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

Geambasu, A.

**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 7

# Conclusion

## 7.1 Thesis summary

The goal of this thesis was to gain new insights into the rule learning capabilities of humans, both infants and adults.

In chapter 2, we studied the ability of infants between the ages of six and nine months to learn grammars similar to the ones used in Marcus et al. (1999). In Experiments 1 and 2 of that chapter, where we used natural speech stimuli, we found no evidence of discrimination between test grammars based on their consistency with the familiarization grammar, and no age-related differences between six- and nine-month-olds. In Experiment 1, we also tested a group of infants using birdsong, with no difference in results. Instead, what we found in both experiments was an overall preference for adjacent repetition items (XXY) during the test, independent of training grammar. In Experiment 3, in an attempted close replication of the original Marcus et al. experiment, we found that seven-months-old performed in a similar manner, preferring the XYY grammar during the test, regardless of whether they were familiarized with XYY or XYX. In both cases, a Bayesian analysis strengthened our findings and allowed us to conclude that infants across the three experiments showed strong evidence of having an adjacent repetition preference.

In chapter 3, we turned to the visual domain. In that chapter we attempted to extend on a study from Saffran et al. (2007), who had shown that infants at seven months were able to generalize Marcus rules in the visual domain when the grammars were composed of pictures of dogs or of cats of different breeds. While their conclusion was that success in this experiment was due to infants' familiarity with the stimuli and their ability to categorize it, we hypothesized that infants may have been successful in that task due to the fact that they probably had a lexical representation for dogs or cats (or similar animate objects) at this age. This would entail that rule learning is still facilitated by linguistic cues. We tested whether there would be a difference between learning Marcus grammars in the visual domain when the visual stimuli consisted of (1) items that 12- to14-month-old infants knew the names for, or (2) of items that were familiar to them but not lexically specified, or (3) of nonsense objects. While we found that they looked longer at stimuli consisting of items they knew the names for, this did not translate to success in learning about the underlying grammatical rules in all but boys familiarized with XYY in the Known condition. This limited finding should be interpreted with caution. While we have theoretical and empirical reasons to trust that there would be differences between groups related to their sex (e.g., Shi, 2007; Mueller et al., 2012; Lany & Gómez, 2008; ter Haar & Levelt, 2018) and training grammar (e.g., S. P. Johnson et al., 2009; Endress et al., 2009), and although the effect was strong, the number of participants was low and the effect should be replicated to confirm the finding. Nevertheless, the work in this chapter suggests that item familiarity in itself does not guarantee generalization of the Marcus rules to the visual domain. A factor that could have contributed to the success in Saffran et al. (2007), which could be tested in future work, is the fact that dogs and cats are animate, which may make them more interesting stimuli for infants of this age to learn from. Neither the abstract shapes in Johnson et al. (2007), nor the pictured items in our experiments were animate.

In chapter 4, we looked at production in infants, seeking to better understand the types of patterns present in their early babbling phase as a means to contextualize the ability (or inability) to learn certain patterns in perceptual tasks. Analysis of more than 60,000 utterances in nine languages from children in their first two years of life showed that contrary to established theories of language production development, completely variegated utterances (i.e., following patterns like XY, XYZ, etc., where the variables represent CV syllables) were present from the earliest utterances and were the most frequent type of utterances across the languages. Of the utterances containing repetition, those containing immediate repetition (or reduplication; XXY and XYY) were far more frequent than those containing non-adjacent repetition (XYX). These results may seem at first glance to contradict our experimental findings, with variegated productions being the most produced, but immediately repeating ones attracting the most attention, i.e., longer looking times. However, repetition has been found to be a learning tool in production, and as such, it follows that repeated utterances attract attention. Utterances such as XYX may be in an unhappy medium, preferred neither in perception nor in production due to the memory constraints necessary both to perceive and to produce them.

Chapters 5 and 6 focused on adult rule learning, using Marcus grammars and the more complex Lindenmayer grammars respectively. Studying adults allowed us to gain insights into factors potentially influencing rule learning that we could not get from infant participants. Our adult studies allowed us to manipulate the task and to look at both implicit and explicit factors. Chapter 5 revealed that Marcus-type rule learning does not proceed unimpeded in adults; they, like infants, can only learn under specific circumstances. Adults could not generalize the simple rules when instructions were too vague. Instead, they benefitted from explicit instructions, from an active learning task with feedback, or from specific test stimuli that could only be processed when generalizing, directing them to the level at which they should attend. On the other hand, chapter 6 revealed that adult participants are able to pick up on regularities in far more complex stimuli, involving hierarchical structures. In both an identification task (Yes/No) and an identification-discrimination task (2AFC), participants showed sensitivity to an L-system grammar instantiated with different types of drum sounds.

