



## SHORT REPORT

# Infant learning is influenced by local spurious generalizations

LouAnn Gerken and Carolyn Quam

Department of Psychology, The University of Arizona, USA

### Abstract

*In previous work, 11-month-old infants were able to learn rules about the relation of the consonants in CVCV words from just four examples. The rules involved phonetic feature relations (same voicing or same place of articulation), and infants' learning was impeded when pairs of words allowed alternative possible generalizations (e.g. two words both contained the specific consonants p and t). Experiment 1 asked whether a small number of such spurious generalizations found in a randomly ordered list of 24 different words would also impede learning. It did – infants showed no sign of learning the rule. To ask whether it was the overall set of words or their order that prevented learning, Experiment 2 reordered the words to avoid local spurious generalizations. Infants showed robust learning. Infants thus appear to entertain spurious generalizations based on small, local subsets of stimuli. The results support a characterization of infants as incremental rather than batch learners.*

### Research highlights

- Infants failed to learn a language rule from a randomly ordered list but learned when the list was reordered to avoid local spurious generalizations.
- Infants appear to be incremental learners who are influenced by local subsets of the input and therefore may not always benefit from global variation in the input.
- Randomly ordered input can contain unintended information that can promote or impede infant learning.

### Introduction

In laboratory experiments, infants have been shown to learn a variety of language-like rules in a very short time, including rules involving the possible relations among speech sounds in a word (i.e. phonotactics; Chambers, Onishi & Fisher, 2003, 2011; Gerken & Knight, 2015; Saffran & Thiessen, 2003; Seidl, Onishi & Cristia, 2014; Wang & Seidl, 2014). Recently, our laboratory demonstrated that 11-month-olds were able to learn one of two phonotactic rules about the relation of two consonants

in consonant-vowel-consonant-vowel (CVCV) non-words from just one token each of four input words, spanning a total of 4.5 sec (Gerken & Knight, 2015). The rules were either that C1 and C2 had to share the same voicing value (either both voiced or voiceless; Voicing Rule) or the same place of articulation (both labial or both coronal; POA Rule).

One interesting observation from this study was that the particular subset of the input that infants heard during pretest affected their responses at test (see Table 1 for Gerken & Knight's stimuli). Different sets of pretest stimuli had been designed to provide different amounts of support for an alternative phonotactic rule: no support (Non-Conflicting Condition), partial support (Partially Conflicting Condition), or full support for the alternative rule (Completely Conflicting Condition). Infants in the Non-Conflicting Condition heard four words in which the C1s and C2s were all different, so that the primary rule (shared POA or shared voicing between the two consonants) was the only linguistically relevant rule present in the stimuli. Infants in the Partially Conflicting Condition also heard four words that all obeyed the relevant rule, but, in addition, pairs of words shared the same C1s and C2s (p/b and d/s for the POA Rule) and the same vowels (o and æ for the POA Rule).

**Table 1** Pre-test stimuli in actual presentation order for the Place of Articulation Rule in three conditions (Gerken & Knight, 2015). POA Test items were: *deta, tæda, bifa, pova, tida, befa, pæva, dota*; Voicing Rule test items were: *deba, tæfa, bida, posa, tifa, bæda, pæsa, doba*. All infants heard all test items on different trials

Non-Conflicting Condition Novelty Preference at test	Partially Conflicting Condition Familiarity Preference at test	Completely Conflicting Condition No Preference at test
desa	poba	pæba
poba	dosa	pæba
bipa	dæsa	poba
tæza	pæba	piba

For infants in the Completely Conflicting Condition, all four words again obeyed the relevant rule, but *all* C1s and C2s were the same (p/b for the POA Rule) and could therefore fully support an alternative rule. Infants in the Non-Conflicting Condition showed a strong novelty preference. Infants in the Partially Conflicting Condition showed a familiarity preference of approximately equal magnitude. Infants in the Completely Conflicting Condition showed no hint of learning at test.

