Infants’ Rule Generalization Is Governed by the Tolerance Principle

Emeryse Emond and Rushen Shi

1. Introduction

A challenging task in child language acquisition is the abstraction of rules from exemplars present in the input and the generalization of rules to novel instances. This task is further complicated by the fact that rules often have exceptions, such as the irregular *go-went* case for the past tense rule of verbs in English. Exceptions are not productive and must be memorized. In this study we examine the effect of the distribution of rule exemplars and exceptions in the input on rule generalization.

Rule learning when the input contains exemplars all conforming to rules (i.e., no exception) has been demonstrated in various experimental studies. For instance, seven- and nine-month-olds were able to learn algebraic-like rules (Marcus, Vijayan, Bandi Rao & Vishton, 1999; Gerken, 2006). In those studies infants were trained with multiple sequences of three syllables that were combined according to a rule: ABB (e.g., *le di di*) for one group of infants and ABA (e.g., *le di le*) for another group. After the training, infants were tested with novel instances composed of all new syllables that had not appeared during the training. These new instances either followed the trained pattern (e.g., ABB for the group that had been trained with this rule) or exhibited a non-trained rule (e.g., ABA for the group that had been trained with ABB). Their results showed that infants discriminated the two patterns in the test trials, showing evidence of abstracting the trained rule and generalizing it to novel instances.

Infants are also able to learn abstract grammatical category rules. In Gómez and Lakusta (2004) 12-month-olds were trained with instances of an artificial grammar such as aXbY, in which X were monosyllabic words (e.g., *deech*) and Y bisyllabic words (e.g., *coomo*). X words were always preceded by an ‘a’ word (either *alt* or *ush*), and Y words were always preceded by a ‘b’ word (either *ong* or *erd*). In the test phase, a and b words were the same as those in the training, but the X and Y words were new monosyllabic and bisyllabic items. The results of the test phase showed that infants discriminated novel rule instances (e.g.,

* Emeryse Emond and Rushen Shi, Université du Québec à Montréal, emond.emeryse@courrier.uqam.ca; shi.rushen@uqam.ca. We thank Charles Yang for insightful discussions and suggestions. This research was supported by NSERC and CFI grants to Rushen Shi.

 Proceedings of the 45th annual Boston University Conference on Language Development, ed. Danielle Dionne and Lee-Ann Vidal Covas, 191-204. Somerville, MA: Cascadilla Press.

new X and Y words co-occurring with a and b words according to the trained aXbY rule) from non-rule instances (e.g., new X and Y words co-occurring incorrectly with a and b words according to the non-trained aYbX pattern).

By 14 months of age, infants can also learn abstract word order rules. In Koulaguina and Shi (2013) the authors used a natural language (i.e., Russian) that was unknown to their participants to construct artificial word order movement rules. Infants were trained with multiple three-word sentences that shifted in word order, that is, ABC to BAC for one group of infants (e.g., *Dozhd' zalil cherdak → Zalil dozhd' cherdak*) and ABC to ACB (e.g., *Dozhd' zalil cherdak → Dozhd' cherdak zalil*) for another group. In the test phase infants discriminated new sentences that followed the trained rule (e.g., ABC→BAC for the group that had been trained with this rule) from those following the other, non-trained rule (e.g., ABC→ACB for the group that had been trained with the ABC→BAC rule). Both groups of infants listened longer to the non-trained rule than the trained rule in the test trials.

The above experiments showed rule generalization when the training input consisted of rule exemplars only. However, most linguistic rules contain exceptions, as mentioned earlier. A few experimental studies have tested infants' rule generalization when the input is inconsistent (i.e., containing exceptions) and they showed that rule generalization can be successful in the presence of certain proportions of exceptions.

Specifically, in the same study of Gómez and Lakusta (2004) described before, the proportions of exceptions in the training exemplars were manipulated in several experiments. The authors used the same artificial rule (e.g. aXbY), and the exceptions, which had a different pattern (e.g., aYbX), were set at a particular proportion relative to rule exemplars. They showed that in an experiment in which the number of exceptions was too high (at 33%, i.e., eight exceptions) relative to rule cases (16 rule exemplars), 12-month-old infants failed to generalize the rule to novel instances. However, a smaller number of exceptions in the input (17%, i.e., four exceptions, relative to 20 rule exemplars) did not prevent rule generalization to novel instances.

