**Author**: Geambasu, A.

**Title**: Simple rule learning is not simple : studies on infant and adult pattern perception and production

**Issue Date**: 2018-12-11

# CHAPTER 2

# Rule learning in infants in the speech and general auditory domains

This chapter has been submitted and is under revision as Geambașu, A., van Renswoude, D., Visser, I., Raijmakers, M., & Levelt, C.C. (under revision). Marcus et al. (1999) revisited: which mechanism underlies infants' abstraction of algebraic rules?

The chapter was the first attempt within an interdisciplinary project to establish a baseline for rule learning in young infants in the speech and the general auditory domains.

## 2.1 Abstract

In this study we attempted to extend on, and then replicate closely, a seminal study from Marcus et al. (1999) showing that seven-month-old infants use algebraic rules to generalize from their input to novel instances. In Experiment 1, we investigated whether infants were able to learn an XYX or XYY pattern and generalize it to different levels of abstractness: (a) familiar syllables in novel combinations, (b) familiar syllables previously heard only in Y position now heard in X position, and (c) completely novel syllables. In Experiment 2, we familiarized infants to the same patterns as in Experiment 1, but tested them only with novel syllables. In these first two experiments, we used a naturally-recorded, phonologically balanced set of phonemes as our stimuli. In Experiment 3, we reverted to a procedure and stimuli as similar as possible

to the original Marcus et al. (1999) study. Across the three experiments, using two different paradigms and using exposure and testing stimuli of gradually fewer degrees of difference from the original study, we were unable to replicate the results of Marcus et al. (1999). We show that indeed learning of these seemingly simple rules is, in fact, quite difficult and may be subject to specific constraints. Our findings showing a consistent preference for triads containing adjacent repetition line up squarely with the theory of perceptual or memory primitives (POMPs) of Endress et al. (2009), who have proposed that much of the artificial grammar learning findings can be explained through the recruitment of domain-general cognitive processes, including the sensitivity for edges and for repetitions (Mehler, Nespor, & Peña, 2008). This sensitivity has been found across species, domains, and age groups. Our results add support to the theory that repetitions, especially at edges, are highly salient for young learners.

## 2.2 Introduction

Learning and representing rules lies at the heart of understanding cognition and cognitive development. Rule learning is implicated in reasoning, in causal learning, in social learning, and in language learning (Bunge & Zelazo, 2006). Specifically in the domain of language and language acquisition, the ability to generalize combinatorial rules allows us to learn and apply properties of language structure beyond the input.

From a young age and without explicit instruction, infants are able to pick up on the regularities in their linguistic input related to frequencies, positions, and combinatorial possibilities of sounds within words, and words within phrases. Very young infants have been found to use statistical properties of language to learn about relevant categories and units. Infants between six and nine months of age have been shown to use the phonotactic patterns in their native language to recognize what is or is not a likely word in the language (Jusczyk & Luce, 1994). Between six and 7.5 months, they are also able to extract monosyllabic words from fluent speech (Jusczyk & Aslin, 1995), and at eight months, they can use transitional probabilities to extract multisyllabic words (Saffran et al., 1996).

Noticing regularities related to word segmentation seems to develop in parallel with the ability to learn how segmented units can be combined. Marcus et al. (1999) proposed that the ability to learn how items can be combined from a specific exposure set, and the ability to apply the extracted rule to novel items in another set, is done by learning abstract, "algebraic" rules, which make use of variables. They showed that already at seven months, infants were able to learn and generalize simple rules (XXY, XYY, XYX – henceforth referred to as Marcus rules) from a two-minute familiarization stream, where the relevant units were delimited by a pause (in Marcus et al., 1999 referred to as "sentences", here referred to as "triads"). This seminal study has been widely cited as proof that infants have the ability to learn algebraic rules, quickly abstracting structure beyond their input.

Marcus and colleagues have since argued that speech is special when it comes to learning and generalizing such rules, as opposed to other domains, such as the visual domain (Frank et al., 2009), the general auditory domain (Marcus et al., 2007), or even the context of sign language (Rabagliati et al., 2012). Marcus et al. (2007) found that while 7.5-month-old infants were unable to learn to generalize the simple Marcus rules when they were formed of tones produced by musical instruments, they were able to apply such rules to the tone test stimuli when they had first been learned on the basis of speech stimuli. The authors concluded that speech might facilitate rule learning (or rule generalization) in domains where infants might otherwise not acquire rules. In a similar vein, Rabagliati et al. (2012) found that hearing infants at the same age were able to learn XYY, but not XXY, when the rules are composed of visual presentations of sign language signs. The authors concluded that the asymmetric results indicate that learning these patterns in a different domain is difficult and that speech may be special, at least to hearing children who are learning from speech (as opposed to non-hearing children for whom sign might be special). Frank et al. (2009) found that five-month-old infants could only learn an XYX or XYY pattern from multimodal, concordant exposure to visual and speech stimuli. Learning from either of the unimodal speech or visual conditions did not occur, indicating that even with speech, the mechanisms for generalization are not yet available to such young infants unless they are supported by redundant information.

Other research has countered the hypothesis that speech is special. Ferguson and Lew-Williams (2016) showed that seven-month-olds could learn XYY rules from tones when they were pre-exposed to a video where tones were used as communicative signals. In the visual domain, Saffran et al. (2007) found that seven-month-old infants were able to learn Marcus rules with images of either different dog breeds or different cat breeds, indicating that the ability is not domain specific. In this case infants succeeded in both an XYX vs. XYY or XYY vs. XXY condition. Ferguson and Waxman (2015) also found that infants as young as three months old could perform the same task, when the visual patterns were accompanied by child-directed speech (not following the same pattern). Adding to the mixed findings, S. P. Johnson et al. (2009) found that eight- and eleven-month-old infants were not always able to learn rules from visual stimuli consisting of shapes. The younger infants were only able to discriminate a non-adjacent repetition when trained on an late repetition (XYY vs. XYX) but not vice versa. The older infants could discriminate early repetitions (XXY) from late repetitions (XYY), but they could only discriminate XXY from XYX when familiarized with XXY. An XYX training grammar thus did not result in rule learning by either age group. From this work we get a more nuanced view that suggests that rule learning is possible in the visual domain but that it may be subject to more cross-developmental variability.

This work also suggests that rules composed of adjacent repetitions are learned more easily.

In the speech domain, Kovács and Endress (2014) showed a similar finding in a rule learning experiment with seven-month-olds in which various syllables were used to create XYX-, XYY-, and XXY-structured words that were subsequently organized hierarchically within either XYY- or XXY-structured sentences. They found that infants could learn the sentence patterns containing adjacent repetitions, but not those containing non-adjacent repetitions. Their work also shows that sensitivity to adjacent repetitions may be an important cue for generalization.

Despite these extensions on the original Marcus et al. (1999) paradigm, the only published replication using the same stimuli and paradigm is presented in Gerken (2006). However, this study still did not constitute a true replication and contains a number of differences with Marcus and colleagues' original work. Gerken could not replicate the original findings with seven-month-olds but was able to do so with nine-month-olds. In addition, the results of Marcus et al. (1999) could only be replicated when an early repetition grammar was used (XXY), instead of the original late repetition grammar (XYY); pilot testing contrasting XXY and XYY yielded no learning and a "marginal" preference for XYY. With respect to the stimuli, Gerken used only one block of four triads in the familiarization as opposed to the 16 in Marcus et al., and only four triads in the test as opposed to three blocks of the four triads as Marcus et al. did. A footnote reveals that using the full set of 16 familiarization triads failed to replicate (when testing XYY vs. XXY). Finally, unlike Marcus et al. , Gerken found a familiarity effect instead of novelty effect.

The mixed findings in experiments that vary in different ways from the original Marcus et al. (1999) experiment are not trivial, as they indicate that the original findings may not be robust or may only hold under specific conditions. The question remains of under what conditions infants are able to learn such rules and what types of cues they may be sensitive to.

### 2.2.1 The present work

In the present work, we attempted to extend on the Marcus et al. (1999) and the Marcus et al. (2007) findings to understand under which conditions Marcus rules are learnable. In Experiment 1, we aimed to test whether age, stimulus type, or specific information in the test trials would either support or hinder rule generalization. To this end we created two exposure and test conditions, with the intention of testing whether infants better learned the rules when they were composed of speech sounds or whether the natural, harmonic properties of birdsong would also support learning. These conditions will henceforth be referred to as the Speech and Song conditions.

In the Speech condition, we created natural recordings[1] of a number of syllables. We hypothesized that the type of learning shown in Marcus et al. (1999) should only be enhanced when using natural speech. We thus took care to balance the syllables' phonological features with the intention of being more representative of natural speech than Marcus et al.'s (1999) set, which uses a very limited set of sounds. Some examples of the limitations include the fact that familiarization triads were composed of only two vowels, /e/ and /i/, both of which are front vowels, that all test consonants are stops, that all familiarization syllables and three of the four test syllables are homorganic[2], that three of the four familiarization consonants are coronal, that four triads had identical vowels when syllables were combined (*wilili, wididi, lewewe, dewewe*), and that one triad had identical consonants when syllables were combined (*wiwewe*). Table 2.1 shows syllables used in Marcus et al. (1999) as compared to the syllables used in our experiments.

In the Song condition, we aimed to directly test the findings of Marcus et al. (2007) that speech is special by testing learning in a non-speech condition with auditory stimuli found in nature[3].

