Developmental Science

Developmental Science (2014), pp 1–10 DOI: 10.1111/desc.12183

PAPER

Surprise! Infants consider possible bases of generalization for a single input example

LouAnn Gerken,[1] Colin Dawson,[1] Razanne Chatila[1] and Josh Tenenbaum[2]

1. Department of Psychology, University of Arizona, USA
2. Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA

Abstract

Infants have been shown to generalize from a small number of input examples. However, existing studies allow two possible means of generalization. One is via a process of noting similarities shared by several examples. Alternatively, generalization may reflect an implicit desire to explain the input. The latter view suggests that generalization might occur when even a single input example is surprising, given the learner's current model of the domain. To test the possibility that infants are able to generalize based on a single example, we familiarized 9-month-olds with a single three-syllable input example that contained either one surprising feature (syllable repetition, Experiment 1) or two features (repetition and a rare syllable, Experiment 2). In both experiments, infants generalized only to new strings that maintained all of the surprising features from familiarization. This research suggests that surprise can promote very rapid generalization.

Introduction

Human infants are amazing generalizers. In dozens of studies, infants who are briefly familiarized with a set of visual or auditory stimuli that embody a pattern or rule listen differentially to new stimuli that are either consistent or inconsistent with the pattern or rule (e.g. Chambers, Onishi & Fisher, 2003; Dawson & Gerken, 2009; Gerken, 2004, 2006, 2010; Gerken & Bollt, 2008; Gerken, Wilson & Lewis, 2005; Gómez, 2002; Gómez & Gerken, 1999; Gómez & LaKusta, 2004; Gómez & Maye, 2005; Marcus, Fernandes & Johnson, 2007; Marcus, Vijayan, Rao & Vishton, 1999; Needham, Dueker & Lockhead, 2005; Quinn & Bhatt, 2005; Quinn, Bhatt, Brush, Grimes & Sharpnack, 2002; Saffran & Thiessen, 2003). Moreover, some of the research cited above, as well as other studies of infants and young children, suggest that generalization reflects a process that looks remarkably like implicit hypothesis selection (Gerken, 2006, 2010; Xu & Denison, 2009; Xu & Garcia, 2008; Xu & Kushnir, 2012; Xu & Tenenbaum, 2007a). For example, Gerken (2010) familiarized 9-month-olds for 2 min to four stimuli that reflected an AA*di* or A*di*A pattern (where the first and second syllables are the same and the third syllable is *di* or the first and third syllables are the same and the second syllable is *di*). The infants failed to generalize to new AAB or ABA test strings, but rather required the syllable *di* to also be present in the proper location in the test strings (suggesting that both consistent features of the input – syllable repetition and the presence of *di* – needed to be included in the generalization). However, when one token of just three AAB or ABA strings not containing *di* were inserted into the same 2 min familiarization set, infants did generalize to new AAB or ABA strings. This result suggests that it isn't the sheer number of strings with the AA*di* or A*di*A pattern that drives generalization, but rather the logical structure of the input, with three counter-examples being enough to change the generalization that infants made.

One question that arises from the research on infants' implicit hypothesis selection concerns the source of the putative hypotheses and what input characteristics are

Address for correspondence: LouAnn Gerken, Department of Psychology, University of Arizona, Tucson, AZ 85721-0068, USA; e-mail: gerken@email.arizona.edu

required to initiate the generalization process. The research reported here focuses on the second part of that question. Previous research on infant generalization has shown that learners are more likely to make the intended generalization when the number of input examples increases from one to three or four (e.g. Gerken, 2006, 2010; Gerken & Bollt, 2008; Gweon, Tenenbaum & Schulz, 2010; Needham *et al.*, 2005; Quinn & Bhatt, 2005; Xu & Tenenbaum, 2007a, 2007b). Such research could be taken to suggest that young learners don't consider *any* possible bases of generalization until they have encountered a small set of input examples. For example, Quinn and Bhatt (2005) found that 3- to 4-month-olds failed to generalize from a single example of letters (Xs and Os) arranged to form a set of vertical bars. However, 3- to 4-month-olds did appear to make the vertical bar generalization when shown three different examples of shapes (Xs and Os, squares and diamonds, Hs and Is) arranged to form vertical bars. However, counter to the view that multiple examples are necessary for generalization, 6- to 7-month-olds tested in the same paradigm just described were able to make the correct generalization when exposed to a single example (Quinn *et al.*, 2002).