Similarly, Koulaguina and Shi (2019) tested the impact of inconsistent input on 14-month-olds' generalization of word order movement rules in multiple experiments. As in their earlier study (Koulaguina & Shi, 2013), their 2019 study used the same natural language (i.e., Russian) that was unknown to their participants. They constructed two artificial word-order rules with three-word sentences. One group of infants was trained with sentences shifting according to the ABC→BAC rule, and another group was trained with the sentences doing the ABC→ACB movement rule. For each group, the exceptions were sentences not going through any movement (i.e., with only the base order ABC). For the test phase, all infants heard trials of new sentences going through the ABC→BAC rule versus the ABC→ACB rule. The two groups were thus controlling for each other; that is, the trained rule for one group was the non-trained rule for the other group, and vice-versa. Moreover, there were two manipulations: in two experiments the level of non-rule exemplars (i.e., the nonmovement cases) in the training was low, only 20% (i.e., two such cases, relative to eight rule exemplars), whereas in two other experiments the level of non-rule exemplars was increased to eight (i.e., 50%, relative to eight rule sentences). Results showed that infants were able to generalize the word order rules in the experiments that included 20% non-rule exemplars in the training input. Infants listened longer to the non-trained rule than to the trained rule in the test trials, the same pattern of results as in their earlier study (Koulaguina & Shi, 2013). In contrast, when the non-rule cases were raised to 50%, rule generalization failed.

Taken together, these results reveal that the proportions of rule instances relative to non-rule instances in the input determine infants' rule generalization. In light of these results, one question to ask is: What is the exact level of exceptions at which rule generalization begins to fail?

The Tolerance Principle (TP) (Yang 2005, 2016) is a theoretical model that addresses this question. The TP algorithm specifies a threshold; that is, the maximum number of exceptions that a rule can tolerate in order to be productive. As defined by the TP, a rule is formed if and only if the number of exceptions (i.e., items not supporting the rule) does not exceed the TP-threshold (θ_N), that is, not exceeding the total number of exemplars (*N*: the sum of rule exemplars and exceptions) divided by the natural log of *N*. The TP is shown in (1):

(1) **Tolerance Principle:** Let *R* be a rule applicable to *N* items, of which *e* are exceptions. *R* is productive if and only iff:
$e \leqslant \theta_N$ where $\theta_N = N/\ln N$ (Yang, 2016: p. 64).

In other words, if the number of exceptions to a rule exceeds this threshold, rule generalization breaks down, and if the number of exceptions to a rule does not exceed the threshold of productivity, rule generalization takes off. The change at the threshold is quantal.

Schuler, Yang and Newport (2016) specifically tested the TP, using an artificial language consisting of nine novel nouns. They calculated the TP-threshold to determine the precise number of regular forms versus exceptions that a productive rule can tolerate. Thus, given a total of nine nouns, the threshold value was 4.096. They then created two training conditions: one in which the number of exceptions was below the TP-threshold (i.e., 5/4 condition: 5 rule exemplars and 4 exceptions), and another in which the number of exceptions was above the TP-threshold (i.e., 3/6 condition: 3 rule exemplars and 6 exceptions). The rule consisted of adding *-ka* to each regular noun to make it plural. The exceptions, i.e., irregular nouns, consisted of other variable plural endings, for example, *-po*. Children aged between 5 and 8 years were trained with nine novel objects, each labeled with a novel word. One group of children was exposed to the 5/4 condition, and another group was exposed to the 3/6 condition. Children saw a picture in each training trial showing a novel object and heard the experimenter produce either the singular or the plural form of the noun, depending on the quantity of the object in the picture. Each noun was heard in its singular and plural forms in different trials. After the exposure, they used a production test with new nouns to examine if children in each condition had formed a productive rule or not. Their results showed that children's performance agreed with the TP predictions. Indeed, when the number of exceptions was below the threshold (i.e., the 5/4 condition), rule generalization succeeded. When the number of exceptions exceeded the threshold (i.e., the 3/6 condition), rule generalization failed.

The TP was proposed for the learning of all types of rules, not just for linguistic rules. This was confirmed in a recent study that tested the TP-threshold in children's learning of mathematical rules (Yang, Lei & Lee, 2019). In particular, the study showed that Cantonese-speaking 3- to 5-year-olds' counting performance was in line with the predictions of the TP-threshold, and their learning of Successor Function was linked to their counting knowledge.

We can also consider the experiments of Koulaguina and Shi (2019) in terms of the TP-threshold values. In their low-exception training condition, 2 out of the 10 total training exemplars were non-rule cases. The tolerance threshold would be 4.34 exceptions (i.e., θ_{10}=10/In10=4.34). Thus, 2 non-rule exemplars were well below this value, and indeed, rule learning succeeded. In their high-exception training condition, 8 out of the total of 16 training exemplars were non-rule cases. The tolerance threshold would be 5.77 (θ_{16}= 16/In16=5.77), so 8 non-rule exemplars well exceeded this value, and as predicted, the rule learning failed. These results were fully consistent with the predictions of the TP.