In both conditions, we tested infants with a variety of test trial types, including previously-heard syllables in novel combinations maintaining the original ordinal positions within triads (Combination), previously-heard syllables in novel positions (Place), and completely novel syllables (Generalization). With these three types of testing conditions, we aimed to draw comparisons between the abilities of infants, adults (Geambașu, Spierings, ten Cate, & Levelt, in prep.), and birds (Spierings & ten Cate, 2016) within a larger comparative project. While zebra finches were able to apply the learned rule only to familiarized sounds (Combination), budgerigars were able to apply the learned rule at all levels (Spierings & ten Cate, 2016). We expected the infants in this experiment would also be able to do this: if algebraic learning has occurred, variables are formed for X and Y items and a rule is formed for how to combine them, irrespective of the familiarity and positions of syllables. However, this experiment failed to deliver evidence of learning. Due to the complex nature of the experiment, we simplified the test and honed in on Generalization items using only speech stimuli in Experiment 2. When this experiment also failed to show evidence of learning, we conducted Experiment 3, a failed attempt at a close replication of Marcus et al.'s original study. In next sections of this paper

---

[1] Since we used natural birdsong we decided to use natural speech stimuli as well. In addition, Saffran and Wilson (2003) pegged the high drop out rate in their experiment on the disinterest of the infants in synthetic synthesized speech. Thus, natural recordings were considered more likely to engage the infants in the task.

[2] Homorganic sounds are those that share place of articulation. For example, coronal consonant /d/, and coronal, close-mid, front vowel /eː/ share the feature of being produced with the coronal part of the tongue in the front of the mouth, making /deː/ a homorganic syllable. Infants produce homorganic syllables first (C. C. Levelt, 1994) and prefer these sounds over non-homorganic ones early in development (ter Haar & Levelt, 2018).

[3] Birdsongs were also used for the purpose of cross-species comparison work conducted by project colleagues (Spierings & ten Cate, 2016).

**Table 2.1:** Syllable chart of possible consonant and vowel combinations. Syllables boxed with double lines were used in Marcus et al. (1999)'s experiment. Syllables with in boxes with polka dot fill were used for as test syllables in Marcus et al. (1999). Familiarization and test syllables were varied for each participant in our experiments. Syllables in bold were used in our experiments 1 and 2. Syllables with boxes with single lines replaced syllables /dʒi/ and /dʒe/ in our Experiment 3. Homorganic syllables used in the experiments are underlined. Unused syllables are grayed out.

| Consonants<br>Vowels | | /a:/ | /e:/ | /i/ | /o:/ | /u/ |
|---|---|---|---|---|---|---|
| Voiced | Features | dorsal, back, open | coronal, front, close-mid | coronal, front, close | labial/round, back, close-mid | labial/round, back, close |
| /m/ | labial | ma: | me: | **mi** | **mo:** | mu |
| /b/ | labial | ba: | be: | **bi** | **bo:** | bu |
| /v/ | labial | va: | ve: | vi | vo: | vu |
| /w/ | labial | wa: | we: | wi | wo: | wu |
| /ʤ/ | coronal | ʤa: | ʤe: | ʤi | ʤo: | ʤu |
| /d/ | coronal | da: | **de:** | di | **do:** | du |
| /l/ | coronal | la: | **le:** | li | lo: | **lu** |
| /n/ | coronal | na: | ***ne:*** | **ni** | **no:** | **nu** |
| /g/ | dorsal | ga: | ge: | gi | go: | gu |
| Unvoiced | | | | | | |
| /f/ | labial | fa: | fe: | **fi** | **fo:** | fu |
| /p/ | labial | pa: | pe: | pi | pɑ: | **pu** |
| /s/ | coronal | **sa:** | se: | si | so: | **su** |
| /t/ | coronal | **ta:** | ***te:*** | **ti** | to: | tu |
| /k/ | dorsal | **ka:** | **ke:** | ki | kɑ: | ku |
| /X/ | dorsal | ga: | **ge:** | gi | **go:** | ***gu*** |

we describe the three experiments in detail and present both a frequentist and Bayesian analysis of the data.

## 2.3 Experiment 1

In Experiment 1, we extended on the foundation of Marcus et al.'s (1999) study. We exposed two age groups (six- and nine-month-old infants) to one of two patterns (XYX or XXY), in a Song or a Speech condition, and tested their ability to apply the learned rule to various levels of novelty during test.

### 2.3.1 Methods

**Stimuli**

**Speech condition.** Because the stimuli used in the original Marcus et al. (1999) study were phonologically limited, in the Speech condition, we expanded the group of syllables used as exposure and test stimuli. We balanced the number of syllables that used voiced and unvoiced consonants, and the number of syllables composed of homorganic and not homorganic sounds. In addition, our triad combinations differed from Marcus et al.'s in that all syllables could be used as either an X or Y item, as long as triad followed the rule.

The stimuli in the Speech condition consisted of natural recordings of consonant-vowel (CV) syllables made by a phonologically trained female native speaker of Standard Dutch (C.C.L.). The original CV recordings were manipulated such that each syllable was approximately 400 ms in total duration, all vowels were similar in length, and all consonants of the same class (stops, fricatives) were similar in length (vowel duration values taken from Koopmans-van Beinum, 1980). A total of eight unique syllables were selected[4], and each syllable was assigned to a letter A through H, in eight different randomizations (see Appendix A1 for a table of all syllables and to which letter they corresponded to in each randomization). These randomizations were made such that infants' performances could not be subject to biases or preferences for specific syllables. The syllables were combined into triads by concatenating three CV syllables as shown in Table 2.2 (see also Appendices B1 and B2 for a complete list of triads per randomization for the familiarization and test phases respectively). The CV syllables within a triad were separated by 100 ms silent pause. This created triads approximately 1.4 seconds in duration. Both the familiarization and test strings were concatenated with a one second pause separating each triad. Each triad's pitch was standardized at a constant 125 Hz such that pitch sounded natural for a female voice and would not highlight any specific syllable in the triad, thus providing no supporting information with respect to the structure of the triad.

---

[4] Eight syllables could be selected so as to match the number of unique zebra finch song elements available.

**Table 2.2:** Each letter A through H was assigned to a unique syllable. Syllables were then combined to form 30 familiarization triads and 18 test triads, half of which followed the XYX pattern and half of which followed the XXY pattern. To avoid influence of preference for specific sounds, each infant heard one of the possible eight combinations (see Appendices A1, B1, and B2).

| Familiarization triads | | Combination test triads | | Position test triads | Position test triads | Generalization test triads | |
|---|---|---|---|---|---|---|---|
| **XYX** | **XXY** | **XYX** | **XXY** | **XYX** | **XXY** | **XYX** | **XXY** |
| ADA | AAD | ACA | AAC | FAF | FFA | GHG | GGH |
| AEA | AAE | BDB | BBD | FEF | FFE | HGH | HHG |
| AFA | AAF | CEC | CCE | | | | |
| BAB | BBA | DCD | DDC | | | | |
| BCB | BBC | EAE | EEA | | | | |
| BFB | BBF | | | | | | |
| CAC | CCA | | | | | | |
| CDC | CCD | | | | | | |
| CFC | CCF | | | | | | |
| DBD | DDB | | | | | | |
| DED | DDE | | | | | | |
| DFD | DDF | | | | | | |
| EBE | EEB | | | | | | |
| ECE | EEC | | | | | | |
| EFE | EEF | | | | | | |

**Song condition.** Stimuli used in the Song condition were also composed of natural recordings of eight zebra finch song elements produced in the Leiden University Institute of Biology. The duration of each element was between approximately 35 and 86 ms. As in the speech condition, triads were concatenated with 100 ms between each element, and strings were concatenated with a one second silent pause between triads. Also as in the Speech condition, all elements could be used as either an X or a Y, as long as the triad followed the rule. Song triads ranged in duration from approximately 400 ms to 560 ms including pauses.

### Participants

A total of 19 six-month-olds and 20 nine-month-olds were tested in the Speech condition; 29 six-month-olds and 27 nine-month-olds were tested in the Song condition.

Infants were recruited via the municipality of Leiden. Parents were briefed on the intention of the study and subsequently gave their informed consent for their infants' participation. Parents and infants were compensated for their voluntary participation with reimbursed travel costs and a gift of a book.

Infants were removed from analysis for the following reasons: experimenter or technical errors (n=6), birth more than three weeks pre-term[5] (n=7), and excessive fussiness (n=4).

Sixteen infants were bilingual to some extent, ranging from hearing some foreign language from a grandparent to having at least one parent who spoke a foreign language to them regularly. Most of the bilingual infants only participated in the Song condition such that the speech stimuli would not be more foreign to them than for the rest of the subject group. The three bilingual infants who were tested in the speech condition had Dutch mothers and heard Dutch most of the time.

While Marcus et al. (1999) tested seven-month-old infants, we tested six- and nine-month-olds. In testing on either side of this age, we expect to see a developmental difference in generalization abilities.

### Procedure

Infants were tested in the Visual Fixation Procedure (VFP; Cooper & Aslin, 1990) using E-Prime software (*E-Prime 2.0, Psychology Software Tools, Inc.*, 2012). While the VFP testing method is different from the Headturn Preference Procedure (HPP) used in Marcus et al. (1999), both methods rely on the same principle that infants will continue to look towards a visual stimulus for as long as they are attending to the auditory stimulus playing at the same time (E. K. Johnson & Zamuner, 2010).