Bhatt and Quinn (2011) hypothesize that variable input (three different examples) was required to 'teach' 3- to 4-month-olds that shape was an organizing principle for the displays they had seen. That is, 3- to 4-month-old infants do not consider shape as the basis for generalization until shape *emerges* as relevant through exposure to multiple input examples. In contrast to the view that variable input teaches or causes to emerge a previously non-existent generalization, a Bayesian hypothesis selection approach to generalization suggests that the younger infants studied by Quinn and Bhatt (2005) might have considered a shape-based generalization after just one example, just as the 6- to 7-month-old infants did (Peterson, 2011). However, something about the model of the visual world entertained by the younger infants made them require more evidence before the shape-based generalization was sufficiently strong to drive behavior (Aslin, 2011; Gerken, 2006, 2010; Xu & Tenenbaum, 2007a).

The view that multiple input examples may be needed to strengthen, but not necessarily to generate, a hypothesis suggests that infants should sometimes demonstrate generalization from a single input example. This prediction is supported by the study described above in which Quinn *et al.* (2002) found that 6- to 7-month-olds were able to generalize to a new test item after seeing just one example of shapes organized in vertical columns or horizontal rows. Such findings raise the question of what allows learners to attach sufficient weight to a particular hypothesis after encountering a single input example. One possible answer can be found in the work of Griffiths and Tenenbaum (2007), who suggest that it is when an input example is inconsistent with the learner's current model of the domain (based on previous experience) and therefore *surprising* that learners take note and seek a new hypothesis to explain the aberrant input (also see Peirce, 1935). According to these authors, such surprising coincidences 'are a rich source of information as to how a theory might be revised, and should be given great attention'.

The role of surprise in generalization falls directly out of Bayes' rule, which says that a hypothesis is bolstered by data precisely to the extent that the hypothesis makes the data less surprising than they otherwise would be. This can be seen by writing Bayes rule in its usual form (1) and dividing both sides by P(H) (2):

(1) $P(H \mid D) = P(H)\, P(D \mid H)/P(D)$
(2) $P(H \mid D)/P(H) = P(D \mid H)/P(D)$

In words, the ratio of posterior to prior plausibility of H is precisely the ratio between the likelihood (unsurprisingness) of the data *with* the hypothesis to the likelihood of the data in general. It should be noted, however, that P(D) here does not necessarily reflect the overall statistical prevalence of the data, but is actually a model-based marginal likelihood, obtained by taking a weighted average over likelihoods under each of several hypotheses: $P(D) = \Sigma\_\{H'\}\ P(H')\ P(D|H')$. The current work tests the prediction that 9-month-olds will generalize from a single input example when the input is surprising from the point of view of the learner's current model of the domain in question.

A corollary to the question of what allows generalization from a single input is what changes over development such that learners of different ages show differences in generalization. Dawson and Gerken (2009, 2011) attempt to answer this question by proposing that part of cognitive development entails splitting the world into smaller and more internally consistent domains. On this view, the 6- to 7-month-olds studied by Quinn *et al.* (2002) might have had a model of the visual world in which objects exist in relatively unorganized relations to each other. Therefore, encountering similarly shaped objects organized into rows or columns might have required 6- to 7-month-olds to rethink their model of the visual world, perhaps to consider 3-D scenes and 2-D graphics as separate domains.

Although we will not directly address the developmental question here, the current work is motivated by Dawson and Gerken's (2009, 2011) explanation of why 4-month-olds, but not 7-month-olds, could generalize AAB or ABA patterns instantiated in musical tones, whereas 7- and 9-month-olds have no problem generalizing the same patterns instantiated in syllables (also see Marcus *et al*., 2007). Dawson and Gerken propose that note repetition is unsurprising on a model of the musical world composed of scalar tones arranged in sequences with small intervals between them (Temperley, 2008). They further propose that while 7-month-olds have begun to develop such a model of the music domain, 4-month-olds have not (e.g. Lynch & Eilers, 1992). On this view, repetition is consistent with the 7-month-olds' music domain model and not deserving of an explanation.