In the present study, we investigated further if rule generalization indeed functions quantally at the TP-threshold. This aspect of the TP needs to be tested with contrasting experiments by setting the number of exceptions close to the threshold value, a manipulation that was not done in prior artificial rule learning studies. We therefore conducted two experiments, one with exceptions just below the threshold, and the other with exceptions just above the threshold.

2. Experiment 1

We used the speech stimuli of Koulaguina and Shi (2013, 2019) to construct the training input of our present experiment. The stimuli were Russian sentences, some serving as rule exemplars (three-word sentences each going through a movement rule, e.g., ABC→BAC or ABC→ACB), and others as non-rule exemplars (i.e., ABC sentences that did not go through any word order shift). In Experiment 1 we exposed infants to an input of 16 exemplars in which the number of exceptions was set just below the TP-threshold. Specifically, we used the TP to calculate the threshold. That is, the number of exceptions that a productive rule can tolerate for 16 exemplars (i.e., the total sample N) is equal to or smaller than $\theta_{16} = N/\ln N$ = 16/In16 = 5.77. Thus, the number of exceptions in the training input of Experiment 1 was set at 5, and the number of rule exemplars at 11. According to the TP, a productive rule should be formed with this input distribution.

2.1.1. Participants

Twenty-four non-Russian-learning 14-month-old infants (mean: 462 days; range: 446-483; 13 boys, 11 girls) completed the experiment. Five other infants were tested, but their data were excluded due to the following reasons: fussiness (1), getting out of the camera range during test trials (1), and lack of interest in the task (3). After the experiment, each child received a toy as a gift for his/her participation.

2.1.2. Stimuli

The stimuli were compiled based on those of previous studies (Koulaguina & Shi, 2013, 2019). They were sentences in Russian, a language that was unknown to our participants. There were 18 sentences in total (see Table 1), 16 of which were used for constructing the training exemplars, and the remaining two sentences served as the test exemplars. Using the 16 training sentences, we constructed the input distribution for the present experiment (i.e., 11 rule exemplars and 5 exceptions). The test exemplars were the same sound files of Experiments 3 and 4 of Koulaguina and Shi (2019).

Table 1. Sentences used for constructing the stimuli in Experiment 1 (adapted from Koulaguina & Shi, 2013, 2019)

Experimental phase	Base ABC sentences for the word-order shift rules (either ABC→BAC or ABC→ACB; each rule instance consisted of a base sentence and its shifted version)	Non-rule sentences (non-shift singletons; ABC)
Training phase	*Dozhd' zalil cherdak* *Veter gnjot derev'ja* *Vorona nashla pugovitsy* *Machty gnutsja lukom* *Zina gladit plat'e* *Pojte pesnju druzhno* *Veter vybil okna* *Dimke snilos' pole* *Chistim tufli vaksoj* *Budesh vilkoj kushat'* *Flagi utrom snjali*	*Stanut reki polny* *Otzvuk smekha sladok* *Seno pahnet volej* *Skrojut tuchi solntse* *Obuv' skinul rezvo*
Test phase	*Vizhu nosik belki* *Snova milyj vesel*	

The rule sentences went through word order shift rules, ABC to BAC (Rule 1, e.g., *Dozhd' zalil cherdak* → *Zalil dozhd' cherdak*), or ABC to ACB (Rule 2, e.g., *Dozhd' zalil cherdak* → *Dozhd' cherdak zalil*). For the non-rule exemplars, the sentences only appeared in the base order (ABC, e.g., *Stanut reki polny*) and did not go through any movement. Such non-rule cases were not overt violations since they did not do a non-rule-like movement. However, they should be treated by infants as exceptions, as shown in Koulaguina and Shi (2019).

Two groups of sound files were compiled. One group were four files each containing the 11 exemplars of one of the rules plus the 5 non-rule exemplars, and the other group were four files each containing the 11 exemplars of the other rule plus the same 5 non-rule exemplars. The 16 cases in each sound file occurred once and were randomized in order. The original sentence (i.e., ABC) and the shifted version (e.g., BAC) of each exemplar (i.e., a sentence pair) always occurred together and were separated by 700 msec approximately. Moreover, there was a separation of approximately 1200 msec between any two exemplars, that is, between any rule exemplar and an exception, between two exceptions, and between two rule exemplars (i.e., between 2 sentence pairs). The duration was 103 sec for each of the four sound files that contained the ABC→BAC rule exemplars and the exceptions, and 103 sec for each of the four sound files containing the ABC→ACB rule exemplars and the exceptions.

