. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In Rovee-Collier, C., & Lipsitt, L. (Eds.) *Advances in Infant Research* (pp. 69-95). Stamford: Ablex.
- Kerkhoff, A., De Bree, E., De Klerck, M., & Wijnen, F. (2013). Non-adjacent dependency learning in infants at familial risk of dyslexia. *Journal of Child Language, 40*(1), 11-28.
- Kemler-Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. A. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development, 18*, 111-116.
- Kramer, J. H., Kaplan, E., Delis, D. C., O'Donnell, L., & Prifitera, A. (1997). Developmental sex differences in verbal learning. *Neuropsychology, 11*(4), 577-584.
- Kramer, J. H., Delis, D. C., & Daniel, M. (1988). Sex differences in verbal learning. *Journal of Clinical Psychology, 44*(6), 905-915.
- Kaiser, A., Haller, S., Schmitz, S., & Nitsch, C. (2009). On sex/gender related similarities and differences in fMRI language research. *Brain Research Reviews, 61*, 49-59.
- Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science, 19*(12), 1247-1252.
- Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F. (1992). Overregularization in language acquisition. *Monographs of Society for Research in Child Development, 57*(4), 1-165.

- Sung, J., Fausto-Sterling, A., Coll, C. G., & Seifer, R. (2013). The dynamics of age and sex in the development of mother-infant vocal communication between 3 and 11 months. *Infancy*, *18*(6), 1135-1158.
- Wallein, M. (2009). Putative sex differences in verbal abilities and language cortex: A critical review. *Brain & Language*, *108*, 175-183.
- Zink, I., & Lejaegere, M. (2002). N-CDIs: Lijsten voor communicatieve ontwikkeling [N-CDIs: Lists for communicative development]. Leuven: Acco.