The sections that follow will more explicitly expand on these findings by categorizing them according to the themes identified in chapter 1.

## 7.2 Domain specificity

While the results of chapters 2 and 3 do show a difficulty to learn the Marcus-type rules, they also lend support to the notion that at least some of the biases and processing mechanisms involved in learning rules are not specific to language. In chapter 2, we tested infants in the auditory domain, using both speech and birdsong stimuli and found no differences between the performance in the two domains. While no general rule learning was found in either condition, infants in both conditions showed the same type of preference in the test phase, namely for test stimuli containing adjacent repetitions. This finding indicates that independent of domain, a bias for repetitions underlies infants' general processing of auditory structures. In the visual rule learning tasks of chapter 3, we found that at least a subset of participants were able to generalize the learned rule, namely male infants familiarized with XYY grammars in the condition in which they had lexical representations for the items in the pictures. While this result is tentative and should be replicated, it adds support to the idea that infants are able to learn rules from more domain-general visual stimuli, but perhaps only when facilitated by a repetition pattern and the availability of a lexical representation of the pictured objects.

In chapter 6, where we tested adult participants using drum sounds rather than speech, participants showed sensitivity to the L-system grammar as compared to a foil grammar with similar surface properties. Although we did not explicitly test domain specificity within our task, comparing our results to those found by Shirley in the speech domain can give us an indication of adults' ability to learn L-systems within and outside the speech domain. Shirley's results were more robust; we can compare the mean correctness scores of participants in her study using speech stimuli and ours using drum sounds in the same type of task (2AFC), and see that participants in her speech task performed with an accuracy rate of nearly 20% higher than participants in our chapter 6. The discrepancy in results adds evidence to the claim that speech has a facilitatory role in aiding rule learning. Nevertheless, our participants did perform above chance level, also suggesting that they did become sensitive to the grammars. One explanation for the differences between chapters 1 through 5 and of chapter 6 may be that with simpler grammars, few differences can be discerned between domains; with more complex grammars, speech may hold an advantage - unsurprising since this is the domain in which humans do perform complex rule learning and production.

In addition, it is relevant to return to the original AGL studies (such as those performed by Reber, 1967, 1976) in discussing this notion of domain specificity or generality. Note that while these first studies on AGL did not focus on domain generality *per se* and were meant to illuminate mechanisms of language acquisition, they were, in fact, instantiated not using spoken or signed language, but with visually-presented letters on a screen whose order of presentation was generated by an underlying grammar. This method of presentation already removes domain-specific learning mechanisms from consideration, as the linguistic value of a string of orthographically presented consonants is questionable. The use of this type of stimuli continues in many AGL tasks with adults (e.g., Kinder & Assmann, 2000; Johnstone & Shanks, 1999; see also Perruchet & Pacton, 2006 for an overview).

The potential to learn patterns across domains corroborates cross-species findings as well. Birds, for example, are able to pick up regularities at some level in auditory stimuli: zebra finches at the level of the sequential ordering of items within a triad, budgerigars even at the more abstract level of rules (Spierings & ten Cate, 2016). Rule learning abilities that humans use for language acquisition may thus have roots in these more domain-general and species-general abilities to process and categorize information. On the other hand, similar outcomes do not necessarily indicate similar processes (Milne, Wilson, & Christiansen, 2018), and how rule learning proceeds online in each domain should be further explained. Nevertheless, there may be shared biases that aid rule learning across domains. One of these mechanisms from which learning seems to benefit is the perceptual sensitivity to repetitions, which has been found across ages, domains, and species (Endress & Bonatti, 2007; Endress et al., 2009; van Heijningen et al., 2013; Chen et al., 2015).

## 7.3 Repetitions

In chapter 2, and to some extent also in chapter 3, we found evidence for an asymmetry in performance after exposure to the types of Marcus grammars containing immediate repetitions (XYY and XXY) and those containing non-adjacent repetitions (XYX). Chapter 4, which looked at the patterns found in infants' earliest utterances, attempts to link these perception results with what infants do around the same period in production. The main finding was that infants produce significantly more variegated utterances (like XYZ) than non-variegated ones (like XYY, XXY or XYX), and consistently so throughout their first year. The secondary finding is that they produce more non-variegated utterances containing immediate repetitions (XYY, XXY) than non-adjacent repetitions (XYX). How can we reconcile these findings and how do they help inform us about what is happening in our perception experiments?