The test stimuli were two novel sentences that were not included in the training input. Each sentence went through the ABC→BAC movement (Rule 1) and the ABC→ACB movement (Rule 2), respectively. There were thus four sound files for the test trials. Within each sound file, the original and the shifted versions of a sentence were separated by about 700 msec, and the pair was repeated three times with a pause of about 1200 msec between the repetitions.

In the training, the average sentence duration was 2.88 sec (*SD* = .46) for Rule 1, 2.86 sec (*SD* = .42) for Rule 2, and 2.48 sec (*SD* = .18) for the exceptions. In the test phase, the average sentence duration was 2.55 sec (*SD* = .09) for Rule 1, 2.55 sec (*SD* = .09) for Rule 2.

The visual stimuli accompanying the speech stimuli included an animation of colorful moving circles and an animation of dynamic blue geometric forms. A jumping star with birds singing served as the attention-getter. An electronic sound of a bouncing ball was also used during the test phase (for the pre-test trial and the post-test trial).

2.1.3. Design and procedure

The experiment started with a training phase in which each infant was exposed to one of the two input conditions: either Rule 1 with exceptions (i.e., ABC→BAC exemplars and non-rule ABC sentences), or Rule 2 with exceptions (i.e., ABC→ACB exemplars and non-rule ABC sentences). Half of the infants were randomly assigned to Rule 1 training condition, and the other half to Rule 2 condition (see Table 2). Each group heard four trials presenting the four training sound files (as describe above), with each trial containing the 16 exemplars (i.e., 11 for the rule and the 5 exceptions). In the first three trials the animation of colorful moving circles was shown on both screens, simultaneously with the speech stimuli. In the last trial, the visual display was the animation of dynamic blue geometric forms, again on both screens and was presented simultaneously with the speech stimuli. Between trials, the attention getter (a jumping star with birds singing) appeared on both screens. Auditory stimuli were played simultaneously from loudspeakers below both screens. The total duration of the training phase was about seven minutes for each group.

The test phase started with a pre-test trial, followed by two introduction trials, then 14 test trials, and finally a post-test trial. The visual stimuli for all these trials were the same animation of the dynamic blue geometric forms. The auditory stimuli for the pre- and post-test trials were repetitions of an electronic sound of a bouncing ball. The auditory stimuli for the two introduction trials were novel test sentences that had not occurred during training (see Table 1), with one sentence going through one rule (e.g., Rule 1) in one trial, and the second sentence going through the other rule (e.g., Rule 2) in the other trial. The two trials were each initiated by infants' look at the central screen, and the sentence pair was played in full (e.g., the original and the shifted version), ensuring that all infants heard one complete version of each test case before the 14 test trials. The order of the two introduction trials and the specific sentence in which the rules appeared were counterbalanced across infants. Specifically, half of the infants heard the Rule 1 trial first, followed by Rule 2, and the other half heard the Rule 2 trial first, followed by Rule 1. Moreover, one sub-group of infants heard the first sentence in Rule 1 and the second sentence in Rule 2, while another sub-group heard the first sentence in Rule 2 and the second sentence in Rule 1. The duration of each introduction trial was 6000 msec.

Table 2. Design of Experiment 1, with exceptions in training below the threshold of the Tolerance Principle, i.e., $e < N/\ln N = 16/\ln 16 = 5.77$

Experimental phase	Group 1	Group 2
Training (e < 5.77 threshold)	11 rule exemplars (Rule 1: ABC→BAC)	11 rule exemplars (Rule 2: ABC→ACB)
	5 exceptions (ABC)	5 exceptions (ABC)
Test (Novel exemplars going through movement rules)	ABC→BAC (trained rule) ABC→ACB (non-trained rule)	ABC→ACB (trained rule) ABC→BAC (non-trained rule)

Subsequently, the test trials began, presenting the same two novel sentences of the two introduction trials. Thus, there were two test trial types: the trained rule versus the non-trained rule. The two main groups were controlled for each other, that is, the trained rule for one group was the non-trained rule for the other group, and vice versa. The two types of test trials alternated (2x7=14 total trials). The test trials were fully infant-controlled, that is, a trial would start if the child looked at the screen, and it would end if he or she looked away for more than 2000 msec. A trial would also end if the maximum trial length (21 sec) was reached. The sentence pair in a trial could be heard for a maximum of three times. The counterbalancing of the first test trial and the specific sentence in which a rule was presented was the same as the two introduction trials.