Crucially for the current work, Dawson and Gerken (2011) make a further observation about the surprising nature of repeated syllables in English. Repeated syllables are statistically rare in English, constituting fewer than 3% of approximately 30,000 caregiver utterances from the CHILDES database. Importantly, the statistics of English are consistent with a model of language in which different syllables are strung together to make words and phrases. Because of the large number of different syllables that can be generated by English phonology, repeated syllables that are not grammatical morphemes (e.g. 'the', 'a', etc.) should be rare. If we assume that learners have formed from their prior experience with English a model like the one just described, they should be surprised when, in the laboratory, they encounter utterances with English phonology but repeated syllables. Because such syllables should be surprising, they are deserving of an explanation. Therefore, repetition will be included in learners' generalization. This view predicts the relative ease with which infants of several ages generalize from AAB or ABA syllable patterns.[1]

Importantly, if syllable repetition violates the infant's model of English utterances, infants might attempt to explain the odd input and therefore generalize from a single AAB or ABA syllable string. Testing this prediction with 9-month-olds was the goal of Experiment 1. The results provide the first evidence of generalization from a single input example generated by a language-like rule. However, repeated syllables also appear to have an *a priori* (i.e. model free) salience for humans and other species. This salience is evidenced by the fact that non-humans and newborn humans are both sensitive to them, despite a lack of relevant experience (Gervain, Macagno, Cogoi, Peña & Mehler, 2008; Murphy, Mondragon & Murphy, 2008). The salience of repetition is also consistent with 4-month-olds noting repetition patterns in musical notes and chords (Dawson & Gerken, 2009). Therefore, Experiment 2 added a second rare feature to the single input example ('zh' as in 'vision') in order to determine whether inconsistency with the infants' likely model of English instead of *a priori* salience causes learners to generalize from certain features of the input. The consonant 'zh' is very low in frequency in English, mostly occurs in '-sion' derivational morphemes (e.g. 'conclusion, decision'), and never occurs at the beginning of stressed syllables. Thus, having this consonant appear in such an unusual context could violate infants' growing model of the morpho-phonology of English. The results suggest that, consistent with the Gerken (2010) experiment described above, learners incorporate both surprising features into their generalization at test. A control experiment (Experiment 3) demonstrated that the results of Experiment 2 are not caused by the total amount of overlap between familiarization and test stimuli. Together, the data support the view that features that are inconsistent with the learner's current model of a domain can promote generalization from a single input example and thereby give us a clearer understanding of what input characteristics are required to initiate the generalization process.

[1]Repetition is inherently salient for many species, which explains the sensitivity in rats and newborn humans (e.g. Gervain *et al*., 2008; Murphy *et al*., 2008). However, it appears that repetition begins to be incorporated with knowledge of the particular domain of interest as human infants develop domain-specific sensitivity (Dawson & Gerken, 2009, 2011, 2012).

Experiment 1

Experiment1 asked whether 9-month-olds would generalize from a single AAB (identical 1st and 2nd syllables) or ABA (identical 1st and 3rd syllables) syllable string to new strings (also see Marcus *et al*., 1999).

Methods

Participants

Participants in Experiment 1 were 20 infants (11 female) from English-speaking homes in the Tucson area. They ranged in age from 8 mos, 15 days to 9 mos, 4 days, with a mean of 9 mos, 2 days. All infants were at least 37 weeks to term, at least 5 lb 8 oz at birth, had no history of speech or language problems in their nuclear family, and were not given medication for an ear infection within one week of testing. Six additional infants participated but were excluded from analysis due to failure to complete the minimum number of test trials (five) and experimenter error (one). Half of the infants were assigned to the AAB familiarization condition and half to the ABA familiarization condition.

Materials

The familiarization and test strings for all experiments are shown in Table 1. In Experiment 1, all infants were familiarized with 16 AAB (always *leledi)* or ABA (always *ledile)* strings. However, because the familiarization strings only played for about 21 sec, they were preceded by about 1.5 min of Andean instrumental music to allow infants to become accustomed to the testing booth and procedure. The familiarization phase was followed by six AAB and six ABA test trials. Test trials comprised six 30 sec trials with the strings *kokoba* and *popoga* in two random orders (AAB test trials) and six trials with *kobako* and *pobapo* in two random orders (ABA test trials). Syllable strings for all conditions 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 three-syllable words, using speech analysis software.

Procedure

The headturn preference procedure (Kemler Nelson, Jusczyk, Mandel, Myers, Turk & Gerken, 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 2 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 2 consecutive seconds or the test trial reached its conclusion.

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 Figure 1. A *t*-test showed that infants listened significantly longer to inconsistent test items (mean = 9.89, $SE = 0.55$) than inconsistent items (mean = 10.76, $SE = 0.69$; $t(19) = 2.43$, $p < .05$). Thus, Experiment 1 showed a novelty preference.[2] Seven infants each in the AAB and ABA conditions showed this preference. Thus, infants were able to generalize from a single input type. Moreover, because a novelty preference has been argued to indicate that infants have learned the familiarization pattern very well (e.g. Hunter & Ames, 1988; Hunter, Ames & Koopman, 1983), the direction of preference might suggest that infants found the AAB or ABA pattern easy to encode and recognize based on one input type. To our knowledge, this is the first demonstration of generalization from such little input in the language domain. It is consistent with a similar demonstration in the visual domain by Quinn *et al.* (2002).

