Cognition 115 (2010) 362–366

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Brief article

Infants use rational decision criteria for choosing among models of their input

LouAnn Gerken *

Department of Psychology, University of Arizona, Tucson, AZ 85721, United States

ARTICLE INFO

Article history:
Received 21 October 2009
Revised 7 January 2010
Accepted 13 January 2010

Keywords:
Infants
Generalization
Rational learner
Bayesian inference

ABSTRACT

Previous work demonstrated that 9-month-olds who were familiarized with 3-syllable strings consistent with both a broader (AAB or ABA) and narrower (AA*di* or A*di*A) generalization made only the latter. Because the narrower generalization is a subset of the broader one, any example that is consistent with the broader generalization but not the narrower one should allow a rational learner to select the broader generalization. The current experiment asked whether infants show evidence of being such learners. Infants who heard the stimuli that previously led to the narrower generalization plus three counterexamples mixed into the last five stimuli made the broader generalization at test. A control condition ruled out the possibility that infants based their generalization on the last five familiarization stimuli. The new findings suggest that infants effectively consider multiple competing models for their input and use rational decision criteria for selecting among these models.

1. Introduction

The past 15 years of infant language research have demonstrated that infants are able to track many linguistically relevant properties of their environment (Gómez, 2002; Saffran, Aslin, & Newport, 1996) and to generalize beyond the specific surface properties of this input to more abstract structure (Chambers, Onishi, & Fisher, 2003; Gerken, 2004; Gerken & Bollt, 2008; Gómez & Gerken, 1999; Marcus, Vijayan, Rao, & Vishton, 1999; Maye, Werker, & Gerken, 2002; Saffran & Thiessen, 2003). What is the basis for infants' ability to generalize? We can take this question to reflect a more basic question about the nature of human mind: Do learners attempt to recover from their input a model of the world that could have generated that input (i.e., a generative model) and use this model to generalize? Or is simply storing the input in the appropriate form sufficient to allow for the types of generalization we see in infant studies?

The view that learners are model builders, which is the focus of the current study, is consistent with recent Bayesian accounts of learning (e.g., Tenenbaum & Griffiths, 2001), and it makes two interrelated predictions. First, because any input could logically have been produced by multiple models, a learner might entertain, at least at some point during learning, more than one possible model for their input. Second and more specifically, if models of the input are logically contradictory, a rational learner should be able to very quickly rule out the contradicted model in favor of the supported one. The experiment reported here reflects an attempt to test these two predictions.

There is a growing body of literature suggesting that infants' and children's behavior is consistent with creating and comparing generative models. For example, Xu and Tenenbaum (2007) found that children presented with a novel word labeling a single example of a basic-level category (e.g., a Dalmatian as an instance of a dog) could interpret the word as referring to either the broader category (dogs) or the narrower category (Dalmatians) consistent with that one example. In contrast, children presented with three examples of the narrower category (e.g., three different Dalmatians) treated the novel category label as

* Tel.: +1 520 621 4327; fax: +1 520 621 9901.
E-mail address: gerken@u.arizona.edu

0010-0277/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2010.01.006

referring to the narrower category (Dalmatians) instead of the broader one (dogs) (also see Gerken & Bollt, 2008; Needham, Dueker, & Lockhead, 2005; Quinn & Bhatt, 2005). The data from the condition in which children were presented with a single input example suggest that they entertained multiple possible referents (or models). The data from the three examples condition suggest that children treat some referents (or models) as more likely than others as additional data are encountered.[1] Xu and Garcia (2008) explored 8-month-olds' capacities for building generative models by testing whether infants can make inferences about the population from which a sample was taken. Infants who saw a researcher take 4 red and 1 white ball from a box subsequently looked longer when shown a box containing more white than red balls (inconsistent with the first sample) than a box containing more red than white balls (consistent with the first sample).