The experiment consisted of an initial attention grabber, a familiarization phase, and a testing phase. The experiment began with the infants being seated

---

[5] According to the protocol of the Many Babies Consortium (unpublished protocol).

on a caregiver's lap in a sound-attenuated booth. Caregivers wore tight-fitting headphones that played a mixture of music and backwards speech to mask the auditory stimuli of the experiment. This ensured that caregivers would not reflexively respond to the auditory stimuli and unwittingly influence the behavior of the infant. The experiment began with an initial attention grabber, composed of a video of green LED lights, to draw the infant's attention to the center. Once the infant looked towards the attention grabber for two continuous seconds, the familiarization phase would begin. The familiarization phase consisted of an approximately two-minute (1 minute 53 seconds) continuous exposure phase. The continuous auditory stream included three repetitions of 15 different triads. While the auditory stimuli played, infants saw a video of a flashing red light on the screen directly in front of them to maintain their attention. Once the familiarization stream ended, the attention grabber played again. Once the infant looked towards the attention grabber for two seconds, the testing phase began. During the testing phase, infants saw the same video of flashing lights as during the familiarization. However, in contrast to the familiarization phase, the test trials only played while the infant was looking. Infants could look to a trial for a maximum time of 15 seconds, after which the trial would automatically end. If the infant looked away during the trial for more than two seconds continuously, the trial would automatically stop and an attention grabber would begin until the infant again focused to center for two continuous seconds.

During the test phase, infants heard three different types of test triads. The first, we refer to as Combination triads, which were made up of syllables infants had heard during familiarization but which were now rearranged into previously unheard combinations. The second, we refer to as Place triads, composed of syllables in which the X test item had only been heard in the Y position during the familiarization. The third type of test type we refer to as Generalization triads. These were, as in the original Marcus et al. (1999) experiment, triads composed of syllables which had not been heard during the familiarization phase. Half of all test triads followed the pattern of the familiarization phase (consistent test items), and half followed the opposite pattern (inconsistent test items). The test items were presented in an unblocked manner and randomized per participant. The test phase included 18 total test trials (see Table 2.2 and Appendix B2) resulting in a test phase of approximately 10 to 15 minutes.

A video camera recorded the infants behavior for offline coding and allowed the experimenter to react to the infants' behavior online. Infants' behavior was coded offline by a coder blind to the stimuli using ELAN software (*ELAN Version 5.0.0*, 2017; Sloetjes & Wittenburg, 2008) and precise looking times were calculated for each trial.

### 2.3.2 Results

**Data preprocessing**

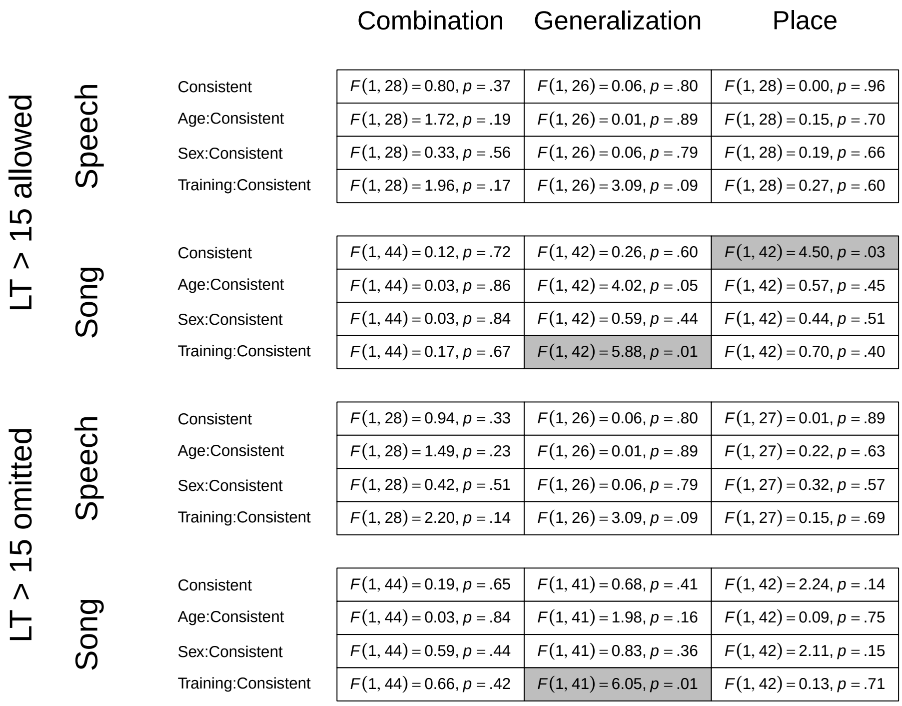
Apart from the inclusion and exclusion of outliers, there are many decision to be made regarding the preprocessing of infant looking data. For instance, it is common practice to exclude looking times (LTs) that are shorter than the time it takes for one repetition of the stimulus to play (Ferguson & Waxman, 2015; Saffran & Wilson, 2003). Following the latter model, the data is analyzed without LTs shorter than 1.5 seconds for the speech and shorter than 0.5 seconds for the sound stimuli (10% of the data).

Another consideration is that there are LTs that exceed the 15 second presentation time of the stimuli (3% of the data). These LTs could be caused by infants not responding to the stimuli but simply staring (or, in the case of the Headturn Preference procedure, reflect a preference for one side of the booth). Ideally, only data points reflecting a preference would be kept and staring behavior would be omitted. However, this distinction cannot be made based on the data, and the literature provides no clear way to handle such cases, although removal of LTs reaching the maximum trial duration have been removed in previous studies (Bernard & Gervain, 2012; Gervain & Werker, 2013). We therefore decided to report the analyses both with and without LTs above 15 seconds.

In all experiments, data was first analyzed using repeated measures ANOVA (labeled 'Frequentist analysis' per experiment) and subsequently using Bayesian paired samples t-tests (labeled 'Bayesian analysis' per experiment). In addition to the more commonly-used frequentist analyses, we choose to include Bayesian analyses as they allow us to quantify relative evidence of competing hypotheses (Wagenmakers, Morey, & Lee, 2016). That is, a certain hypothesis (H1) can be more likely than another hypothesis (H0) or vice versa, given the data. In our studies H0 and H1 are clearly defined: "Infants will show preferential looking based on the test grammar's inconsistency with the familiarization grammar" (H1) or "Infants Infants will not show preferential looking based on the test grammar's inconsistency with the familiarization grammar" (H0). An additional benefit of this procedure is that null-results can be interpreted, since in a Bayesian analysis, H0 can be 'x' times more likely than H1 given the data. Such an interpretation is impossible with frequentist analyses, since the truth of the null-hypothesis is an assumption of frequentist analyses (Cohen, 1994).

**Frequentist analysis**

There were three conditions (Combination, Generalization, and Place) with two type's of stimuli (Song and Speech) creating six outcome options. Together this results in 12 LTs to compare between the Consistent and Inconsistent condition. To make the matter more complicated, there are also possible interactions with sex, age, and type of training stimuli (i.e., XYY or XXY) that should be



| | | | Combination | Generalization | Place |
|---|---|---|---|---|---|
| LT > 15 allowed | Speech | Consistent | $F(1, 28) = 0.80, p = .37$ | $F(1, 26) = 0.06, p = .80$ | $F(1, 28) = 0.00, p = .96$ |
| | | Age:Consistent | $F(1, 28) = 1.72, p = .19$ | $F(1, 26) = 0.01, p = .89$ | $F(1, 28) = 0.15, p = .70$ |
| | | Sex:Consistent | $F(1, 28) = 0.33, p = .56$ | $F(1, 26) = 0.06, p = .79$ | $F(1, 28) = 0.19, p = .66$ |
| | | Training:Consistent | $F(1, 28) = 1.96, p = .17$ | $F(1, 26) = 3.09, p = .09$ | $F(1, 28) = 0.27, p = .60$ |
| | Song | Consistent | $F(1, 44) = 0.12, p = .72$ | $F(1, 42) = 0.26, p = .60$ | $F(1, 42) = 4.50, p = .03$ |
| | | Age:Consistent | $F(1, 44) = 0.03, p = .86$ | $F(1, 42) = 4.02, p = .05$ | $F(1, 42) = 0.57, p = .45$ |
| | | Sex:Consistent | $F(1, 44) = 0.03, p = .84$ | $F(1, 42) = 0.59, p = .44$ | $F(1, 42) = 0.44, p = .51$ |
| | | Training:Consistent | $F(1, 44) = 0.17, p = .67$ | $F(1, 42) = 5.88, p = .01$ | $F(1, 42) = 0.70, p = .40$ |
| LT > 15 omitted | Speech | Consistent | $F(1, 28) = 0.94, p = .33$ | $F(1, 26) = 0.06, p = .80$ | $F(1, 27) = 0.01, p = .89$ |
| | | Age:Consistent | $F(1, 28) = 1.49, p = .23$ | $F(1, 26) = 0.01, p = .89$ | $F(1, 27) = 0.22, p = .63$ |
| | | Sex:Consistent | $F(1, 28) = 0.42, p = .51$ | $F(1, 26) = 0.06, p = .79$ | $F(1, 27) = 0.32, p = .57$ |
| | | Training:Consistent | $F(1, 28) = 2.20, p = .14$ | $F(1, 26) = 3.09, p = .09$ | $F(1, 27) = 0.15, p = .69$ |
| | Song | Consistent | $F(1, 44) = 0.19, p = .65$ | $F(1, 41) = 0.68, p = .41$ | $F(1, 42) = 2.24, p = .14$ |
| | | Age:Consistent | $F(1, 44) = 0.03, p = .84$ | $F(1, 41) = 1.98, p = .16$ | $F(1, 42) = 0.09, p = .75$ |
| | | Sex:Consistent | $F(1, 44) = 0.59, p = .44$ | $F(1, 41) = 0.83, p = .36$ | $F(1, 42) = 2.11, p = .15$ |
| | | Training:Consistent | $F(1, 44) = 0.66, p = .42$ | $F(1, 41) = 6.05, p = .01$ | $F(1, 42) = 0.13, p = .71$ |

**Figure 2.1:** Multiverse of $p$-values for Experiment 1.

assessed, creating 48 $p$-values to evaluate. This expansion is called the data multiverse (Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016) and is very common in many psychological experiments, although it is almost never made explicit that there were so many outcome options. In order to increase transparency we decided to report all results. Figure 2.1 highlights $p$-values below .05 in gray and clearly shows an overall failure to replicate earlier findings (Gerken, 2006; Marcus et al., 1999). The only significant main effect (if we do not consider multiple comparison issues) occurs in the Place testing trials in the Song condition, and is in the opposite direction: LTs are longer during consistent than inconsistent trials. The significant interaction effects between training and condition are visualized in Figure 2.2 with error bars of ±1 standard error. These effects indicate that infants have a preference for adjacent repetition (i.e., a preference for XXY versus XYX), independent of familiarization grammar.