The importance of syllable repetition in prompting generalization might be due to the statistical rarity of repeated syllables in English and the model of English that underlies this rarity (Dawson & Gerken, 2011). Alternatively, repetition might be salient without any basis in prior experience (e.g. Endress, 2013). Experiment 2 asked whether other types of statistically rare, and therefore surprising, stimuli could elicit generalization as well.

Experiment 2

As noted above, syllable repetition in caregiver utterances is quite rare in English, and it is possible that the statistical rarity of this repetition, and the model of English that it supports, is what prompted infants in Experiment 1 to

[2]The current Experiment 1 in which infants were familiarized with 16 tokens of a single input type yielded a novelty preference, whereas the somewhat similar diagonal condition of Gerken (2006) yielded a familiarity preference. Several differences in the two studies might explain the difference in response. One possibility is that it is actually easier to generalize from fewer examples than from more examples. Other research in our lab suggests that this is the case (Gerken, in preparation). Therefore, the novelty preference seen in the current experiment might reflect the relative ease that infants had in generalizing from a single input example. However, a more mundane difference between the two experiments is that Gerken (2006) presented only four test trials, while the current studies employed 12. An examination of the first four test trials of Experiment 1 indeed shows a non-significant familiarity preference. Therefore, we cannot directly compare the current experiments and previously published ones.

Table 1 *Familiarization, test stimuli, shared features in familiarization and test, listening preference in all experiments*

Experiment	Familiarization stimuli	Test stimuli	Number of 'surprising' features in familiarization, and do test items contain all of the same surprising features from familiarization?	Listening preference at test
1	*leledi* or *ledile*	*kokoba*, *popoga*	1, yes	Novelty preference
2a	*lelezhi* or *lezhile*	*kokoba*, *popoga*	2, no	No preference
2b	*lelezhi* or *lezhile*	*kokozhi*, *popozhi*	2, yes	Familiarity preference
3	*leledi* or *ledile*	*kokodi*, *popodi*	1, yes	Novelty preference

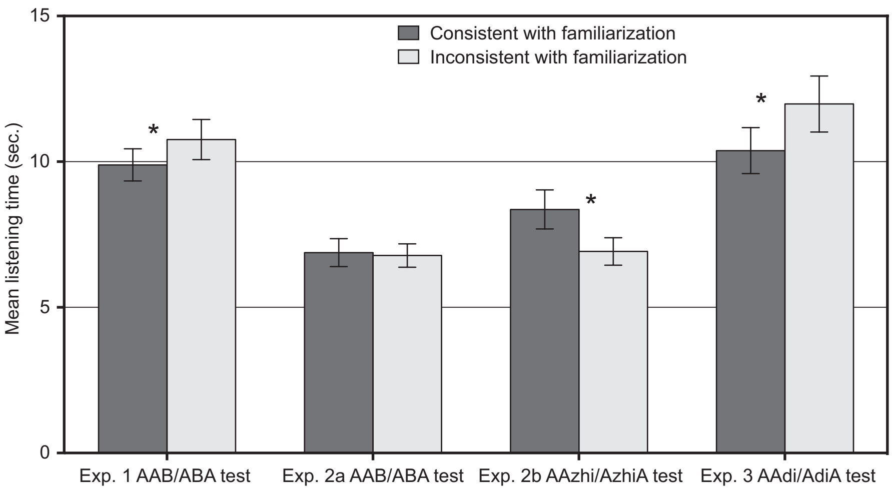

Figure 1 *Mean listening time and SE for infants in Experiments 1, 2a, 2b, and 3. Asterisks indicate a significant difference between listening times for consistent vs. inconsistent test items. See Table 1 for a list of familiarization and test stimuli.*