Repetition in perception has long been identified as an important factor supporting learning. Gómez et al. (2000) identify repetition as the basis of the ability to generalize (or transfer) rules. While Gomez et al. addressed learning and generalization of a finite state grammar, the finding is relevant for rule learning in general. Repetition is highly salient for humans from the beginning of life (Gervain et al., 2008), and is relevant for learning across domains (S. P. Johnson et al., 2009).

The role of repetition in production is linked to learning too. Based also on corpus results, Lipkind et al. (2013) argued that repetitions in production serve a memory function, helping to solidify the memory of a newly-acquired syllable in the infant's inventory. Experimental work from Fagan (2015) also supports this notion of the role of repetition in infants' productions. Observations of normally-hearing infants and those learning language after receiving a cochlear implant indicated that repetitions in production emerged as vocal exercises, strengthening the perceptual memory of new phonemes and syllables.

Returning to rule learning, we can see that if syllable repetition is a learning tool for infants, both in perception and in production (and its interaction) it follows that the rules containing immediate repetition in the Marcus et al. (1999) paradigm would facilitate learning in that task too.

Where does this leave seemingly non-preferred utterances such as XYX containing non-adjacent repetitions? Gervain et al. (2008) found asymmetries between discrimination abilities of newborns tasked with discriminating XXY vs. XYZ and XYX vs. XYZ. While the former produce evidence of discrimination abilities, the latter did not. The fact that in perception XYX and XYZ are not discriminated, and that in production XYX is so rarely produced in comparison to all other utterances, may indicate these are perceived (and produced, perhaps accidentally) as a subset of the variegated XYZ.

Turning to adults, chapter 5 and 6, show that repetition is not the only basis for abstraction and that perhaps adults are less sensitive to repetition than infants. In chapter 5, we found no evidence of asymmetry in performance between adults familiarized with repetition- versus non-repetition grammars. In addition, in chapter 6, using a far more complex grammar, we found that participants were able to correctly identify and discriminate the grammatical from ungrammatical strings even though both streams shared a similar surface distribution of immediate repetition bigrams. In order to discriminate this similar non-grammatical foil from that following an L-system, adults in this task must have learned far more about the grammar than the fact that it includes repetition. Even if participants in this task were to use a strategy that has an expectation of repetition, such a strategy would still require an anticipation of where the repetition can and cannot be in order to correctly identify and discriminate the correct grammar from the incorrect one. Therefore, an implicit understanding of the structure must have been formed. Thus, while repetition may be helpful and salient throughout development, and may aid in learning, the bias for simple repetition grammars found in our infants can be overcome. In the case of our adults, this can presumably come as a result of a number of cues, several of which are discussed further in section 7.5.

## 7.4 Familiarity

The work in chapter 3 attempted to clarify the level at which stimuli had to be familiar in order to facilitate rule learning. Saffran et al. (2007) claimed that their success in showing rule learning in the visual domain and the failures and asymmetries found in other visual rule learning tasks (e.g., S. P. Johnson et al., 2009) was a result of the level of the infants' familiarity with the objects presented (high for dogs in Saffran et al., 2007 and low for shapes in S. P. Johnson et al., 2009). However, as mentioned in chapter 1, familiarity may be at the level of the pictured item itself or at the level of the item's lexical representation. In chapter 3, while we did find that infants were significantly more attentive to items with which they had lexical familiarity, this behavior did not translate to better overall learning. While we found some evidence of learning in the condition in which participants knew the picture labels, overall performance in that condition was not strongly indicative of learning, since it occurred in only one sex (male), trained with one grammar (XYY). This result needs to be replicated with more participants in order to draw stronger conclusions. What can be said, however, is that familiarity at the item level did not seem to facilitate generalization; compared to nonsense items, which could not have been familiar, infants did not perform better with familiar stimuli.

The findings of chapter 2 also support the conclusion that item familiarity does not make a clear difference in learning performance. Infants in this case performed similarly, whether exposed to speech or to birdsong stimuli, the latter of which they would not be highly familiar with.

The argument for familiarity being beneficial for generalization, therefore does not find support in the work presented here.