### Bayesian analysis

It is difficult to draw conclusions from the frequentist analysis presented above, because null results cannot be interpreted within the frequentist framework.

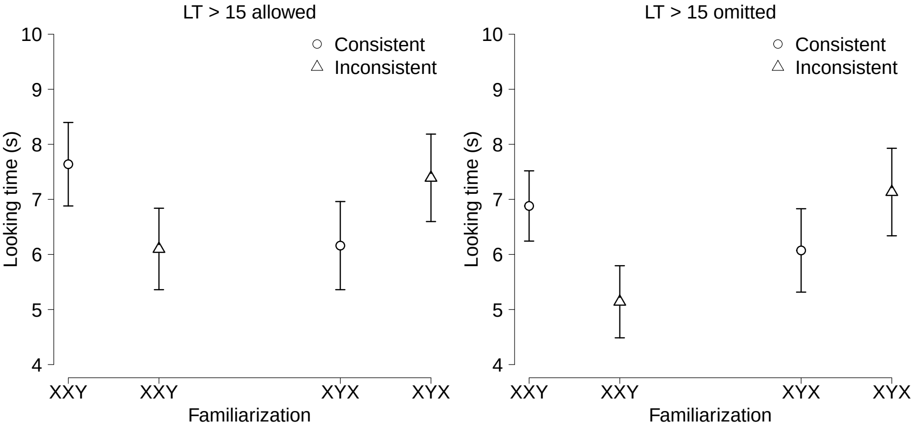


**Figure 2.2:** Interactions of familiarization and test grammars in Experiment 1. The left panel shows results when LTs of longer than 15 seconds are allowed in the analysis, while the right panel shows results when LTs of longer than 15 seconds are omitted from analysis. inclusions or omission did not change the pattern of results: infants familiarized with XXY look longer to the consistent test grammar, while infants familiarized with XYX look longer to the inconsistent test grammar.

Yet it would certainly help our understanding of infants' behavior if we could interpret these null effects. The Bayesian framework provides the possibility to accumulate evidence in favor of the null hypotheses and interpret how likely the null hypothesis is in comparison to other hypotheses (Jeffreys, 1961). This is possible by calculating Bayes Factors (BFs), which indicate how many times more likely one hypothesis is over another. In this case, we use BFs to quantify how much more likely the null hypothesis (that looking times are equal for both the consistent and inconsistent sounds) is than the alternative hypothesis (that infants look longer when they hear inconsistent sounds).

To make this comparison we use JASP (JASP Team, 2017), an alternative for SPSS using Bayesian analyses. In the Bayesian framework it is possible to accumulate evidence with every new data point. This can be seen in Figures 2.3 and 2.4 where the number of participants is plotted on the x-axis and the BF is plotted on the y-axis. With the data of every new participant, the BF is updated and the evidence accumulates. The sequential analyses for the Speech (Figure 2.3) and Song (Figure 2.4) conditions for the generalization test trials are shown, with LTs longer than 15 seconds allowed. Using BFs does not solve the problem that there are many possible comparisons that can be made. In Table 2.3 we report Bayes Factors that indicate how much more likely the null hypothesis is than the alternative hypothesis for three types of test trials (Combination, Generalization, Place), two stimulus conditions (Speech, Song), and two types of preprocessing options (with and without looking times longer than 15 seconds). Note that for every combination the null hypothesis is

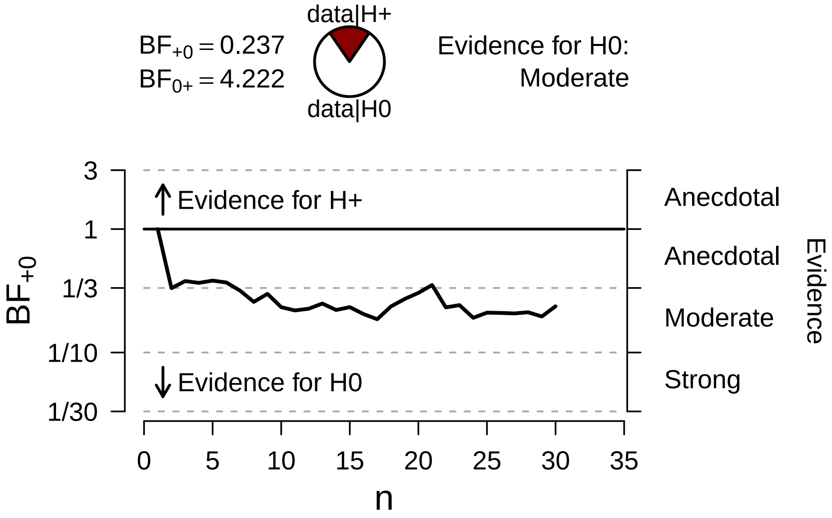


**Figure 2.3:** Sequential analyses for Experiment 1 for the speech stimuli in the generalization conditions with LTs longer than 15 seconds allowed.

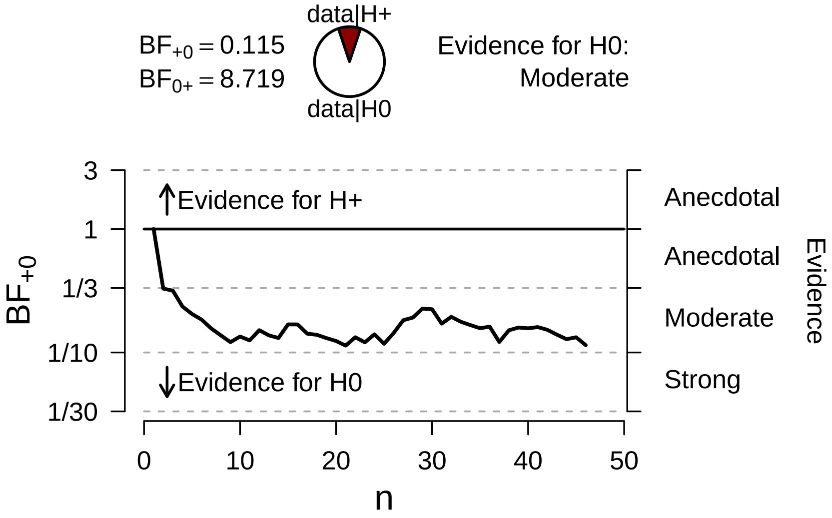


**Figure 2.4:** Sequential analyses for Experiment 1 for the song stimuli in the generalization conditions with LTs longer than 15 seconds allowed.

more likely than the alternative hypothesis as all BFs are greater than 1. Also note that the highest BFs can be found in the cells with lowest $p$-values. This may seem counter-intuitive, but is the case because the small effects that are found are in the opposite direction. Infants look longer when consistent patterns are presented than when inconsistent patterns are presented, providing more evidence against a preference for inconsistent sounds.

**Table 2.3:** Bayes Factors that indicate how much more likely H0 (that infants do not have a preference) is than H1 (that infants have a preference for inconsistent sounds), presented for three conditions (Combination, Generalization, Place), two types of stimuli (Speech, Song), and two types of preprocessing options (with and without looking times longer than 15 seconds).

| | Combination | Generalization | Place |
|---:|---:|---:|---:|
| Speech, all LT | 9.23 | 4.22 | 5.47 |
| Song, all LT | 4.70 | 8.72 | 18.38 |
| Speech, LT $< 15$ | 9.56 | 4.22 | 5.78 |
| Song, LT $< 15$ | 4.31 | 10.26 | 15.08 |

### 2.3.3 Discussion

The analysis did not show evidence of differential looking to consistent or inconsistent test patterns overall except in the specific case of Place test items in the Song condition if LTs over 15 seconds were allowed. If those long looking times were removed the effect was no longer found, indicating that it was not a strong effect to begin with. No other main effects were found.

There were, however, interaction effects between the familiarization pattern and the consistency of the test pattern. Infants preferred the XXY pattern, independent of the pattern with which they were familiarized.

Counter to our initial hypotheses, we did not find evidence for better performance when infants were exposed to speech, nor did we find any effect of development. We recognized that the testing conditions presented in this experiment may have placed an unrealistically high cognitive load on the infants and that this may have obscured any potential evidence of learning. Including test items with various degrees of difference from the familiarization items, and in unequal numbers, may have been ecologically valid, as infants hear a variety of familiar and novel items intermixed on a daily basis, but the amount of variety may have confused them, not allowing them to show differential attention on the basis of pattern alone. We thus hypothesized that if the testing sets were simpler and included only generalization test items, as in Marcus et al. (1999), infants might be more prone to show differential looking between test items consistent or inconsistent with the familiarization pattern.