incorporate repetition into their generalization. That is, by incorporating repetition into the generalization (H), the surprising nature of the input is reduced. In Bayesian terms, P(D | H) / P(D) is greater than in the case where D is more probable overall. If the surprise of a rare event that is inconsistent with the current model of a domain prompts generalization from a single input example, then adding a second surprising feature to the AAB and ABA strings might cause infants to incorporate that feature into their generalization. In Experiment 2, the second surprising feature was the initial consonant on the B syllable of the AAB and ABA strings. In English, the consonant 'zh' as in 'vision' or 'decision' is very rare, occurring in fewer than 1% of words. As noted above, the consonant also only occurs in a restricted set of contexts. In contrast, the consonant 'd' that appeared in the B-syllable of Experiment 1 is very frequent. Thus, to directly contrast the strings in Experiment 2 with those in Experiment 1, we replaced the syllable 'di' from Experiment 1 with the syllable 'zhi'. We familiarized all infants with 16 *lelezhi* or *lezhile*. Note that these strings contain two surprising features: syllable repetition and the syllable 'zhi'. If learners attempt to explain what is surprising in their input, they should include both surprising features in their generalization. Therefore, at test, they should not generalize to test stimuli that reflect just one of those two features, but only generalize to test stimuli that reflect both features. This notion is similar to the one found by Gerken (2006, column condition), except that here, what is surprising about the stimuli is based entirely on infants' prior experience with English. Therefore, infants in Experiment 2a were tested on the same strings as in Experiment 1 (with only repetition), whereas infants in Experiment 2b were tested on stimuli that contained both repetition and 'zh'.

Methods

Participants

Participants in Experiments 2a and 2b were 40 infants (20 females) from English-speaking homes in the Tucson area. They ranged in age from 8 mos, 19 days to 9 mos, 27 days, with a mean of 9 mos, 18 days. The same inclusion criteria used in Experiment 1 were used in Experiment 2. Eleven additional infants participated but were excluded from analysis due to failure to listen through the familiarization phase (one), failure to complete the minimum number of test trials (eight), experimenter error (one), and parental interference (one). As in Experiment 1, half of the infants were assigned to the AAB familiarization condition and half to the ABA familiarization condition. Infants in Experiment 2a were tested on new AAB and ABA strings with no phonetic overlap with the familiarization strings, while infants in Experiment 2b were tested on new AAB and ABA strings in which the B syllable was the same for familiarization and test.

Materials

All infants were familiarized with 16 AAB (always *lelezhi)* or ABA (always *lezhile)* strings. As in Experiment 1, the familiarization strings were preceded by about 1.5 min of Andean instrumental music to allow infants to become accustomed to the testing booth and procedure. The familiarization phase was followed by six AAB and six ABA test trials. For infants in Experiment 2a, test trials comprised six 30 sec trials with the strings *kokoba* and *popoga* in two random orders (AAB test trials) and six trials with *bakoba* and *gapoga* in two random orders (ABA test trials). For infants in Experiment 2b, test trials comprised six 30 sec trials with the strings *kokozhi* and *popozhi* in two random orders (AAB test trials) and six trials with *kozhiko* and *pozhipo* in two random orders (ABA test trials). Syllable strings were generated as in Experiment 1.

Procedure

The procedure was identical to that of Experiment 1.

Results

As in Experiment 1, test trials were classified as consistent vs. inconsistent with the familiarization stimuli for each infant. Infants' listening times for consistent vs. inconsistent test trials are shown in Figure 1. A 2 consistency (consistent vs. inconsistent) × 2 test stimuli (2a vs. 2b) ANOVA showed the predicted consistency × test trial interaction ($F(1, 38) = 5.92$, $p < .05$. The interaction was followed up with separate *t*-tests for Experiments 2a and 2b. As predicted by the view that infants incorporated the surprising presence of the rare syllable 'zhi' into their generalization from the familiarization string, infants in Experiment 2a showed no hint of generalization at test (mean consistent = 6.88, $SE = 0.48$; mean inconsistent = 6.78, $SE = 0.40$; $t(19) < 1$, *ns*). In Experiment 2a, seven infants in the AAB condition and five infants in the ABA condition listened longer to consistent test items. In contrast, infants in Experiment 2b, whose test items contained the same two surprising features present in the familiarization stimuli, showed significant generalization at test (mean consistent = 8.36, $SE = 0.67$; mean inconsistent = 6.92, $SE = 0.47$; $t(19) = 2.60$, $p < .02$). In Experiment 2b, eight infants in the AAB condition and seven in the ABA condition listened longer to consistent test items. Thus, infants in Experiment 2b were able to generalize from a single input type, just as infants in Experiment 1 were.