Turning to linguistic structure, Gerken (2006) suggested that 9-month-olds entertain multiple models about the appropriate generalization to make over a set of linguistic strings and give more weight to the model that best fits the particular distribution of input. Because that study forms the basis of the experiments presented here, it will be described in some detail. Nine-month-olds were familiarized with 2 min of stimuli from either the first column or the diagonal of Table 1 (Marcus et al., 1999).[2] Infants were then tested on four test trials of novel stimuli. The test stimuli *kokoba* and *popoga* were consistent with the AAB familiarization stimuli, while the test stimuli *bakoba* and *gapoga* were consistent with the ABA familiarization stimuli. A result indicating generalization might be that infants familiarized with AAB stimuli listened longer to AAB test stimuli, while infants familiarized with ABA stimuli listened longer to ABA test stimuli (i.e., a familiarity preference). The results showed that only infants familiarized with the diagonal showed such generalization; infants familiarized with the column stimuli did not. Gerken hypothesized that, because the column stimuli exhibited a narrower generalization (AA*di* or A*di*A), infants in that condition ultimately settled on this local generalization. To test that hypothesis, Gerken familiarized another group of infants with the column stimuli, but tested them on stimuli in which the A syllables were novel and the B syllable was the syllable *di*. Infants generalized, suggesting that they made the narrower generalization.

One interpretation of the previous results that is consistent with Bayesian inference is that infants entertained both the broader (AAB or ABA) and the narrower (AA*di* or A*di*A) models of the input during familiarization. Under a Bayesian account, it would be a "suspicious coincidence" that none of the stimuli violated the narrower generalization *if* the broader generalization was in fact the correct one (Tenenbaum & Griffiths, 2001). This suspicious coincidence could have biased infants toward the narrower generalization. However, because the narrower generalization is a subset of the broader one, stimuli that are consistent with the broader but not the narrower generalization should rule out the narrower generalization for a rational learner. For example, if you are given the set of numbers 80, 10, 200, 30, 120, 60, 90, 170, and 40 as examples of some rule, you will probably consider the hypothesis that all of these numbers are divisible by 10 more likely than the hypothesis that all of the numbers are divisible by 2. That conclusion is based on the fact that 'divisible by 10' is a subset of 'divisible by 2', and all of the numbers that you have so far encountered are consistent with the narrower hypothesis. However, if the numbers 58, 142, and 6 are added to the list, you should quickly abandon your 'divisible by 10' hypothesis in favor of the 'divisible by 2' hypothesis, no matter that the large majority of the numbers you've encountered are consistent with the narrower hypothesis.

Table 1
AAB familiarization stimuli used by Marcus et al. (1999). The first column and diagonal, as well as the ABA version of the stimuli, were used as familiarization stimuli by Gerken (2006).

A	B			
	di	je	li	we
le	leledi	leleje	leleli	lelewe
wi	wiwidi	wiwije	wiwili	wiwiwe
ji	jijidi	jijije	jijili	jijiwe
de	dededi	dedeje	dedeli	dedewe

The experiment reported here asked if 9-month-old infants show a similarly rapid change in their favored generalization based on a small number of counterexamples (three). Infants in the *column-plus-3* condition were exposed to 2 min of the stimuli from the column of Table 1 or the ABA parallel, in which the B element was always the syllable *di*. This condition led the infants tested by Gerken (2006) to make only the narrower generalization. Mixed into the last five strings in the familiarization stimuli were three strings from the diagonal of Table 1, the form of which are only consistent with the broader AAB or ABA generalization, not the narrower AA*di* or A*di*A model. Just as in the 'divisible by 10' vs. 'divisible by 2' example above, infants should use the three counterexamples to the narrower generalization to infer that the broader generalization is more appropriate. A control condition (*music-plus-5*) was also run on another group of infants to ensure that infants were not simply generalizing based on the last few stimuli they heard prior to test.