## 2.4 Experiment 2

### 2.4.1 Methods

#### Stimuli

The stimuli in Experiment 2 differed from those used in Experiment 1 only with respect to the test items used. In this experiment, only the generalization items, GGH, GHG, HHG, and HGH were used in test (see Table 2.2 and Appendix B2). Because we found no difference with respect to the use of speech or song in Experiment 1, during this experiment, only speech stimuli were used.

#### Participants

We tested 25 six-month-olds and 22 nine-month-olds. Recruitment and compensation was identical to Experiment 1.

Infants were removed from analysis for the following reasons: technical errors (n=5) and excessive fussiness (n=1). Ten of the infants in this experiment were exposed to another language in addition to Dutch, with all infants having one Dutch parent and being exposed to Dutch both in- and outside the home.

#### Procedure

The procedure was identical to Experiment 1, with the exception of the type and number of test trials. Four generalization test items were used, two that were consistent with the training pattern, and two that were inconsistent with the training pattern. These four test trials were repeated three times, for a total of 12 test trials (cf. Marcus et al., 1999).

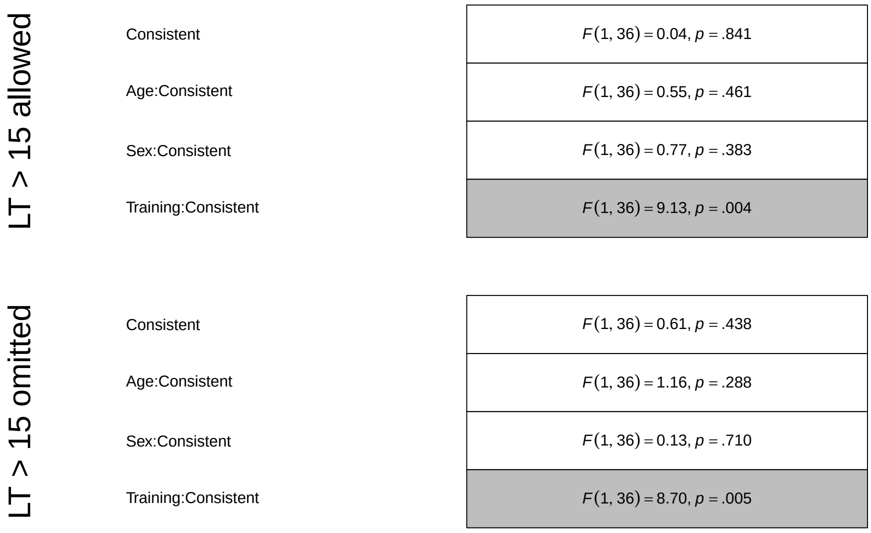
### 2.4.2 Results

#### Data preprocessing

As in Experiment 1, LTs that were shorter than the duration of one triad (1.5 seconds) were excluded (4% of the data). Again, there were LTs that exceed the 15-second presentation time of the stimuli (6% of the data) and we report the analyses both with and without these LTs.

#### Frequentist analysis

The multiverse analysis (Figure 2.5) did not find any differences between including or omitting LTs longer than 15 seconds. In both cases, there were no significant effects of consistency, but there was an interaction effect between training and consistency, similar to the interaction effect found in Experiment 1. In order to interpret this interaction effect, the mean LTs for the consistent and inconsistent test stimuli and familiarization stimuli are visualized in Figure



| | | |
|---|---|---|
| LT > 15 allowed | Consistent | $F(1, 36) = 0.04, p = .841$ |
| | Age:Consistent | $F(1, 36) = 0.55, p = .461$ |
| | Sex:Consistent | $F(1, 36) = 0.77, p = .383$ |
| | Training:Consistent | $F(1, 36) = 9.13, p = .004$ |
| LT > 15 omitted | Consistent | $F(1, 36) = 0.61, p = .438$ |
| | Age:Consistent | $F(1, 36) = 1.16, p = .288$ |
| | Sex:Consistent | $F(1, 36) = 0.13, p = .710$ |
| | Training:Consistent | $F(1, 36) = 8.70, p = .005$ |

**Figure 2.5:** Multiverse of $p$-values for Experiment 2.

2.6 with error bars of ±1 standard error. As in Experiment 1, infants seemed to have a preference for patterns including adjacent repetitions (XXY).

### Bayesian analysis

To help our understanding of the null effect we also performed a sequential Bayesian analysis, using BFs to quantify how much more likely the null hypothesis is than the alternative hypothesis. Figures 2.7 and 2.8 shows how evidence accumulates in favor of the null hypothesis for both the data with and without LTs longer than 15 seconds included.

## 2.4.3 Discussion

Infants in Experiment 2, like those in Experiment 1, did not show differential looking between consistent and inconsistent test items overall. As in Experiment 1, we found an interaction between the familiarization pattern and consistency of the testing pattern. This result indicates that in Experiment 2, infants also preferred to look more to XXY test items during the test phase,

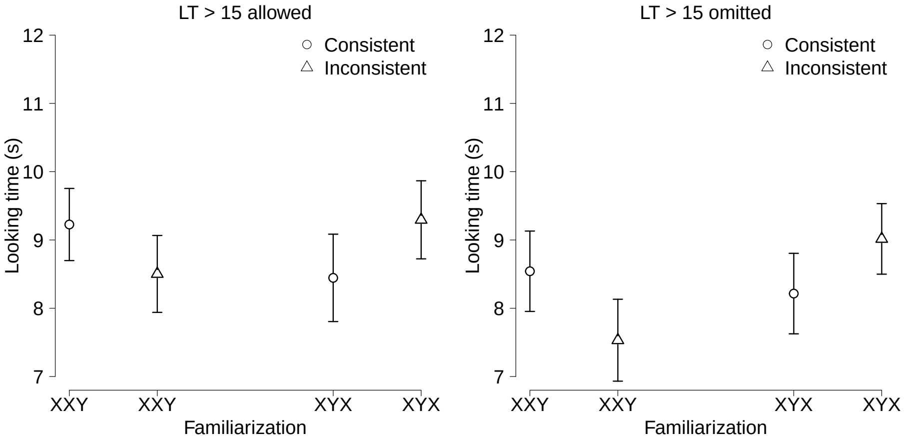


**Figure 2.6:** Interactions of familiarization and test grammars in Experiment 2.

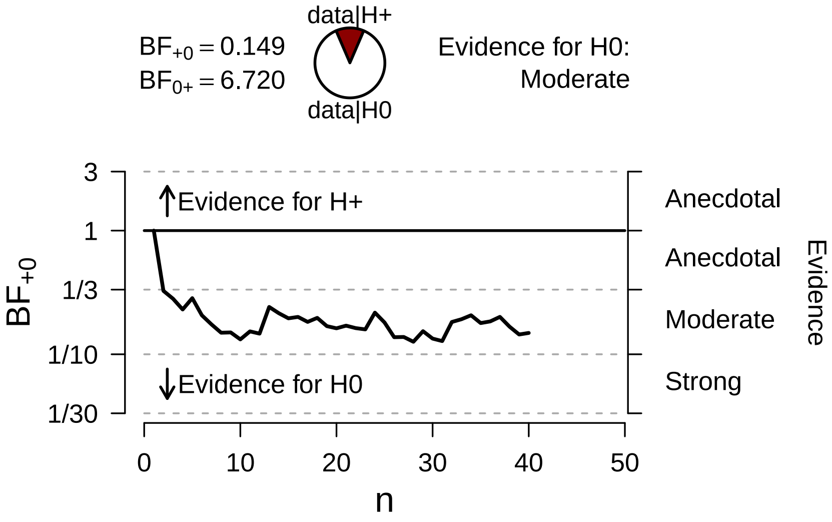


**Figure 2.7:** Sequential analyses for Experiment 2 with LTs longer than 15 seconds allowed.

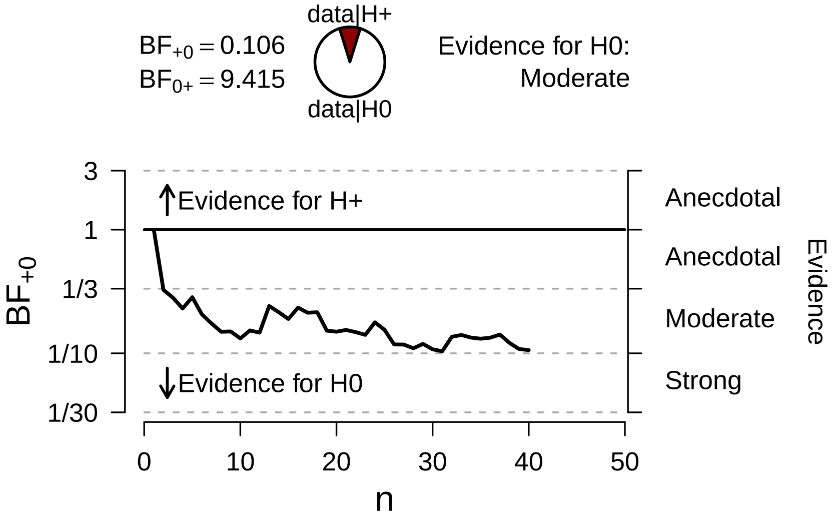


**Figure 2.8:** Sequential analyses for Experiment 2 with LTs longer than 15 seconds not allowed.

independent of which pattern they had been exposed to. While this is an interesting result consistent with the idea that infants may have a bias for adjacent repetitions, it does not replicate the results of Marcus et al. (1999) and does not show clear evidence of the ability to learn rules.