It is important to note that Experiment 2 rules out an alternative interpretation of Experiment 1. We claimed that infants generalize the repeated syllables of the AAB or ABA because syllable repetition is surprising given their existing model of English morpho-syntax. They explain the surprising input by incorporating syllable repetition into their generalization. In contrast, it is possible that infants in Experiment 1 (and previous experiments that employ the AAB and ABA grammars) based generalization on the perceptual salience of syllable repetition (e.g. Gervain *et al.*, 2008; Murphy *et al.*, 2008). On this view, infants might have responded at test based on the perceptual salience of syllable repetition in Experiment 2a and the perceptual salience of syllable repetition plus the surprising nature of the syllable 'zhi' in Experiment 2b. However, this view would predict that generalization based on repetition should have been independent of generalization based on the surprising nature of 'zhi'. Therefore, infants in Experiment 2a should have shown the same novelty preference shown by infants in Experiment 1. The fact that they did not supports the view that infants treat the two surprising features in Experiment 2 as linked (not independent). Interestingly, infants in Experiment 2b listened longer to consistent test trials (a familiarity preference), whereas infants in Experiment 1 showed a novelty preference. We tentatively take the reversal in preference to indicate that noting and generalizing based on two surprising features of an input string is more difficult in some way than noting and generalizing from a single surprising input feature. The finding that generalizations that entail encoding multiple input features are more difficult is supported by other research from our lab and suggests an interesting line of future research (Quam & Gerken, in preparation).

In order to ensure that the switch in direction of preference was not an artifact of having more overlap between familiarization and test in Experiment 2b (repetition and 'zh') than Experiment 2a (just repetition), we conducted a third and final experiment in which infants were familiarized with the same input string as in Experiment 1 (*leledi* or *ledile*) but tested on strings that maintained the repetition pattern and the 'di' syllable.

Experiment 3

The goal of Experiment 3 was to provide further evidence that it was the presence of one surprising feature in Experiment 1 and two surprising features in Experiment 2b that caused infants to demonstrate a novelty and familiarity effect in these two experiments, respectively. Specifically, we sought to rule out the possibility that the switch in preference seen between Experiments 1 and 2b was due to the increased overlap between familiarization and test stimuli in the latter experiment.

Because 'd', which was the onset of the non-repeated syllable in Experiment 1, is a very frequently occurring consonant in English (about 5% of all onsets) and one that can occur in most morpho-phonological contexts, we anticipated that infants in Experiment 3 would not find it surprising and therefore would not incorporate it into their generalization at test. Therefore, we predicted that infants in Experiment 3 would show the same significant novelty preference demonstrated by infants in Experiment 1. Such a result would serve to bolster our interpretation of Experiment 1 vs. Experiment 2b: Encoding and generalizing based on two surprising input features is more difficult than generalizing based on a single input feature.

Methods

Participants

Participants in Experiment 3 were 20 infants (nine females) from English-speaking homes in the Tucson area. They ranged in age from 8 mos, 20 days to 9 mos, 21 days, with a mean of 9 mos, 1 day. The same inclusion criteria used in Experiments 1–2 were used in Experiment 3. Four additional infants participated but were excluded from analysis due to failure to complete the minimum number of test trials (one), parental interference (one), and total listening times more than 2 *SD* above the group mean (two). Half of the infants were assigned to the AAB familiarization condition and half to the ABA familiarization condition.

Materials

The familiarization phase was identical to that used in Experiment 1. It was followed by six AAB and six ABA test trials. The test trials were identical to Experiments 1–2 in presentation format, but now contained the strings *kokodi* and *popodi* in two random orders (AAB test trials) and *kodiko* and *podipo* in two random orders (ABA test trials). Syllable strings were generated as in Experiments 1–2.

Procedure

The procedure was identical to that of Experiments 1–2.

Results

As in Experiments 1–2, test trials were classified as consistent vs. inconsistent with the familiarization stimuli for each infant. Infants' listening times for consistent vs. inconsistent test trials are shown in Figure 1. As predicted, the data from Experiment 3 replicated the novelty preference seen in Experiment 1 (mean consistent = 10.38, $SE = 0.79$; mean inconsistent = 11.98, $SE = 0.96$; $t(19) = 2.11$, $p < .05$). Eight infants in the AAB condition and seven infants in the ABA condition showed this preference. As a parallel to the statistical analysis of Experiments 2a and 2b, we performed a 2 consistency (consistent vs. inconsistent) × 2 test stimuli (Experiment 1 with completely new test syllables vs. Experiment 3 with non-repeated syllable at test matching familiarization) ANOVA on the data from Experiments 1 and 3. As predicted, there was only a main effect of consistency ($F(1, 38) = 8.71$, $p < .01$). Neither the main effect of test stimuli nor the interaction were significant (Fs < 1).

Discussion

The results of the three experiments presented here begin to paint a picture of very rapid generalization by 9-month-old infants. Experiment 1 showed that infants were able to generalize to new input types from multiple tokens of a single input example. As noted above, this is the first time, to our knowledge, that generalization from a single input example has been found in the domain of language. The finding is consistent with work on infant generalization of visual patterns, in which 6- to 7-month-olds are able to generalize from a single display of shapes in columns to a new display of vertically oriented bars (Quinn *et al*., 2002).