## 7.5 Ecological validity

While it is surprising that infants struggle to learn simple Marcus-type grammars when they can learn the complex grammar of natural language, it is of note that the tasks infants are asked to perform in rule learning studies are not representative of the natural language or natural learning environments in which either infants or adults actually learn. While infants are able to learn a great deal implicitly, their learning environment is not only implicit but, also includes explicit feedback, repetitions of words and utterances, exaggerated intonation contours that direct attention to specific structures, and positive (verbal and non-verbal) reinforcement from interlocutors. Adult learning of natural language (e.g., a foreign language), on the other hand, is often not implicit but, depending on the learner, often benefits from varied formal instruction and feedback. While AGL tasks do not pretend to be representative of natural learning circumstances, the end goal of performing such experiments is to be able to tap into learning abilities used in natural grammar learning. Yet if we show multiple failures of "simple" learning, we fail at that end goal, and the ecological validity of such tasks can be questioned on multiple fronts.

### 7.5.1 Variability

One way in which we attempted to make our rule learning tasks more ecologically valid was to manipulate the amount of variability in the input. Previous Marcus-type rule learning experiments have typically been limited with respect to the phonological variability of the auditory stimuli. In Experiments 1 and 2 of chapter 2 with infants, and in chapter 5 with adults, we used a set of syllables that were phonologically balanced to be more representative of the natural language to which participants are normally exposed. When comparing the results of Experiments 1 and 2 of chapter 2 with those of Experiment 3, where infants heard the low-variability Marcus et al. stimuli, we found no difference in performance. Increasing the variability of the phonemes and syllables to make them more naturalistic did not produce any advantage in learning. In addition, in chapter 5, we manipulated the amount of variability in the stream by varying the number of triads presented to the participant: either 15 triplets repeated three times or three triplets repeated 15 times. We found that while learning was still successful when participants were exposed to three triplets, a boost was seen as a result of being exposed to 15 phonologically different triplets.

However, the fact remains that learning in adults was indeed still possible beyond chance level with a low-variability stream, and infants did not receive any boost in performance whether they were exposed to our phonologically well-balanced set of syllables and triplets versus those used in Marcus et al.'s (1999) original study. Together these results from infants and adults indicate that variability may play a role in learnability, but it is probably not a key factor. This finding fits well with results from Gerken (2010; 2015) that indicate that infants are able to learn a rule from a handful of stimuli, or even from just one stimulus, given the correct circumstances.

### 7.5.2 Task and instruction

In chapter 5, we systematically varied task and instruction in a series of Marcus-type rule learning tasks with adults. When presented with vague instructions, resulting in a passive learning task in which they were unaware of what level of the stimulus stream they should attend to, adults struggled to generalize the underlying rule to novel stimuli. This passive learning task was arguably the most similar to the experimental conditions that infants undergo, in which no instruction can be given and nothing in particular directs their attention to anything specific about the stimuli. On the other hand, adults succeeded in the tasks when their attention was directed to the relevant level of information (the grammar) by manipulations in the instruction or in the type of task presented. An active, Go-Left/Go-Right task, analogous to the Go/No-Go task animals typically perform, yielded highly successful results.

However, in our L-systems experiments, task did not seem to make a difference in participants' ability to discriminate grammatical from non-grammatical strings. In this case, there were no differences between a 2AFC task and a Yes/No tasks overall : participants in both cases identified the correct strings at a rate higher than would be expected by chance. In this case, instructions were more explicit than those provided during some of the experiments in chapter 5, calling attention to the fact that a rhythmic pattern would be heard and need to be classified later in test. Yet the tasks were similar to those in chapter 5, consisting of an identification task and a discrimination task. Participants here performed above chance in both tasks. And since each participant performed both tasks, we were able to study the influence of the first on the second task. Participants performed better on the 2AFC task when it was performed first, although there was no effect of order of tasks on performance in the Yes/No task. These task differences give us an indication that identification tasks may be more robust while performance on discrimination tasks may become weaker over time after exposure to a different type of task.

Like the adults in the passive learning task, infants struggled to learn simple grammars as well; whether their difficulty was a result of the task or not is still an empirical question. However when comparing the complex structures that infants can make sense of in the natural world with what they are unable to do in the lab, the question arises of whether we give infants a fair chance to show us what they can do. In an experimental environment, as we have seen above, adults need instruction or feedback to learn. Animals also receive feedback in AGL tasks, both positive (e.g., a food reward) and negative (e.g., a short period of darkness), and they are typically trained until they reach a certain criterion of performance (e.g., Spierings & ten Cate, 2016). Infants in the real world also receive feedback from their caregivers, both verbal and non-verbal. While short passive exposure tasks might be sufficient for simple discrimination studies (e.g., Maye et al., 2002; Maye, Weiss, & Aslin, 2008), the lack of feedback in the rule learning experiments might thus impede infants' learning, not allowing them to show their full potential in grammar learning tasks. Alternatively, if infants are able to learn the grammar of a natural language implicitly, but not a Marcus grammar, this could imply that the learning of Marcus grammars actually does not model natural language learning very well.