A number of infant cognition studies indicate that robust pre-test generalization causes infants to attend more to the non-conforming test items than to the conforming test items (i.e. show a novelty preference), while less robust pre-test generalization causes infants to attend more to the conforming test items (familiarity preference) (Aslin, 2007; Gerken, Dawson, Chatila & Tenenbaum, 2015; Hunter & Ames, 1988; Hunter, Ames & Koopman, 1983). Based on these prior studies, Gerken and Knight interpreted the pattern of test behavior in their study to indicate that learning of the relevant rule in the Non-Conflicting Condition was most robust, and learning in the Partially Conflicting Condition was less robust. They further suggested that infants in the Completely Conflicting Condition actually learned a rule about particular consonants instead of the relevant Voicing or POA rules. If we for the moment accept this general line of thinking, two questions arise – one more theoretical and one more methodological.

The theoretical question concerns how to interpret infants' pattern of test behavior in the Non-Conflicting vs. Partially Conflicting Conditions. In both conditions, the full set of four input words was consistent only with the relevant rule. Therefore, if infants' preference at test was based on the generalization that described the entire input set, they should have responded similarly in both conditions. There are three explanations for infants' apparently less robust generalization in the Partially

Conflicting Condition. The explanation offered by Gerken and Knight (2015) is that pairs of input words shared properties (e.g. two words that contained the same consonants such as *dosaldæsa*; Table 1), which allowed *local spurious generalizations*. These spurious generalizations might have interfered with the relevant generalization that could be made across the entire set of input.

The second and third explanations for infants' apparently less robust generalization in the Partially Conflicting Condition both concern the fact that the set of four words exhibited less variability on the irrelevant (particular consonants) dimension than did the words in the Non-Conflicting Condition (e.g. only two C1s and two C2s instead of four of each). This lack of variability might have caused less robust generalization (e.g. Thiessen & Pavlik, 2013). The *local variability* version of this explanation assumes that infants consider only small subsets of input at a time, and if variability across the irrelevant dimension (e.g. particular consonants) in the subset is lower, generalization on the relevant dimension (e.g. the feature voicing) is less robust. The *global variability* version of the explanation asserts that learners are influenced by variability on irrelevant stimulus dimensions over the entire input set. Importantly, the global version of the variability explanation predicts that a more varied input set should allow more robust generalization.

Both the local spurious generalization and local variability explanations of infants' less robust generalization in the Partially Conflicting Condition of Gerken and Knight (2015) are consistent with infants being *incremental* learners, who update the weight of their generalizations as they encounter each new input. In contrast, the global variability explanation is consistent with a *batch learner*, who only generalizes after encoding all of the input examples (see Yu & Smith, 2012, for further discussion of incremental vs. batch language learners). The experiments reported here contrast an incremental learner who is influenced by local stimulus properties with a batch learner who is influenced by global stimulus variability. The latter should demonstrate more robust generalization when the input is more globally varied on irrelevant dimensions than the input presented to infants in the Partially Conflicting Condition of Gerken and Knight.

The methodological question raised by the Gerken and Knight (2015) findings concerns what these findings say about standard infant artificial language learning paradigms, in which a long sequence of stimuli is presented in random order. On the one hand, being presented with more input and more varied input (than those used by Gerken and Knight) should help learners

who are able to respond based on a larger input set to rule out particular surface properties in favor of the relevant abstract phonetic featural properties. On the other hand, when a larger number of input items is selected randomly, the likelihood increases that spurious local generalizations will arise. The full stimuli selected by Gerken and Knight came from a larger set of 32 words (4 C1s each paired with 2 C2s = 8 C1/C2 sequences each paired with 4 vowels), which were created for a more general line of research concerning the amount of input required for learning. From this set of 32 words, eight words were held out for test items (the same eight test words used by Gerken and Knight, see Table 1), leaving 24 possible pre-test words.