Moreover, the first introduction trial and the first test trial were always correspondent. For example, if a child heard a sentence in Rule 1 in the first introduction trial, the first test trial would also present the same sentence in Rule 1. The same attention getter used in the training phase (a jumping star with birds singing) was also used between any two trials during the test phase.

During the training phase, the infant and the parent were invited into an acoustic room. There were two small screens, one on the far left of the room and the other on the far right, a sofa in the center facing each screen diagonally, and soft toys in a box to the side of the sofa. Parents were instructed not to talk to their child, but that they could play silently with their child using the toys in the box. The infant could also move around in the room during this phase. The parent wore headphones to hear masking music. A researcher in the adjacent room observed the baby on a monitor (connected with a camera in the acoustic room), and initiated each training trial when the baby looked at a screen.

After the training phase, the parent and the infant immediately moved to another acoustic room in which there was no toy. In this room, there was a centered TV monitor and a sofa. The infant was invited to sit on his or her parent's lap on the sofa, facing the central screen, which displayed the visual stimuli. Auditory stimuli were also played centrally. The parent heard masking music through headphones and was asked not to interact at all with his or her child. A camera filmed the baby and sent the video simultaneously to a monitor in the adjacent room, where the experimenter observed the infant while running the experiment. A computer program presented the stimuli and recorded the looking times during each trial (Oakes, Sperka, Debolt & Cantrell, 2019). Throughout the experiment, the experimenter could not hear or see any stimuli.

2.1.4. Predictions

Since the number of exceptions to the rule was below the TP-threshold, we predicted that infants should learn the rules and generalize them to novel instances. In particular, we predicted that looking times in the trained rule test trials and those in the non-trained rule test trials should differ. Furthermore, we expected a novelty preference (i.e., longer looking towards the non-trained rule), given that rule-learners of the same age (14 months) had showed this looking pattern consistently in multiple experiments involving similar stimuli (see Koulaguina & Shi, 2013, 2019).

2.2. Results

Infants' looking times (i.e., listening times) during the two types of test trials were analyzed. Specifically, for every infant, the average looking time per trial for the trained rule trials and that for the non-trained rule trials were calculated separately. As we had an a priori prediction that infants should show longer looking time to the non-trained rule (see Section 2.1.4), we conducted a one-tailed paired *t*-test to compare the looking times of the two types of test trials. As predicted, infants discriminated the two types of test trials, and moreover, they looked longer while listening to the non-trained rule than to the trained rule, t (23) = -2.816, p = .005, one-tailed. The average looking time per trial was 7.22 sec (SE = .63) for the trained rule and 9.52 sec (SE = .73) for the non-trained rule (see Figure 1). This result indicates that infants generalized the rules when the number of exceptions was just below the TP-threshold.

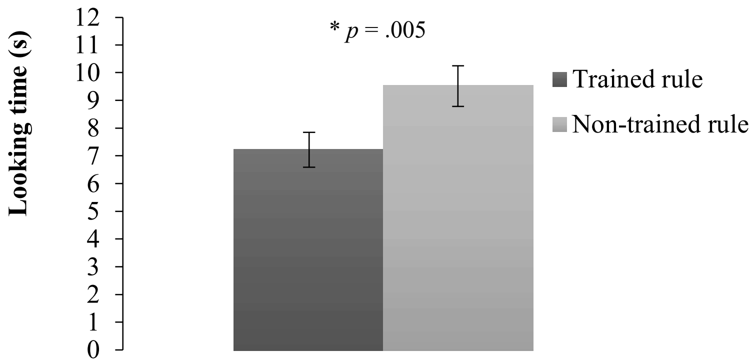

Figure 1. Means and standard errors of the average looking time per test trial for trained rule trials versus non-trained rule trials. Looking time to the non-trained rule was significantly longer than that to the trained rule.

3. Experiment 2

In the second experiment we used the same artificial rules as in Experiment 1, and we manipulated the training input such that the number of exceptions was set just above the TP-threshold. As in Experiment 1, we had a total of 16 exemplars in the training input, with the predicted TP-threshold value at 5.77. Here we increased the exceptions to 6, just above the threshold of 5.77, and reduced the rule exemplars to 10. According to the TP, infants should not form a productive rule since the exceptions have exceeded the threshold.