After failing to show rule learning in two experiments using the same familiarization phase and procedure, we identified several points on which our experiment differed from the original Marcus et al. (1999) study that may have negatively impacted the ability of infants to learn or to show their learning. First, our carefully controlled stimuli may have introduced a high level of variability which did not allow the infants to learn the rule quickly enough. This may have been due to the fact that our stimuli were phonologically complex as compared to those used in Marcus et al. The presence of more heterorganic syllable patterns in our stimuli might have attracted attention to the syllable structure rather than to the pattern.

A second reason for infants' failure to generalize may have been the fact that our patterns were constructed from uncategorized elements – any syllable could appear in any position, except for specifically the "F" syllable which changed position in Experiment 1, from appearing exclusively in the Y position during familiarization to the X position during test. We know from literature that categorization allows learners to attend to more than just the individual syllables and to form a rule. Braine (1987; summarized in Gerken, Wilson, &

Lewis, 2005) showed that the first step in solving an AGL task was to learn that there were categories in the grammar. Gómez and Gerken (2000) also argued that being able to make abstract categories is crucial for language learning and productivity.

In addition to the differences with respect to the stimuli, we used a different experimental paradigm. While VFP and HPP are based on the same principle of looking-while-listening, there may be differences with respect to how active the infant has to be in each task, the HPP being slightly more active (details below).

With these considerations in mind we ran a third experiment to try to replicate Marcus et al. (1999) as closely as possible, using similar stimuli and a similar procedure.

## 2.5 Experiment 3

### 2.5.1 Methods

#### Stimuli

In this experiment, the stimuli were similar to those used in Marcus et al.'s (1999) Experiments 2 and 3, with two small exceptions. As our infant participants were Dutch native speakers, a Dutch (rather than an American English) synthesizer was used to create the syllables (synthesized in Praat; Boersma & Weenink, 2007). In addition, the voice affricate /dʒ/ present in syllables *ji* and *je* does not exist in Dutch. Because there are no voice affricates in Dutch, these syllables were replaced with with voiced fricative /v/ to produce *vi* and *ve*. The triads used are shown in Table 3.

#### Participants

We tested 30 monolingual Dutch seven-month-old infants. Recruitment and compensation was identical to Experiments 1 and 2.

Infants were removed from analysis for the following reasons: technical errors (n=5), excessive fussiness (n=6), being born more than 3 weeks preterm (n=1), and parental decision to stop the experiment (n=2).

#### Procedure

Infants were tested in the HPP as in Marcus et al. (1999). As in Experiments 1 and 2, infants were placed on their caregiver's lap in a sound attenuated booth. The caregiver wore headphones playing non-rhythmical music from a female vocal artists to mask the auditory stimuli of the experiment. At a distance of 110 cm on each side of the room was one circular, red LED light (the side lights). In the center, directly in front of the infant at a distance of 110 cm was a circular, green LED light (the attention grabber). Directly below the attention

**Table 2.4:** Familiarization and test items following the XYY and XYX pattern respectively.

| XYY | XYX |
|---|---|
| **Familiarization** | |
| dedidi | dedide |
| deveve | devede |
| delili | delide |
| dewewe | dewede |
| vididi | vidivi |
| viveve | vivevi |
| vilili | vilivi |
| viwewe | viwevi |
| ledidi | ledile |
| leveve | levele |
| lelili | lelile |
| lewiwi | lewile |
| wididi | widiwi |
| wiveve | wivewi |
| wilili | wiliwi |
| wiwewe | wiwewi |
| **Test** | |
| bapopo | bapoba |
| kogaga | kogako |

grabber was a wide-angle video camera (Go Pro Hero 3), which allowed the experimenter to view and react to the infant's behavior and which recorded the experiment for off-line coding. Directly above the side lights were speakers, hidden below white sheets of fabric to minimize distraction away from the lights.

The experiment began with an attention grabber. As in the previous two experiments, when the infant focused on the attention grabber for two seconds, the familiarization phase began. During this phase, the familiarization stream, consisting of 16 triads repeated three times in different randomizations, played uninterrupted for approximately two minutes (1 min 46 sec). In addition, the side lights were illuminated when the infants looked at them and were extinguished when the infant looked away, in order to familiarize the infants with the fact that the lights were contingent on their attention. This was done independently of the auditory stream.

When the familiarization phase ended, the attention grabber played until the infants again looked towards it for a total of two seconds. Then the test phase began with one of the side lights flashing. The side of the first light was counterbalanced between infants. As soon as the infants turned their heads towards the flashing side light, the auditory test stimuli started to play. As in Experiments 1 and 2, the test stimuli would continue playing as long as the infants looked, for a maximum of 15 seconds. When the infants looked away for less than two seconds, the auditory stimuli would briefly pause, but would resume if the infants turned back within that time window. If the infants looked away for more than two seconds, the auditory stimuli and the blinking light would stop, the trial would end, and the attention grabber would begin again. This repeated until the infants finished the test phase, or until the caretaker or experimenter determined that an infant was too fussy (not orienting to attention grabber for an extended period of time) or distressed (crying or screaming while ignoring the lights). As in Experiment 2, the test phase consisted of four generalization test items, two corresponding to a pattern consistent with the familiarization and two corresponding to a pattern inconsistent with the familiarization. These four test items were repeated for three blocks, and randomized per block, totaling 12 test trials.

## 2.5.2 Results

### Data preprocessing

As this was a replication, we used the same procedure as Marcus et al. (1999), LTs shorter than the duration of the stimuli were omitted (3% of the data) and LTs longer than 15 seconds were kept (16% of the data).

### Frequentist analysis

A repeated measures ANOVA with consistency, training and their interaction on the mean looking times did not yield a significant main effect for consistency [$F(1,14) = 1.74$, $p = .208$)]. There was a near-significant interaction effect between training and condition [$F(1,14) = 4.05$, $p = .064$, see Figure 2.9]. Infants who were trained with XYX had a preference for inconsistent trials, whereas infants trained with XYY had a preference for consistent trials. This implies an overall preference for XYY over XYX, similar to the results of Experiments 1 and 2, where there was a preference for XXY over XYX.

### Bayesian analysis

As in the previous two experiments, we performed a sequential Bayesian paired samples t-test to quantify how much more likely the null hypothesis is than the alternative hypothesis that infants do look longer when inconsistent sound are played. Figure 2.10 shows how evidence accumulates in favor of the null hypothesis.

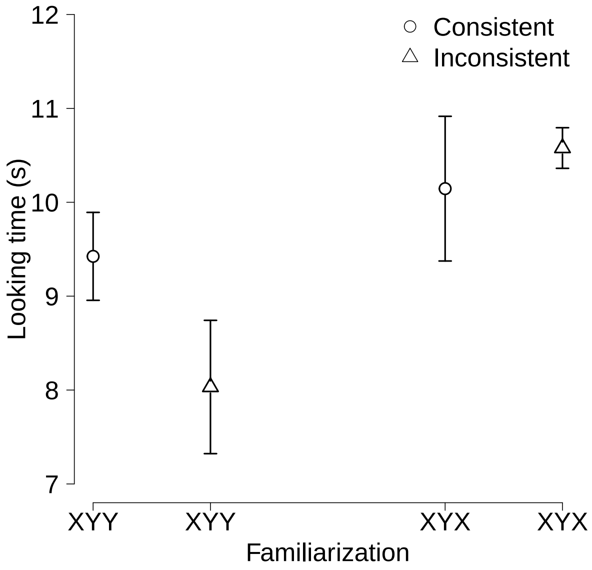


**Figure 2.9:** Interactions of familiarization and test grammars in Experiment 3.

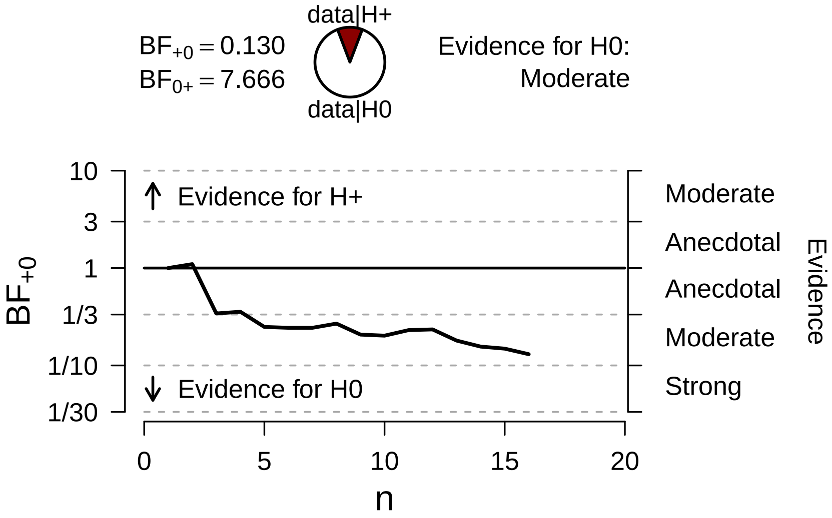


**Figure 2.10:** Sequential analyses for Experiment 3.

### 2.5.3 Discussion

Counter our expectations, infants in this experiment did not perform differently from those in the previous two experiments: they did not show differential looking based on consistency of test items with the familiarization grammar, indicating that they did not generalize the rule to novel input. They did, however, show longer looking to the test pattern containing adjacent repetition, the XYY pattern in this case. Despite using the similar syllables and triads, the same procedure, and the same age group as in the original Marcus et al. (1999) experiment, we could not replicate the results with Dutch infants.