To see how many spurious local generalizations might arise from randomly ordering these 24 words, we followed the usual procedure in our lab for creating a random sequence of pre-test words. We assigned each of the words a random number, using the RAND function of Microsoft Excel, and we assigned the same random order for the list of words for each rule. We then reordered the words from both rules based on their random number (lowest to highest) to give us the pre-test

stimulus order. The randomly ordered words appear in Table 2. This random order resulted in eight adjacent pairs of words for each rule, many of which shared surface properties (one pair shared C1s: bæfa/bopa; two shared C1/C2, e.g. toda/tæda; one shared C1/V: dæsa/dæta; and five shared vowels, e.g. dæsa/pæba). These randomly ordered pre-test stimuli were used in Experiment 1 of the present study. Of course, if we looked at longer strings of adjacent words, there would be more possible overlap (e.g. word 1 and word 3 of a 3-word string). However, we have restricted our analysis to adjacent words, since we know that 12-month-olds (who are just older than the age tested here) can track adjacent relations (Gómez & LaKusta, 2004) but cannot yet track nonadjacencies (Gómez & Maye, 2005) without additional scaffolding (Lany & Gómez, 2008).

To foreshadow, we employed the randomly ordered words just described in Experiment 1 and found no evidence of learning. Therefore, we reordered the words in Experiment 2 to eliminate local spurious generalization between adjacent words. In the latter experiment, infants showed a strong novelty preference.

**Table 2** Ordered stimuli for the Place of Articulation (POA) rule in Experiments 1 and 2. In Experiment 1, where stimuli were randomly ordered, adjacent pairs of words share rule-irrelevant (spurious) properties

Random Order Exp. 1	Words N & N-1 shared C1s in Random Order	Words N & N-1 shared V1s in Random Order	Words N & N-1 shared C2s in Random Order	Avoid Spurious Generalization Order Exp. 2
piva				piva
bofa				teza
dæsa				bofa
pæba		ɛ		tiza
bæpa				bopa
toza				tæza
peva				bipa
toda				toza
tæda	t		d	bepa
pæba				dæta
dita				peva
poba				dita
bæfa	b			tæda
bopa				bæfa
piba				toda
tiza		i		peba
bipa		i		disa
dosa				poba
teza				dæsa
bepa		ɛ		piba
disa				dæsa
dæsa	d		s	pæba
dæta	d	æ		dosa
tæza		æ		bæpa

## Experiment 1

If more input and more varied stimuli promote generalization, infants should generalize at least as well from the 24 randomly ordered words in Experiment 1 as they did from the four words in the Non-Conflicting Condition of Gerken and Knight (2015). However, if infants are affected by local spurious generalizations, no matter how large and varied the input set, infants in Experiment 1 might show poorer learning than infants in the Non-Conflicting Condition of the previous experiment, either showing a familiarity preference at test or no preference.

### Methods

#### Participants

Participants were 20 infants (11 females) from English-speaking homes, ranging in age from 10.6 to 11.8 mos, with a mean of 11.1 mos. We chose to study 11-month-olds because this was the age studied by Gerken and Knight (2015), and because our task involves learning new phonotactic generalizations, which previous studies have found in 9-month-olds (Saffran & Thiessen, 2003) and 16-month-olds (Chambers *et al.*, 2003). Pilot testing by Gerken and Knight with 9-month-olds on the four words from the Non-Conflicting Condition revealed learning only in girls, therefore, we moved to a slightly older age group.

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 not given medication for an ear infection within one week of testing. Two additional infants were tested but not included because their mean listening time was more than  $2SD$  above the group mean ( $N = 1$ ), or because they had fewer than 10 useable trials once trials longer than  $2SD$  above the group mean for that trial were excluded ( $N = 1$ ).

### Materials

Materials for Experiment 1 were 24 pre-test words generated by the Voicing Rule for one group of infants and 24 words generated by the POA Rule for the other group. Both sets of 24 words were placed in the same random order in the manner described in the Introduction and therefore had the same number and type of spurious local generalizations. Test words were the same 16 words used by Gerken and Knight (2015; see Table 1): POA Test items were: *deta, tæda, bifa, pova, tida, befa, pæva, dota*. Voicing Rule test items were: *deba, tæfa, bida, posa, tifa, beda, pæsa, doba*. All infants heard all test items on different trials. Two different orders of test words were created for each rule, yielding four total test trials. These four trials were repeated across three blocks, for a total of 12 test trials.