In addition, there is ambiguity in interpreting infant looking times in the sense that we are not always sure what it means with respect to the individual infant, the stimuli, and the task as a whole (Kidd et al., 2010). There are many hidden variables that can contribute to this one dependent measure (Aslin, 2007). The preference that we measure is a proxy for discrimination, but an infant may be able to perceive a difference between two stimuli, while preferring neither of the two. This may be why no preference between XYX and XXY grammars was found when pilot-testing a pure preference task, presenting infants with either XYX or XXY stimuli without a preceding familiarization phase (see discussion of chapter 2). Their bias for the repetition grammar thus emerges with respect to what they heard during familiarization. Research has shown that infants' preferences are volatile, and that the stimuli they are provided with must contain the exact right amount of information to draw and hold their attention (the "Goldilocks effect"; Kidd et al., 2010).

### 7.5.3 Grammar complexity

In chapters 1 and 2, we discuss the notion that while immediate repetition is found throughout the world's languages at the word and phrase level, the XYX rule is not typically found in natural language. This lack of naturalness in the stimuli is reflected in the fact that, as mentioned above, this grammar was not learned or preferred by infants in chapters 2 and 3, nor was it readily learned by adults in chapter 5 without certain task, instruction, or attention conditions being met. This leads us to the conclusion that these grammars are not well suited for our greater understanding of rule learning in infants or in adults. And it should not be surprising, since they do not reflect much about the task that we are attempting to explore.

In chapter 6 we experimented with a far more complex grammar, generated by simple recursive rules. The work in this chapter gives us an indication that we may be more successful in better understanding how participants learn about language and structure by employing artificial grammars that are more ecologically valid. The L-systems that were explored in chapter 6 have not been widely used in AGL work up until now. The task is difficult and outside of the linguistic domain, making use of drum sounds rather than speech. Yet participants do show some evidence of sensitivity to the difference between grammatical and ungrammatical strings during test at a rate higher than would be expected by chance, in both an identification and a discrimination-identification task. While these results may be explained by a more superficial error detection mechanism rather than a process of learning about the underlying structure of the grammar as a whole, and while they are not evidence of generalization to novel stimuli or to a novel domain, they do indicate that there is some mechanism within participants that presumably allows them to pick up on some level of regularity even in this complex stimuli. While more exploration is clearly needed, it is likely the route of more complexity that will provide us with more interesting and applicable results. The fact that this grammar has rewrite rules that allow for recursive patterns brings it much closer to natural languages than what is used in the simple rule learning experiments typically used in infant research (e.g., Marcus et al., 1999; Gómez, 2002). Future experiments with these grammars may attempt to use linguistic stimuli to study whether infants are as sensitive to these complex grammars as adults were in the work of Shirley (2014).

## 7.6 Conclusion

The work presented in this thesis has shown that generalization is not easy, and simple rule learning is not so simple. Work on Marcus-type rule learning has proved difficult to replicate, and failures to replicate are, in turn, difficult to interpret. Perhaps these grammars are not the best model grammars to use for understanding rule learning in natural language, or for understanding how or why humans pick up on regularities in the input. If these experiments are difficult to replicate, if they have small effect sizes when they do succeed, or if successes only occur under specific circumstances, how can we identify the cognitive mechanisms humans use to learn the structure of natural language? AGL should be a reliable tool for investigating how we learn natural language, and when infants fail to learn in an AGL experiment, we want to be sure it is because of their cognitive mechanism and not because of the method we are using.

Future studies into rule learning capabilities should address several of the issues raised in this thesis. First, artificial grammars used in these studies can and should be more complex and naturalistic. In addition, to get away from some of the methodological pitfalls identified above, future work could focus more on online measurements rather than the indirect measures such as the looking times typically collected from infants and the Yes/No-type responses typically collected from adults. In addition to EEG and NIRS studies, this could include online tasks involving serial reaction times (SRT), pupil dilation, or recall tasks (e.g., statistically-induced chunking recall, Isbilen, McCauley, Kidd, & Christiansen, 2017). Online measurements allow us to gain further insights into what participants find to be informative or surprising stimuli, and into the time-course of processing, which would inform us about the difficulty of the stimuli or task.

Finally, future work should include Bayesian analyses for better understanding of the data, particularly for understanding null results. Along with tools such as meta-analyses of experimental data and corpus analyses of production data, such analyses of null results and failures can help reshape the picture we have of how learning occurs in infants as well as in adults.