## 2.6 Combination of the 3 experiments

Taken together, these three experiments seem to imply that infants do not have a preference for inconsistent over consistent patterns, but that they do have a preference for repeating patterns over alternating patterns. To illustrate these two points we performed a combined Bayesian analysis with the data of the three experiments combined. Using a Bayesian approach is especially well suited, as the outcomes can be framed as competing hypotheses, allowing us to combine data from multiple experiments.

### 2.6.1 Results

**Data preprocessing**

Although there are many possible options to preprocess the data, the multiverse analyses (Steegen et al., 2016) of Experiments 1 and 2 indicated most choices did not matter. Therefore, we kept the combined analyses close to the original Marcus et al. (1999) paper and use the same choices made in Experiment 3. That is, LTs longer than 15 seconds were kept while LTs shorter than the duration of one triad were omitted.

**Bayesian analysis - Overall**

Figure 2.11 shows how evidence accumulates in favor of the null hypothesis. Together the three experiments provide strong evidence in favor of the null hypothesis that infants have no preference for either consistent or inconsistent sounds. The data of 132 infants shows that the null hypothesis is approximately 14 times more likely than the alternative hypothesis that infants have a preference for inconsistent patterns.

**Bayesian analysis - Repetition**

For all of the three experiments we calculated the mean LTs during the test phase for the repeating patterns (XXY and XYY) and the non-repeating pattern (XYX). These LTs were then used to perform a sequential Bayesian paired

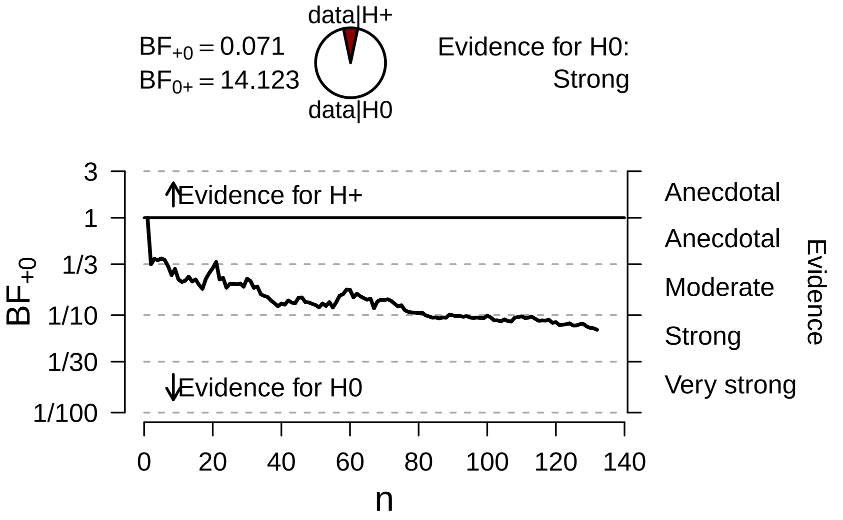


**Figure 2.11:** Sequential analyses for all three experiments showing evidence for the null hypothesis.

samples t-test. Figure 2.12 shows how evidence accumulates in favor of the repeating patterns hypothesis. Together the three experiments provide extremely strong evidence in favor of a preference for repeating over non-repeating patterns. The data of 134 infants shows that a preference for repeating patterns is 1177 times more likely than no preference or a preference for non repeating patterns.

## 2.7 General Discussion

Across three experiments with over 100 infants, we attempted to extend on, and eventually simply to replicate, the work of Marcus et al. (1999). The original study had shown that seven-month-old infants are able to learn simple XYX, XYY, and XXY rules on an abstract level, generalizing them to novel stimuli. In the original study, infants showed significantly more interest in test patterns that were inconsistent with the familiarization pattern than those that were consistent with it, independent of which familiarization pattern was used. In our experiments, Bayesian statistical analyses showed moderate evidence in favor of the null hypothesis in each experiment, indicating that infants did not discriminate the familiarized pattern from the novel pattern based on consistency. In Experiments 1 and 2, infants trained on XXY looked longer to the consistent XXY, while infants trained on XYX looked longer to the inconsis-

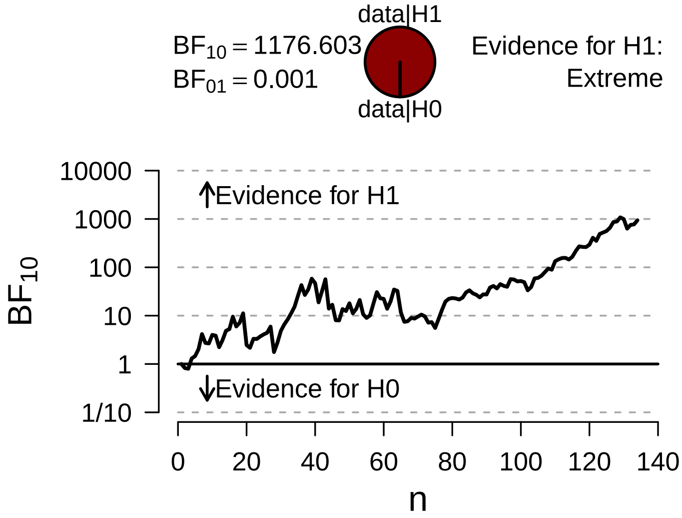


**Figure 2.12:** Sequential analyses for all three experiments showing evidence for the alternative hypothesis that infants look differently based on the presence of repetition.

tent XXY. Similarly, in Experiment 3, infants trained on XYY preferred the consistent XYY test pattern, while infants trained with XYX preferred the inconsistent XYY test pattern. We interpret the difference in looking towards the consistent and inconsistent patterns based on the familiarization pattern to indicate a general preference for patterns containing adjacent repetition.

Indeed, a Bayesian analysis of all the data confirmed that there is strong evidence for the null hypothesis that infants have no preference based on consistency of the test pattern with the familiarization pattern, and extremely strong evidence for the alternative hypothesis that infants are more likely to prefer adjacent repetition than to have no preference. The current set of studies is a good illustration of the practical benefits of a Bayesian analysis approach. First, we were able to quantify the plausibility of the null hypothesis that infants have no preference for either consistent or inconsistent sounds. Furthermore, the presented cumulative analyses show how incrementing the amount of data strengthens conclusions. The latter feature of Bayesian analysis introduces a new way of conducting infant experiments, which would entail testing until either the null-hypothesis or the alternative hypothesis gathers enough evidence. This would constitute a practical solution to the expensive and complicated practice of infant research.

### 2.7.1 Repetition biases

In our experiments, we found evidence for a repetition bias in infants. If infants have a general preference for repetition patterns, we might expect to find that they show this preference already during familiarization – infants trained on XXY might show longer looking times during familiarization than infants trained on XYX. Yet no such preference was found. One explanation may be that the preference is not found during familiarization simply because of the nature of the familiarization paradigm: the auditory stimulus continues independently of infant's interest in all three experiments such that even if they were not interested in the stimuli, infants had no incentive to modulate their behavior accordingly and show their preference or lack thereof. On the other hand, it might also be indicative that the preference simply does not manifest itself except in relation to the familiarization pattern. When presented with a reference point first, infants later prefer the test items that are similar or different based on the presence of repetition[6]. Such an explanation *would* require learning at some level, at least learning of the presence or absence of repetition. In either case, the conclusion we can draw from this work is not that speech is special but that repetition is special.

This type of bias for repetitions had been shown in previous literature but the preference has manifested itself in a different way. Most notably, Gervain et al. (2008) found that newborn infants' brains already react differently to the presence of repetitions in sound sequences than to the absence of repetitions. In addition, there is evidence that the presence of repetitions is favorable for rule learning. In the visual domain, Johnson et al. (2009) found that 11-month-old infants are able to learn XYY and XXY rules, but not XYX rules; eight-month-old infants were only able to learn the XYY rule. These findings show that there is a hierarchy of difficulty with respect to even these very simple patterns. Our results line up with these empirical findings on infants, with cross-species findings showing a bias for repetitions (Chen et al., 2015; van Heijningen et al., 2013), and with the theoretical arguments of Endress, Mehler, and colleagues who identify repetitions, particularly at edges, as one of the main biases that aids in learning structures and patterns across domains (Endress et al., 2009; Endress, 2013). While we could not conclude that algebraic rule learning was occurring in our experiments, the results provide evidence for a preferential status of repetitions in perception in the first months of life. From our results, we conclude that the ability to learn and generalize simple Marcus rules may be far more fragile than previously thought, even in the speech domain (whether natural or synthetic).

---

[6] A small pilot experiment in our lab conducted without a familiarization phase did not reveal a preference for one type of pattern over another either, adding support for this account.

### 2.7.2 Future work

The results raise several possible paths for future work. First, a possible reason for discrepancy between Experiment 3 and the original Marcus et al. (1999) study could be the difference in native language of the participants. In Experiments 1 and 2, we used natural Dutch syllables that did not produce real words. In Experiment 3, except for the replacement of one consonant, we used the stimuli from Marcus et al. (1999) produced by a Dutch synthesizer. The frequency distribution of the particular syllables used in the original study may have been different in English than in Dutch. On a related note, the similarity between phonemes and between syllables may be different in English than in Dutch. Simple perceptual cues and frequencies may play a role in the ability to notice different patterns, may modulate attention, and thus may influence learning ability during the familiarization phase. How similar or different individual items in a to-be-learned triad should be is worth exploring further in the quest to understand what conditions do and do not allow learning to occur. However, if learning outcomes indeed turn out to be highly dependent on these cues, it would show again that this type of rule learning is fragile, and one could question whether this type of algebraic rule learning is possible outside the lab at all.