### Procedure

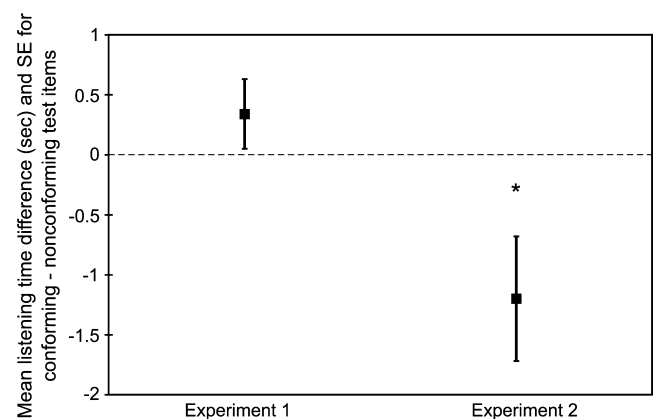
The headturn preference procedure (Kemler Nelson, Jusczyk, Mandel, Myers, Turk *et al.*, 1995) was used. Infants were seated on a parent's lap in a small room. The parent listened to pop music through headphones in order to mask the stimuli heard by the infants and prevent inadvertent influence on the infant. During the pre-test phase, during the presentation of the auditory stimuli, 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 pre-test stimulus, which played uninterrupted to its conclusion. In this stage there was no correspondence between infants' looking behavior and the stimuli. Because the 24 pre-test words only played for about 28 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.

After the pre-test sequence ended, the test phase began immediately. The flashing lights behaved in 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 (each with a duration of about 23.5 sec), continuing until either the infant looked away for 2 consecutive seconds or the test trial reached its conclusion. In keeping with the standard practice in our lab, test trials shorter than 2 seconds were excluded from the analysis because it was unlikely that infants were able to fully encode the nature of the stimulus on such a short trial (e.g. Gerken & Knight, 2015).

### Results and discussion

Eight test trials from seven infants (out of 240 total trials) were excluded from the analysis because they were longer than  $2SD$  above the group mean for that trial. From the remaining trials, each infant's mean for test trials that conformed vs. failed to conform to their pre-test rule was calculated, and these means were subjected to a 2 rule-type (Voice vs. POA)  $\times$  2 test-item-conformity (conforming vs. nonconforming with pre-test words) ANOVA. Neither of the two main effects nor the interaction approached significance (all  $F_s < 1$ ; Mean conforming = 6.42 ( $SE = .35$ ); Mean nonconforming = 6.08 ( $SE = .30$ )). (See Figure 1.)

Infants in Experiment 1 failed to show any sign of generalizing the Voicing or POA rule. Thus at the global level (the entire input set), more input and more varied input does not seem to help generalization. The result of Experiment 1 is surprising given the robust generalization



**Figure 1** Mean differences in listening times to test stimuli that conform vs. fail to conform to pre-test items in Experiment 1 (randomly ordered stimuli) and Experiment 2 (stimuli ordered to avoid spurious generalizations between adjacent pairs of stimuli).



observed in the infants in the Non-Conflicting and Partially Conflicting Conditions of Gerken and Knight (2015; see Table 1), who were familiarized with only four of the 24 words employed in Experiment 1. However, the results are consistent with an incremental learner who is influenced by local stimulus properties.

An alternative explanation for infants' failure to generalize in Experiment 1 is that they became fatigued listening to 24 pre-test words and were unable to demonstrate any learning at test. Similarly, the processing load entailed in discovering a generalization in a set of 24 words might be greater than in a set of four words. The mean listening times in test trials were about 0.5 sec shorter in the current experiment than in the Non-Conflicting and Partially Conflicting Conditions from Gerken and Knight (2015), perhaps suggesting that infants were more fatigued at test. However, other researchers have found learning with similar numbers of pre-test words. For example, Chambers *et al.* (2003) presented 25 words, each repeated six times.