Experiment 2 explored whether the rapid generalization observed in Experiment 1 occurred because syllable repetition in English is rare and surprising based on an underlying model of language in which syllables are organized into words and sentences with almost no repeated syllables except for grammatical morphemes. In both Experiments 2a and 2b, infants were familiarized with an input string that contained two rare features: syllable repetition as in Experiment 1, and a syllable that began with the rare consonant 'zh'. This consonant typically occurs in derivational morphemes and never occurs at the beginning of a stressed syllable in English. Therefore, its presence in our stimuli should have been surprising based on infants' growing model of English morpho-phonology. In Experiment 2a, the test strings contained only one of the surprising features (repetition), while in Experiment 2b, the test strings contained both syllable repetition and 'zh'. Only the infants in Experiment 2b generalized at test, strongly suggesting that infants did not treat these two features as independent, but rather linked, in their generalization. Furthermore, while infants in Experiment 1, who only generalized on a single feature, showed a novelty preference at test, infants in Experiment 2b, who generalized based on two features showed a familiarity preference at test. The difference in generalization based on one vs. two input features has been replicated in other research in our lab (Quam & Gerken, in preparation).

Although our proposal that the number of input features entailed in a generalization affects infant responses to test stimuli must remain tentative, we did rule out one obvious artifact in Experiment 3. Here, infants were familiarized with the same strings as in Experiment 1, which contained only a single surprising feature (repetition). The test items, however, contained both repetition and the *un*surprising syllable 'di', which had also occurred in the familiarization stimuli. If infants in Experiment 2b showed a familiarity preference simply because there was more overlap between familiarization and test strings, they should have shown a similar familiarity preference in Experiment 3. However, the fact that infants showed a novelty preference in Experiment 3, as they had in Experiment 1, further bolsters our proposal that the number of surprising features, not the total amount of overlap between familiarization and test, determines whether infants respond with a novelty or familiarity preference.

One question that remains from this research concerns the circumstances under which infants are able to generalize from a single input example (the current experiment and Quinn *et al.*, 2002) and when no generalization is observed for fewer than three input examples (e.g. Gerken & Bollt, 2008; Needham *et al.*, 2005; Quinn & Bhatt, 2005). At the very least, the data presented here suggest that generalization from a single input example occurs when that input is either surprising given the learner's current model of a domain or is perceptually salient. But what of the cases where more input is required for generalization? We noted two possible reasons in the Introduction, which we will elaborate on here. One possibility concerns learning via induction in which a basis or bases of generalization emerge as more input is encountered. On this view, input examples are stored in a multidimensional space and clusters of similar inputs suggest which dimensions are signal and which are noise. This view is consistent with Bhatt and Quinn's suggestion that more input is required to 'teach' 3- to 4-month-olds the relevant features for generalization. Although this view can account for situations in which learners require more than one input, it doesn't account well for situations like the one presented here, in which learners generalize from a single input.

The other possible reason that learners might require more than a single input item concerns learning by abduction, which is consistent with the Bayesian hypothesis selection proposal that we are entertaining here. On this view, learners might consider a relevant basis or bases of generalization from a single input, but the generalization (hypothesis) is not sufficiently strong to drive behavior (e.g. Gerken, 2006, 2010; Peterson, 2011; Xu & Tenenbaum, 2007a). For example, when a child is given *Fep* as a label for a spotted dog, the child might consider both dogs and spotted dogs as possible referents for *Fep*. However, upon hearing the same label applied to three different spotted dogs, the hypothesis that *Fep* refers to spotted dogs is weighted heavily enough to yield consistent behavior (Xu & Tenenbaum, 2007a). Similarly, if an infant hears the string *leledi*, she might consider several possible bases of generalization including syllable repetition, the presence of the syllable 'di', etc. Because repetition is surprising on the learner's existing model of English, a hypothesis based on surprise is weighted heavily, because this hypothesis decreases surprise (Griffiths & Tenenbaum, 2007). As we saw, infants are able to generalize from a single input example in such situations. However, Gerken (2006) found that infants were also able to generalize based on the presence of the syllable 'di' when they encountered four examples containing this syllable (e.g. *leledi*, *wiwidi*, etc.). One way to think about the change in generalization from a single input example to four examples is that, if these input examples were *not* generated by a rule that contained the syllable 'di', the examples constitute a suspicious coincidence (Tenenbaum & Griffiths, 2001). That is, getting four examples containing 'di' in a row is surprising under one hypothesis, and not surprising under the hypothesis that includes both repetition and 'di'. Thus, the latter hypothesis is weighted heavily enough to drive behavior. Although this surprise reduction framework can explain why generalization can sometimes occur from one example and other times with three to four examples, we must ultimately begin to predict what input will be surprising to learners in particular developmental stages.