2. Methods

Participants were 36 infants (20 female) ranging in age from 8 months 17 days to 9 months 23 days, with a mean of 9 months 4 days. All infants were at least 37 weeks to term, at least 5 lbs 8 oz at birth, had no history of speech or language problems in their nuclear family, and were

[1] Aside from learners' responses to situations in which they are presented with a single input example, it is quite challenging to distinguish between an account in which learners entertain *simultaneously* more than a single model from one in which they switch *serially* from one model to another as the familiarization data are presented, showing evidence of the last model they entertained at test. This interpretational challenge applies to the current study as well. However, given that the single input example situation does suggest the availability of more than one model, the simultaneous model may be more parsimonious until evidence to the contrary is produced.

[2] While Marcus et al. (1999) tested 7-month-olds, Gerken (2006, footnote 1) found that 9-month-olds yielded more robust data.

not given medication for an ear infection within 1 week of testing. Half of the infants were assigned to the experimental (column-plus-3) condition and half to the control (music-plus-5) condition. Within each condition, half of the infants heard familiarization stimuli consistent with an AAB grammar and half with an ABA grammar. An additional 17 infants were tested and not included in the study due to failure to provide listening times longer than 2 s on all four test trials (15) or parental distraction (2).

2.1. Materials

All infants were exposed to 2 min of familiarization stimuli followed by four test trials. In the familiarization phase, half of the infants were assigned to the column-plus-3 (experimental) condition, and within that condition, half of the infants heard AAB and half heard ABA strings. Familiarization strings in this condition comprised 2 min of a randomly ordered sample of the 3-syllable nonsense items *leledi, wiwidi, jijidi,* and *dededi* (AAB) or *ledile, widiwi, jidiji,* and *dedide* (ABA). Before the last two stimuli of the 2 min familiarization sequence, three stimuli from the diagonal of Table 1 or the ABA equivalent were added in random order, such that the last five stimuli were *wiwije, jijidi, dedewe, jijili, wiwidi* (AAB) or *wijewe, jidiji, dewede, jlijii, wididwi* (ABA).

The other half of the infants were assigned to the music-plus-5 condition, such that during the familiarization phase, they heard 2 min of recorded Andean music followed by the same five AAB or ABA stimuli that ended the familiarization sequence for infants in the column-plus-3 condition. Thus, the familiarization stimuli immediately preceding test were identical for both groups of infants. Syllable strings for both experiments were generated with the speech function of a Power Macintosh, system 8.6, using the Victoria voice at the default rate. One sec. pauses were inserted between the 3-syllable words, using speech analysis software.

Regardless of familiarization condition, test items were identical across infants and identical to those used by Gerken (2006). They comprised two 30 s trials with the strings *kokoba* and *popoga* in different random orders (AAB test trials) and two trials with *bakoba* and *gapoga* in different random orders (ABA test trials).

2.2. Procedure

The headturn preference procedure (Kemler Nelson et al., 1995) was used. Infants were seated on a parent's lap in a small room. The parent listened to pop music through headphones in order to mask the stimuli heard by the infants and prevent inadvertent influence on the infant. During the familiarization phase, a light directly in front of the infant flashed until the observer, blind to the experimental condition and unable to hear the stimuli, judged the infant to be looking at it, at which point a light on the left or right would begin flashing. When the infant looked first at the side light and then away for two consecutive seconds, the center light would resume flashing, and the cycle would begin again. This continued for the duration of the familiarization stimulus, which played uninterrupted to its conclusion. In this stage there was no correspondence between infants' looking behavior and the stimuli.

After the familiarization sequence ended, the test phase began immediately. The flashing lights behaved the same way except that now the sound was contingent on the infant orienting to a side light. Each time a side light began flashing and the infant oriented toward it, one of the four test trials would play, continuing until either the infant looked away for two consecutive seconds or the test trial reached its conclusion.