To further test the view that infants are influenced by local stimulus properties, it is important to determine whether there is something about the 24 stimuli used in Experiment 1 that prevented infants from demonstrating learning. In particular, we hypothesized that the random ordering of the 24 words created spurious potential generalizations based on similarities between adjacent words (see Table 2). Therefore, Experiment 2 reordered the stimuli from Experiment 1 to avoid adjacent words sharing vowels or consonants.

## Experiment 2

If it was the random order of the 24 words in Experiment 1 that prevented infants from demonstrating generalization, and not the number of words per se, then reordering the words in Experiment 2 should allow generalization. If it is possible to interpret direction of preference at test in experiments such as these, we might predict a novelty preference specifically. A novelty preference would match the finding in Gerken and Knight's Non-Conflicting Condition, which was similar to the present experiment in that segmental overlap was also avoided (Table 1). Thus, in both of these cases, spurious possible generalizations are minimized by the stimulus design, maximizing the chances that infants will find the correct generalization (the Voicing or POA rule). However, any direction of preference will at least demonstrate that infants are able to generalize from the 24 pre-test words of Experiment 1 if the order is changed, in contrast to the lack of learning shown in Experiment 1.

## Methods

### Participants

Participants were 20 infants (10 females) from English-speaking homes, ranging in age from 10.3 to 11.8 mos, with a mean of 11.3 mos. Infants met the same inclusion criteria used in Experiment 1. Three additional infants were tested but not included because they were fussy or cried ( $N = 1$ ), their mother said they were having an unusual day ( $N = 1$ ), or because they had fewer than 10 useable trials once trials longer than  $2SD$  above the group mean for that trial were excluded ( $N = 1$ ).

### Materials

Materials for Experiment 2 were 24 pre-test words generated by each of the Voicing and POA rules, ordered such that no adjacent word pair shared vowels or consonants (see Table 2). Test words and test block structure were identical to those of Experiment 1. Although it is possible that 11-month-olds are sensitive to spurious local generalizations spanning a larger window than two input items, we focused on adjacent items in Experiment 2.

### Procedure

The procedure was identical to that used in Experiment 1.

## Results and discussion

As in Experiment 1, eight test trials from seven infants (out of 240 total trials) were excluded from the analysis because they were longer than  $2SD$  above the group mean for that trial. Infants' mean listening times in seconds to test trials that conformed vs. failed to conform to their pre-test rule were subjected to a 2 rule-type (Voice vs. POA)  $\times$  2 test-item-conformity (conforming vs. nonconforming with pre-test words) ANOVA. There was a significant main effect of conformity ( $F(1, 18) = 5.15, p < .04; d = 0.54$ ), such that infants listened longer to non-conforming test items (mean = 7.96 ( $SE = .55$ )) than conforming test items (mean = 6.76 ( $SE = .45$ )), indicating a novelty preference (see Figure 1). The interaction of rule-type and test-item-conformity did not approach significance, suggesting that infants learned both the POA and the voicing rules ( $F(1, 18) = 1.05, p < .40$ ). There was a main effect of rule type, such that infants familiarized with the POA Rule (mean listening time = 8.29,  $SE = 0.51$ ) showed longer listening times than infants familiarized with the

Voicing Rule (mean = 6.42,  $SE = 0.43$ ;  $F(1, 18) = 6.06$ ,  $p < .03$ ;  $d = 0.88$ ).

To compare the two experiments reported here, a 2 Experiment (1 vs. 2)  $\times$  2 test-item-conformity (conforming vs. nonconforming) ANOVA was performed on the combined data. Given the complete lack of generalization in Experiment 1 and the significant generalization in Experiment 2, we expected to find an interaction. The main effect of conformity was not significant ( $F(1, 38) = 2.11$ ,  $p > .15$ ). As expected, there was an interaction between Experiment and conformity ( $F(1, 38) = 6.71$ ,  $p < .02$ ), such that only infants in Experiment 2 demonstrated a significant listening difference between conforming and non-conforming test trials. There was also a main effect of Experiment, such that the mean listening times in Experiment 2 (mean = 7.36,  $SE = 0.37$ ) were significantly longer than Experiment 1 (mean = 6.25,  $SE = 0.37$ ;  $F(1, 38) = 4.52$ ,  $p < .05$ ,  $d = 0.58$ ).