3.1.1. Participants

Another 24 non-Russian-learning infants aged 14 months (mean: 459 days; range: 435-478; 12 boys, 12 girls) completed the experiment. Three other infants were tested, but their data were excluded due to the following reasons: lack of interest in the task (2), and fussiness (1). After the experiment, each child received a toy as a gift for participating in the experiment.

3.1.2. Stimuli

The stimuli were similar to those in Experiment 1, the only difference being the changed distribution of rule exemplars and exceptions in the training input.

The speech stimuli were 18 Russian sentences (see Table 3) that were taken from Koulaguina and Shi (2013, 2019). Sixteen of these sentences were the training exemplars. The test sentences were identical to those in Experiment 1.

Table 3. Sentences used for constructing the stimuli in Experiment 2 (adapted from Koulaguina & Shi, 2013, 2019)

Experimental phase	Base ABC sentences for the word-order shift rules (either ABC→BAC or ABC→ACB; each rule instance consisted of a base sentence and its shifted version)	Non-rule sentences (non-shift singletons; ABC)
Training phase	*Dozhd' zalil cherdak* *Veter gnjot derev'ja* *Vorona nashla pugovitsy* *Machty gnutsja lukom* *Zina gladit plat'e* *Pojte pesnju druzhno* *Dimke snilos' pole* *Chistim tufli vaksoj* *Budesh vilkoj kushat'* *Flagi utrom snjali*	*Stanut reki polny* *Otzvuk smekha sladok* *Seno pahnet volej* *Skrojut tuchi solntse* *Obuv' skinul rezvo* *Tanets veren bubnu*
Test phase	*Vizhu nosik belki* *Snova milyj vesel*	

Like the training stimuli in Experiment 1, the rule exemplars each went through word order shift rules (Rule 1: ABC→BAC; Rule 2: ABC→ACB) and, for the exceptions, the sentences only appeared in the base order (ABC). Similar to the last experiment, two groups of sound files were compiled for each training condition. One group were four files each containing the 10 exemplars of Rule 1 plus the 6 non-rule exemplars, and the other group were four files each containing 10 exemplars of Rule 2 plus the same 6 non-rule exemplars. The 16 cases in each sound file were randomized in order. The inter-stimulus pauses were set in the same way as in Experiment 1. The duration was 100 sec for each sound file that contained the ABC→BAC rule exemplars and the exceptions, and 100 sec for each sound file containing the ABC→ACB rule exemplars and the exceptions.

In the training input, the average sentence duration was 2.91 sec (*SD* = .47) for Rule 1, 2.89 sec (*SD* = .43) for Rule 2, and 2.43 sec (*SD* = .20) for the exceptions.

3.1.3. Design and procedure

The design and the procedure were the same as in Experiment 1. The only difference was the training distribution that changed with an increase of one non-rule exemplar and a decrease of one rule exemplar relative to Experiment 1 (see Table 4). The test phase was exactly the same as in Experiment 1.

Table 4. The design of Experiment 2, with exceptions in training above the threshold of the Tolerance Principle, i.e., $e > N/\ln N = 16/\ln 16 = 5.77$

Experimental phase	Group 1	Group 2
Training (e > 5.77 threshold)	10 rule exemplars (Rule 1: ABC→BAC)	10 rule exemplars (Rule 2: ABC→ACB)
	6 exceptions (ABC)	6 exceptions (ABC)
Test (Novel exemplars going through movement rules)	ABC→BAC (trained rule) ABC→ACB (non-trained rule)	ABC→ACB (trained rule) ABC→BAC (non-trained rule)

3.1.4. Predictions

Since the number of exceptions to the rules was above the TP-threshold, infants should fail to generalize the rules to novel instances. Specifically, we predicted that their looking times in the trained rule test trials and those of the non-trained rule test trials should not differ.

3.2. Results

As in Experiment 1, each infant's average looking time per trial in the trained rule trials and in the non-trained rule trials were calculated respectively. As predicted, a paired t-test showed that the looking times for the two types of test trials were not different, t (23) = -.975, p = .340, two-tailed. Infants did not discriminate the two rules in the test trials. Their looking time was 6.56 sec (SE = .57) for the trained rule trials, and 7.10 sec (SE = .55) for the non-trained rule trials (see Figure 2). Thus, infants did not show evidence of generalizing the rules when the number of exceptions was just above the TP-threshold.