3. Results

Test trials were classified as consistent vs. inconsistent with the familiarization stimuli for each infant. For example, an AAB test trial was classified as consistent for an infant who heard AAB familiarization stimuli, but as inconsistent for an infant who heard ABA familiarization stimuli. Infants' listening times for consistent vs. inconsistent test trials are shown in Fig. 1. In addition to the listening times for the two conditions of the current experiment, times are shown for the column only condition of Gerken (2006), because it is the comparison between this condition and the column-plus-3 condition of the current experiment that is of greatest interest. Infants in the music-plus-5 condition were tested on syllable stimuli after listening to only five similar stimulus strings during familiarization, while infants in the column-plus-3 condition were tested on syllable stimuli after listening to 2 min of similar stimulus strings. Not surprisingly, therefore, infants in the music-plus-5 condition exhibited listening times that were more than 50% longer (mean = 15.30 s, SD = 6.87) than infants in the column-plus-3 condition (mean = 9.76, 4.13; $t(27) = 3.53$, $p < 0.002$, 2-tailed). Because the large difference in overall listening times might obscure differences in how infants in the two familiarization conditions responded to the test items, the listening times in each condition were made comparable by converting them to *z*-scores. In order to compare the results from the two conditions of the current experiment to the column only condition of Gerken (2006), the latter listening times were also converted to *z*-scores, and all three conditions (column, column-plus-3, music-plus-5) were entered into a 3 familiarization condition $\times$ 2 consistency (test stimuli consistent vs. inconsistent with familiarization stimuli) ANOVA. Neither the main effect of familiarization condition nor the main effect of consistency was significant (both Fs < 1). However, as predicted, there was a significant familiarization $\times$ consistency interaction ($F(2, 49) = 3.76$, $p < .05$). Because the crucial prediction concerned the difference in listening times between consistent and inconsistent test items in Gerken (2006) column only condition vs. the current column-plus-3 condition, a second ANOVA was performed on the *z*-scores for these two conditions. As predicted, there was again a significant familiarization $\times$ consistency interaction ($F(1, 32) = 4.85$, $p < 0.05$). Follow-up *t*-tests on the mean listening times revealed that infants listened significantly longer to the consistent test items in the column-plus-3 familiarization condition ($t(17) = 2.34$,

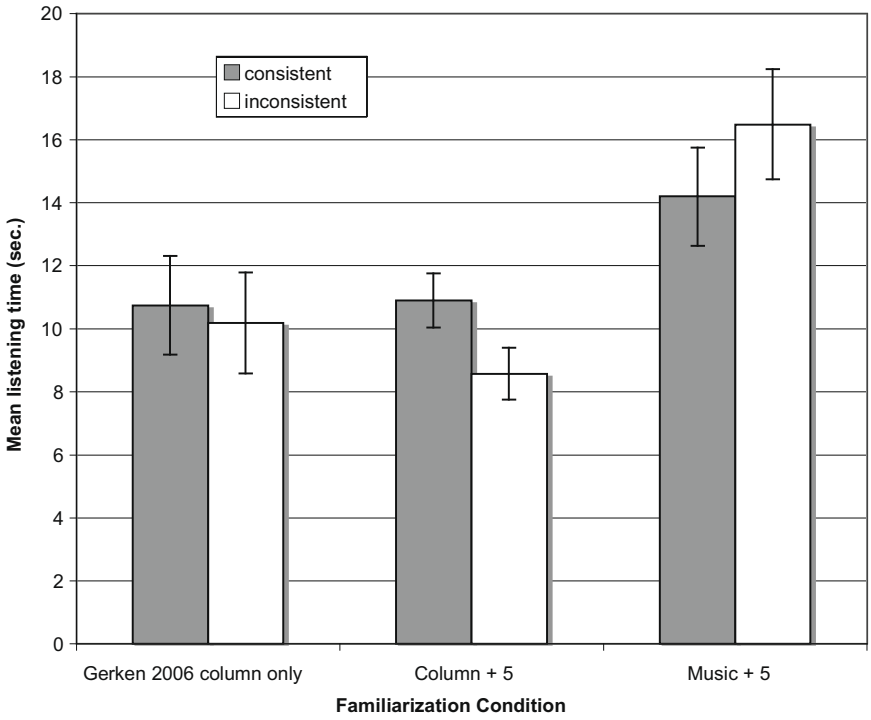

Fig. 1. Mean listening times and SE for the consistent vs. inconsistent test stimuli across three conditions.