Infants in Experiment 2 demonstrated generalization at test after receiving the same 24 words that did not permit infants in Experiment 1 to demonstrate generalization.

## General discussion

Infants in Experiment 1 failed to generalize from a set of 24 input words generated by either the Voicing or POA Rule. This result is surprising, given that the infants studied by Gerken and Knight (2015) generalized the same rules from just four input words, which were included in the 24 words presented in Experiment 1. However, the result was predicted by the view that infants are incremental learners who are influenced by local stimulus properties, which resulted from the randomly ordered stimuli in Experiment 1. To ensure that it was the order of input and not the input more generally that led to the generalization failure observed in Experiment 1, Experiment 2 presented infants with the same 24 words reordered to avoid adjacent words containing spurious local generalizations. The reordered familiarization resulted in significant generalization, lending support to the view that local spurious generalizations make learning the relevant rule more difficult.

The current data have at least one theoretical implication for understanding language learning. As noted in the Introduction, models of language learning can either assume a batch learner, who is able to store many input examples before settling on a preferred generalization, or an incremental learner whose generalizations are based on only a few input items at a time (e.g. Yu & Smith, 2012). Such an incremental learner is consistent with the proposal made by Gerken and Knight that infants are

influenced by local spurious generalizations involving shared segments between adjacent familiarization words. On this view, local unintended similarities between adjacent or near-adjacent input items, which are irrelevant to the intended generalization, impede learning, perhaps because the surprise of encountering adjacent similar items causes an implicit search for a new (incorrect) generalization (e.g. Gerken *et al.*, 2015).

The view that encountering adjacent similar items affects learning and generalization is consistent with data from a number of studies. For example, adult artificial language learning experiments have demonstrated that local invariance in structure that is *relevant* to the learning task helps learners find word forms. Even when the overall learning sets contain identical exemplars, a training set that is designed to provide local repetition of a particular word form (in different phrasal contexts) leads to better learning than a training set without such local overlap (Onnis, Waterfall & Edelman, 2008). Developmental research further suggests that such local overlap might promote some aspects of language learning (Waterfall, 2006).

Waterfall and colleagues' facilitative effects of local overlap on a relevant dimension are consistent theoretically with the present investigations. However, importantly, the current experiments and the ones by Gerken and Knight (2015) manipulate variability on *irrelevant* dimensions. The relevant dimension is always invariant, since all words in both sets conform to the voicing or POA rule. In Gerken and Knight's Non-Conflicting Condition, the *only* similarity (invariance) among adjacent items was the relevant feature (Voicing or POA). Adjacent items always varied (differed) on irrelevant dimensions. In contrast, Gerken and Knight's Completely Conflicting and Partially Conflicting Conditions and the current Experiment 1 introduced similarity – i.e. removed variability – on irrelevant dimensions between adjacent items (which could overlap in their segments).

In another line of previous research that is germane here, one machine-learning model of word learning depends on the recurrence of a word form and its potential referent within the same Short Term Memory (STM) window (Roy & Pentland, 2002; see below for more discussion of window size). Some recurring form-reference pairs created in the STM reflect spurious generalizations (e.g. 'the' or 'yeah' connected to a dog-shaped referent). These spurious generalizations must be further analyzed and discarded in long-term memory that takes into account variability (e.g. 'the' co-occurs with a number of other potential referents in addition to dog-shaped ones). The important aspect of this research to consider here is that local *similarity* initiates the generalization process, with some generalizations being

correct and others spurious. The current study adds to prior work by Waterfall and colleagues and by Roy and Pentland (2002) by intentionally manipulating the degree of local *irrelevant* variability and demonstrating its impact on infants' learning.