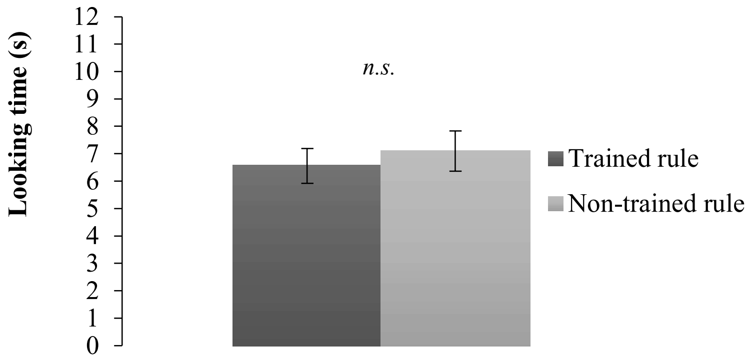

Figure 2. Means and standard errors of the average looking time per test trial for trained rule trials versus non-trained rule trials. Looking times to the two types of test trials did not differ significantly.

We conducted more analysis to further address the question of the quantal effect of the TP. Recall that according to the TP, when the number of exceptions to a rule exceeds the threshold, rule generalization should fail, and when the number of exceptions to a rule does not exceed the threshold, rule generalization succeeds. The TP predictions emphasize an abrupt change in learning at the threshold. In Experiment 1 we predicted that babies should learn and generalize the rules, by showing a novelty preference in test trials (i.e., longer looking towards non-trained rule), given the results of previous experiments that used similar stimuli and task (Koulaguina & Shi, 2013, 2019). This looking pattern was confirmed in Experiment 1. In Experiment 2 we predicted that infants should yield no evidence of rule generalization, by showing no differential looking times in the trained rule and non-trained rule trials during the test phase. This looking pattern was confirmed in Experiment 2. To further assess the quantalness of the TP, we compared the two experiments. In particular, we first calculated the differential score of the test trials for each baby, i.e., the average looking time to the non-trained rule minus the average looking time to the trained rule. For Experiment 1, the differential scores should be an overall positive value since more looking to the non-trained trials was expected (i.e., a novelty preference). In Experiment 2, we expected the differential scores to be close to zero. Thus, given the quantal nature of the TP, the differential scores should be higher in Experiment 1 than in Experiment 2. As predicted, an unpaired *t*-test revealed that the differential score was significantly higher in Experiment 1 (M = 2.3 sec, SE = .82) than in Experiment 2 (M = .54 sec, SE = .55), t (46) = 1.785, p = .04, one-tailed.

Taken together, our results provide supporting evidence for the quantal nature of the TP. Infants' rule learning showed sensitivity to small changes in the number of rule exemplars and that of exceptions close to the tolerance threshold.

4. Discussion

Previous studies showed that children can learn and generalize rules when the input contains certain numbers of exceptions (Gómez & Lakusta, 2004; Koulaguina & Shi, 2019; Schuler et al., 2016; Yang et al., 2019). The results of these studies and those of our present study are fully consistent with the predictions of the TP. When the number of exceptions is high relative to rule exemplars, rule learning fails. When the number of exceptions is low, rule generalization succeeds. Our experiments thus add further data showing that the level of non-rule exemplars affects children's rule abstraction.

One crucial component of the TP is the quantal nature of rule productivity at the threshold. The experiments of Koulaguina and Shi (2013, 2019) and Gómez and Lakusta (2004) did not test this aspect. Schuler et al. (2016) provided a partial answer. In their study, rule productivity was observed in children when the number of exceptions was set closely below the tolerance threshold. The tolerance threshold was 4.096 for their sample, and they included

4 exceptions. In contrast, in the condition in which children failed to learn the rule, the threshold was kept at the same level (i.e., 4.096) while they increased the number of exceptions to 6, which seemed somewhat far away from the threshold. In our study here, we tested the quantal nature of rule productivity by manipulating the numbers of exceptions close to the TP-threshold in both experiments. To do so, we set the total sample *N* equally for both experiments (16), thus yielding the TP-threshold at 5.77, and we slightly varied the number of exceptions around this value, i.e., 5 in Experiment 1 and 6 in Experiment 2. Our results thus provided supporting evidence for the quantal prediction of the TP; that is, infants generalized the rules when the number of exceptions was just below the TP-threshold, and they failed to show evidence of rule generalization when the number of exceptions was just above the threshold.