$p < 0.05$, 2-tailed). In contrast, the difference in listening times for consistent and inconsistent test trials for the original Gerken column only condition ($t(15) = 0.57$, $p < 0.60$, 2-tailed) and music-plus-5 familiarization condition ($t(17) = 1.13$, $p < 0.30$, 2-tailed) did not approach significance.

4. Discussion and conclusion

As predicted, infants in the column-plus-3 condition generalized to new AAB or ABA test stimuli that were composed of entirely different syllables than the familiarization items. Their generalization in this condition stands in contrast to the lack of generalization observed in the column condition of Gerken (2006), despite the fact that the familiarization stimuli in the current study included just three different items from those in the previous study. Therefore, at the very least, the new research coupled with the previous finding indicates that infants can be moved from showing no hint of a particular generalization to making it robustly based on just three input stimuli. This observation by itself is remarkable and suggests that infants might be better described in terms of incremental, as opposed to batch, learning models.

The music-plus-5 familiarization condition was included to rule out the possibility that infants have a very small window over which they generalize and in particular that the broader generalization was not made simply on the basis of the last five stimuli. Infants failed to generalize in this condition, indicating that a window of five input stimuli is not enough to support generalization. The explanation offered here for infants' pattern of behavior across the three conditions shown in Fig. 1 is that the three counterexamples *in the context of the preceding stimuli* in the column-plus-3 condition led to the broader generalization. Taken together, the results shown in Fig. 1 are consistent with the view that infants in the column only and column-plus-3 conditions entertained both the broader and narrower models of the input (but see footnote 1). In the column only condition of Gerken (2006), infants assigned a higher likelihood to the narrower generalization, perhaps based on the suspicious coincidence of hearing only stimuli consistent with the narrower generalization.[3] The higher likelihood of the narrower generalization appears to have determined their behavior at test. However, in the current study, just three counter examples to the narrower generalization caused infants to assign a higher probability to the broader generalization, which they demonstrated at test. In short, these data give license to further explore learning architectures in which infants consider multiple possible models that might have generated the input that they encounter and choose among these models using rational decision criteria.

Acknowledgments

This research was supported by NICHD #R01 HD042170. The author is grateful for helpful comments

[3] Infants in the music-plus-5 condition may also have begun to entertain models of their very limited input. However, the model that would have supported the generalization tested may not have been robust enough to allow discrimination of the consistent and inconsistent test items (work in progress with Josh Tenenbaum sparked this observation).

on earlier versions from Rebecca Gómez, Laura Schulz, and Josh Tenenbaum, and three anonymous reviewers.

References

Chambers, K. E., Onishi, K. H., & Fisher, C. L. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition, 87*, B69–B77.

Gerken, L. A. (2004). Nine-month-olds extract structural principles required for natural language. *Cognition, 93*, B89–B96.

Gerken, L. A. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition, 98*, B67–B74.

Gerken, L. A., & Bollt, A. (2008). Three exemplars allow at least some linguistic generalizations: Implications for generalization mechanisms and constraints. *Language Learning and Development, 4*(3), 228–248.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431–436.

Gómez, R. L., & Gerken, L. A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*(2), 109–135.

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

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

Maye, J., Werker, J. F., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101–B111.

Needham, A., Dueker, G., & Lockhead, G. (2005). Infants' formation and use of categories to segregate objects. *Cognition, 94*, 215–240.

Quinn, P. C., & Bhatt, R. S. (2005). Learning perceptual organization in infancy. *Psychological Science, 16*(7), 511–515.

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

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology, 39*, 484–494.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24*, 629–640.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 5012–5015.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*, 245–272.