This begs the question of whether these types of rule learning tasks are actually successfully tapping into learning abilities that we know infants have. The role of artificial grammar learning is to shed light on the cognitive building blocks that are at work in allowing for rule learning to occur in natural language. Yet we must ask whether these types of patterns are successful tools in researching these abilities. While the patterns XXY and XYY do occur as reduplication at different levels of language (word-internally, and within and across sentences), the pattern XYX in is more unrealistic as a representation of anything occurring in natural language at the word or sentence level[7,8]. In attempting to interpret the theoretical consequence of infants' failure to learn such unnatural patterns, we are left with the question of why infants cannot do something simple when we know they can do far more complicated computations and abstractions. Future work might move away from the use of these Marcus patterns and focus on more complex yet still simple structures from which clearer parallels with natural language can be drawn.

## Acknowledgements

We would like to thank all who helped in setting up appointments and collecting data: Marijn van 't Veer, Renske Jacobs, Rabia Mahboeb, Astrid Gilein,

---

[7] Examples of exceptions in English would be phrases such as *Dog eat dog*, *Dust to dust*, *Brother against brother*, etc. Idiomatic expressions and phrases aside, this remains a rarely found phenomenon.

[8] See also chapter 4 for an analysis of differences between English and Dutch at the level of the three-syllable patterns used here.

Mirjam Bahlmann, Isabella Jordanoska, Gijsbert Westland. We thank Reinhard Schillberg, Vincent Buurman, and Saskia Lensik for analysis advise along the way. We are very grateful to Carel ten Cate for comments on this paper, and to Michelle Spierings, Carel ten Cate, Raquel Garrido Alhama, and Jelle Zuidema for discussion and feedback throughout this project. The research reported here was supported by NWO Vrije Competitie grant 360.70.452 to Clara C. Levelt.

## 2.8 Appendices

### 2.8.1 Appendix A

**Table A1:** Each letter within the triads presented in Table 2 corresponded to a different syllable for each participant according to the following scheme.

| | Participant number | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Syllable label | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| A | do | fi | ka | le | mo | nu | pu | sa |
| B | fi | ka | le | mo | nu | pu | sa | do |
| C | ka | le | mo | nu | pu | sa | do | fi |
| D | le | mo | nu | pu | sa | do | fi | ka |
| E | mo | nu | pu | sa | do | fi | ka | le |
| F | nu | pu | sa | do | fi | ka | le | mo |
| G | pu | sa | do | fi | ka | le | mo | nu |
| H | sa | do | fi | ka | le | mo | nu | pu |

### 2.8.2 Appendix B

**Table B1:** Familiarization triads for Experiments 1 and 2, per grammar.

| | Familiarization Grammar XXY | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| AAD | do do le | fi fi mo | ka ka nu | le le pu | mo mo sa | nu nu do | pu pu fi | sa sa ka |
| AAE | do do mo | fi fi nu | ka ka pu | le le sa | mo mo do | nu nu fi | pu pu ka | sa sa le |
| AAF | do do nu | fi fi pu | ka ka sa | le le do | mo mo fi | nu nu ka | pu pu le | sa sa mo |
| BBA | fi fi do | ka ka fi | le le ka | mo mo le | nu nu mo | pu pu nu | sa sa pu | do do sa |
| BBC | fi fi ka | ka ka le | le le mo | mo mo nu | nu nu pu | pu pu sa | sa sa do | do do fi |
| BBF | fi fi nu | ka ka pu | le le sa | mo mo do | nu nu fi | pu pu ka | sa sa le | do do mo |
| CCA | ka ka do | le le fi | mo mo ka | nu nu le | pu pu mo | sa sa nu | do do pu | fi fi sa |
| CCD | ka ka le | le le mo | mo mo nu | nu nu pu | pu pu sa | sa sa do | do do fi | fi fi ka |
| CCF | ka ka nu | le le pu | mo mo sa | nu nu do | pu pu fi | sa sa ka | do do le | fi fi mo |
| DDB | le le fi | mo mo ka | nu nu le | pu pu mo | sa sa nu | do do pu | fi fi sa | ka ka do |
| DDE | le le mo | mo mo nu | nu nu pu | pu pu sa | sa sa do | do do fi | fi fi ka | ka ka le |
| DDF | le le nu | mo mo pu | nu nu sa | pu pu do | sa sa fi | do do ka | fi fi le | ka ka mo |
| EEB | mo mo fi | nu nu ka | pu pu le | sa sa mo | do do nu | fi fi pu | ka ka sa | le le do |
| EEC | mo mo ka | nu nu le | pu pu mo | sa sa nu | do do pu | fi fi sa | ka ka do | le le fi |
| EEF | mo mo nu | nu nu pu | pu pu sa | sa sa do | do do fi | fi fi ka | ka ka le | le le mo |

| | Familiarization Grammar XYX | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| ADA | do le do | fi mo fi | ka nu ka | le pu le | mo sa mo | nu do nu | pu fi pu | sa ka sa |
| AEA | do mo do | fi nu fi | ka pu ka | le sa le | mo do mo | nu fi nu | pu ka pu | sa le sa |
| AFA | do nu do | fi pu fi | ka sa ka | le do le | mo fi mo | nu ka nu | pu le pu | sa mo sa |
| BAB | fi do fi | ka fi ka | le ka le | mo le mo | nu mo nu | pu nu pu | sa pu sa | do sa do |
| BCB | fi ka fi | ka le ka | le mo le | mo nu mo | nu pu nu | pu sa pu | sa do sa | do fi do |
| BFB | fi nu fi | ka pu ka | le sa le | mo do mo | nu fi nu | pu ka pu | sa le sa | do mo do |
| CAC | ka do ka | le fi le | mo ka mo | nu le nu | pu mo pu | sa nu sa | do pu do | fi sa fi |
| CDC | ka le ka | le mo le | mo nu mo | nu pu nu | pu sa pu | sa do sa | do fi do | fi ka fi |
| CFC | ka nu ka | le pu le | mo sa mo | nu do nu | pu fi pu | sa ka sa | do le do | fi mo fi |
| DBD | le fi le | mo ka mo | nu le nu | pu mo pu | sa nu sa | do pu do | fi sa fi | ka do ka |
| DED | le mo le | mo nu mo | nu pu nu | pu sa pu | sa do sa | do fi do | fi ka fi | ka le ka |
| DFD | le nu le | mo pu mo | nu sa nu | pu do pu | sa fi sa | do ka do | fi le fi | ka mo ka |
| EBE | mo fi mo | nu ka nu | pu le pu | sa mo sa | do nu do | fi pu fi | ka sa ka | le do le |
| ECE | mo ka mo | nu le nu | pu mo pu | sa nu sa | do pu do | fi sa fi | ka do ka | le fi le |
| EFE | mo nu mo | nu pu nu | pu sa pu | sa do sa | do fi do | fi ka fi | ka le ka | le mo le |

**Table B2:** Test triads. For Experiment 1, all items were used, while in Experiment 2 only triads labeled "Generalization" were used.

| | | Test Phase | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Combination | AAC | do do ka | fi fi le | ka ka mo | le le nu | mo mo pu | nu nu sa | pu pu do | sa sa fi |
| | ACA | do ka do | fi le fi | ka mo ka | le nu le | mo pu mo | nu sa nu | pu do pu | sa fi sa |
| | BBD | fi fi le | ka ka mo | le le nu | mo mo pu | nu nu sa | pu pu do | sa sa fi | do do ka |
| | BDB | fi le fi | ka mo ka | le nu le | mo pu mo | nu sa nu | pu do pu | sa fi sa | do ka do |
| | CCE | ka ka mo | le le nu | mo mo pu | nu nu sa | pu pu do | sa sa fi | do do ka | fi fi le |
| | CEC | ka mo ka | le nu le | mo pu mo | nu sa nu | pu do pu | sa fi sa | do ka do | fi le fi |
| | DDC | le le ka | mo mo le | nu nu mo | pu pu nu | sa sa pu | do do sa | fi fi do | ka ka fi |
| | DCD | le ka le | mo le mo | nu mo nu | pu nu pu | sa pu sa | do sa do | fi do fi | ka fi ka |
| | EEA | mo mo do | nu nu fi | pu pu ka | sa sa le | do do mo | fi fi nu | ka ka pu | le le sa |
| | EAE | mo do mo | nu fi nu | pu ka pu | sa le sa | do mo do | fi nu fi | ka pu ka | le sa le |
| Position | FFA | nu nu do | pu pu fi | sa sa ka | do do le | fi fi mo | ka ka nu | le le pu | mo mo sa |
| | FAF | nu do nu | pu fi pu | sa ka sa | do le do | fi mo fi | ka nu ka | le pu le | mo sa mo |
| | FFE | nu nu mo | pu pu nu | sa sa pu | do do sa | fi fi do | ka ka fi | le le ka | mo mo le |
| | FEF | nu mo nu | pu nu pu | sa pu sa | do sa do | fi do fi | ka fi ka | le ka le | mo le mo |
| Generalization | GGH | pu pu sa | sa sa do | do do fi | fi fi ka | ka ka le | le le mo | mo mo nu | nu nu pu |
| | GHG | pu sa pu | sa do sa | do fi do | fi ka fi | ka le ka | le mo le | mo nu mo | nu pu nu |
| | HHG | sa sa pu | do do sa | fi fi do | ka ka fi | le le ka | mo mo le | nu nu mo | pu pu nu |
| | HGH | sa pu sa | do sa do | fi do fi | ka fi ka | le ka le | mo le mo | nu mo nu | pu nu pu |