In the present study as well as in the previous work of Koulaguina and Shi (2019), the non-rule exemplars were base sentences (i.e., ABC) that did not go through any movement, so they were not overt violations of the rules. In principle, such exemplars could be considered either as instances that did not have a chance to apply the rules, or as true exceptions to the rules. Both scenarios exist in natural languages. It is hence interesting to know how infants treat such cases. This question was addressed in Koulaguina and Shi (2019). Across four experiments, they showed that infants perceived such non-shifting exemplars as exceptions. Had infants not treated them as exceptions, they should have learned the rules in the high-exception experiments as they did in the low-exception experiments. Instead, rule learning was impeded when the number of such non-rule exemplars was high. That is, the failure in rule generalization indicated that the non-rule exemplars were treated as exceptions. In comparison, the exceptions in all other previous studies (Gómez & Lakusta, 2004; Schuler et al., 2016; Yang et al., 2019) were overt violations, which were clearly different from the form of rule exemplars. The results of our present experiments replicated those Koulaguina and Shi (2019), further confirming that any deviant instances are perceived by infants as exceptions. The kind of exceptions in our studies functions in the same way as overt violations, operating according to the TP.

The frequency of exemplars was uniform in our present experiments, as in those of Koulaguina and Shi (2013) and Gómez and Lakusta (2004). That is, all exemplars had equal frequency of occurrences. Our rule and non-rule exemplars each occurred four times during training. However, in natural language this is not the case. Different words and word combinations have different frequencies, some occurring very often, and some less often (e.g., Vallian, Solt & Stewart, 2009; Yang, 2013). This kind of unequal frequencies was reflected in the exemplars of Schuler et al. (2016). In Koulaguina and Shi (2019) rule instances were of equal token frequencies (four times for each instance) in the input across experiments. The token frequency for each exception was equal to that of each rule instances in two of their experiments, while it was much higher (16 times each) in the other two experiments. Interestingly, they found that token frequencies did not have an impact on infants' rule generalization. This agrees with the TP, which concerns only types. In the TP-threshold calculation (i.e., $\theta_N = N/\ln N$), the N refers to the number of types (i.e., the total number of different exemplars), and the threshold for the exceptions also refers to the number of types. More experiments are needed to better understand how token frequency distribution might affect rule generalization.

In conclusion, the results of the present study show that linguistic rule generalization is governed by the TP, even in infants as young as 14 months, an age when grammatical acquisition is just beginning. Furthermore, we demonstrate that the TP functions quantally in infants; that is, the number of exceptions just below versus just above the TP-threshold value leads to the predicted success versus failure in rule generalization.

References

Gerken, LouAnn. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition, 98*(3), B67-B74.

Gómez, Rebecca L., & LaKusta, Laura. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science, 7*(5), 567–580.

Koulaguina, Elena, & Shi, Rushen. (2013). Abstract rule learning in 11-and 14-month-old infants. *Journal of Psycholinguistic Research, 42*(1), 71-80.

Koulaguina, Elena, & Shi, Rushen. (2019). Rule generalization from inconsistent input in early infancy. *Language Acquisition, 26*(4), 416-435.

Marcus, Gary F., Vijayan, Sugumaran, Bandi Rao, Shoba, & Vishton, Peter M. (1999). Rule learning by seven-month-old infants. *Science, 283*, 77–80.

Oakes, Lisa M., Sperka, Daniel, DeBolt, Michaela C., & Cantrell, Lisa M. (2019). Habit2: A stand-alone software solution for presenting stimuli and recording infant looking times in order to study infant development. *Behavior Research Methods, 51*(5), 1943-1952.

Schuler, Kathryn, D., Yang, Charles, & Newport, Elissa L. (2016). Testing the Tolerance Principle: children form productive rules when it is more computationally efficient to do so. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 2321–2326). Austin, TX: Cognitive Science Society.

Valian, Virginia, Solt, Stephanie, & Stewart, John. (2009). Abstract categories or limited-scope formulae? The case of children's determiners. *Journal of Child Language, 36*(4), 743-778.

Yang, Charles. (2005). On productivity. *Linguistic Variation Yearbook, 5*(1), 265-302.

Yang, Charles. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences, 110*(16), 6324-6327.

Yang, Charles. (2016). *The price of linguistic productivity.* Cambridge, MA: The MIT Press.

Yang, Charles, Lei, Margaret, & Lee, Thomas. (2019). Languages shapes children's understanding of number. Presentation at the 44th Annual Boston University Conference on Language Development, Boston, MA.

Proceedings of the 45th annual Boston University Conference on Language Development

edited by Danielle Dionne
and Lee-Ann Vidal Covas

Cascadilla Press Somerville, MA 2021

Copyright information

ISSN 1080-692X
ISBN 978-1-57473-067-8 (2 volume set, paperback)

Ordering information

To order a copy of the proceedings or to place a standing order, contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, sales@cascadilla.com, www.cascadilla.com