In summary, the results presented here clearly demonstrate that infants can generalize to new input types from multiple tokens of a single input example. They further suggest that surprise, either in the form of the rarity of raw statistics or in terms of the learner's current model of a domain, can be one factor that drives generalization.

Acknowledgements

This research was supported by NICHD #R01 HD042170 and NSF 0950601 to LAG.

References

Aslin, R.N. (2011). Perceptual organization of visual structure requires a flexible learning mechanism. *Infancy*, **16** (1), 39–44.

Bhatt, R.S., & Quinn, P.C. (2011). How does learning impact development in infancy? The case of perceptual organization. *Infancy*, **16** (1), 2–38.

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

Dawson, C., & Gerken, L.A. (2009). Learning to learn differently: the emergence of domain-sensitive generalization in the second six months of life. *Cognition*, **111**, 378–382.

Dawson, C., & Gerken, L.A. (2011). When global structure 'explains away' evidence for local grammar: a Bayesian account of rule induction in tone sequences. *Cognition*, **120** (3), 350–359.

Dawson, C., & Gerken, L.A. (2012). Can rational models be good accounts of developmental change? The case of language development at two time scales. In F. Xu & T. Kushnir (Eds.), *Rational constructivism in cognitive development* (pp. 95–124). Oxford: Elsevier Publishing.

Endress, A.D. (2013). Bayesian learning and the psychology of rule induction. *Cognition*, **127**, 159–176.

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. (2010). Infants use rational decision criteria for choosing among models of their input. *Cognition*, **115** (2), 362–366.

Gerken, L.A. (in preparation). Less really is more: infants learn better from fewer input examples.

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.

Gerken, L.A., Wilson, R., & Lewis, W. (2005). 17-month-olds can use distributional cues to form syntactic categories. *Journal of Child Language*, **32**, 249–268.

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences, USA*, **105** (37), 14222–14227.

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.

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

Gómez, R.L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency. *Infancy*, **7** (2), 183–206.

Griffiths, T.L., & Tenenbaum, J.B. (2007). From mere coincidences to meaningful discoveries. *Cognition*, **103** (2), 180–226.

Gweon, H., Tenenbaum, J.B., & Schulz, L.E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences, USA*, **107** (20), 9066–9071.

Hunter, M., & Ames, E. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, **5**, 69–95.

Hunter, M., Ames, E., & Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. *Developmental Psychology*, **19** (3), 338–352.

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.

Lynch, M.P., & Eilers, R.E. (1992). A study of perceptual development for musical tuning. *Perception & Psychophysics*, **52** (6), 599–608.

Marcus, G.F., Fernandes, K., & Johnson, S.P. (2007). Infant rule learning facilitated by speech. *Psychological Science*, **18** (5), 387–391.

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

Murphy, R.A., Mondragon, E., & Murphy, V.A. (2008). Rule learning by rats. *Science*, **319** (5871), 1849–1851.

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

Peirce, C.S. (1935). Pragmatism and abduction. In C. Hartshorne (Ed.), *Collected papers of Charles Sanders Peirce* (pp. 112–135). Cambridge, MA: Harvard University Press.

Peterson, M.A. (2011). Variable exemplars may operate by facilitating latent perceptual organization. *Infancy*, **16** (1), 52–60.

Quam, C.M., & Gerken, L.A. (in preparation). Exploring the role of variability in infant attention to relevant vs. irrelevant dimensions of language.

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

Quinn, P.C., Bhatt, R.S., Brush, D., Grimes, A., & Sharpnack, H. (2002). Development of form similarity as a Gestalt grouping principle in infancy. *Psychological Science*, **13** (4), 320–328.

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

Temperley, D. (2008). A probabilistic model of melody perception. *Cognitive Science*, **32** (2), 418–444.

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

Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition*, **112**, 97–104.

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

Xu, F., & Kushnir, T. (Eds.) (2012). *Rational constructivism in cognitive development*: Oxford: Elsevier Publishing.

Xu, F., & Tenenbaum, J.B. (2007a). Sensitivity to sampling in Bayesian word learning. *Developmental Science*, **10**, 288–297.

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

Received: 15 January 2013
Accepted: 5 February 2014