In focusing on local spurious generalizations, we are considering both relevant and irrelevant *similarity* between adjacent (or nearby) items and how these affect the likelihood that learners will make the relevant generalization. However, as noted in the Introduction, we can also view generalization problems in the current Experiment 1 and in Gerken and Knight's two conflicting conditions as stemming from a lack of local *variability* (e.g. Thiessen & Pavlik, 2013), which is particularly relevant for an incremental learner. Future research is needed to determine whether similarity and variability are simply two expressions of the same concept, and if they are not, what the role of each is for generalization.

To explain the current results, we suggest that infants make generalizations based on small subsets of adjacent data. If the local subset supports a generalization that is not correct for the entire set of input, infants may or may not be able to recover and find the generalization that can describe the full input set. We cannot determine from the current experiments the size of the window in which infants detect local spurious generalizations, since we only focused on adjacent input words when eliminating possible spurious generalizations in Experiment 2. However, the fact that infants in the Partially Conflicting Condition of Gerken and Knight (2015) showed generalization, even though all three adjacent word pairs demonstrated local spurious generalizations, might be taken to suggest that infants are looking for potential generalizations within a window of three adjacent items. The first two POA Rule input words from the Partially Conflicting Condition of Gerken and Knight (see Table 1) shared a vowel (*pobaldosa*). The second pair did not share a vowel but rather C1 and C2 (*dosaldæsa*), and the final pair shared a different vowel from the first pair (*dæsalpæba*). The fact that infants generalized but showed a familiarity preference for this set of four input items (compared with the novelty preference shown by infants in the Non-Conflicting Condition) might be taken to suggest that infants in the Partially Conflicting Condition were considering spurious generalizations based on adjacent words (e.g. the shared vowel in *pobaldosa*), but were able to rule them out when the next word negated the generalization based on the previous two. In other words, across a window of three words, *pobaldosaldæsa* did not share any properties but the relevant POA Rule. The size of infants' processing window and the manner in which local spurious gener-

alizations are considered and rejected is clearly a subject for future research (also see Roy & Pentland, 2002).

If, as suggested above, infants *do* consider a window of three adjacent items, their failure to generalize in Experiment 1 might be due partially or entirely to the last four items in the randomly ordered list. Words 21–23 in both POA and Voicing Rule lists shared C1 (*disa*, *dæsa*, *dæta*), and words 22–24 shared a vowel (*dæsa*, *dæta*, *tæza*).<sup>1</sup> Whatever the correct explanation for infants' failure to generalize in Experiment 1 and their success in Experiment 2, it must address how local spurious generalization or lack of variability can cause generalization failure even when the input set as a whole supports the relevant generalization.

We should also point out that the local spurious generalizations we believe prevented generalization in Experiment 1 might be based on shared phonological features as well as shared consonants and vowels. For example, the three adjacent words *poba*, *bæfa*, *bopa* from the POA Rule of Experiment 1 share labial stops as C1 and labial consonants (stops or fricatives) as C2. The ordering of words in Experiment 2 reduced such feature-based local spurious generalizations compared with Experiment 1, such that the mean shared features between adjacent word pairs was 3.70 in Experiment 1 vs. only 0.65 in Experiment 2. Again, future research must determine which sorts of local spurious generalizations influence infants' ability to detect a generalization that describes the entire set of input words.

We noted in the Introduction that the current work makes two points, one theoretical and one methodological. We now turn to the methodological point. A common practice in behavioral research is to present stimuli to participants in random order. This practice is true of experiments that test infants' ability to generalize from linguistic and other forms of input. The current pair of experiments examined the potential consequences of this practice by presenting infants with pre-test input that was either randomly ordered (Experiment 1) or ordered to avoid potential spurious generalizations between pairs of adjacent words (Experiment 2). The complete lack of generalization for the randomly ordered stimuli, coupled with the significant generalization for the more intentionally ordered stimuli, suggests that particular random stimulus orders might lead to an incorrect conclusion that infants at a particular age cannot make a particular linguistic generalization.

<sup>1</sup> The last four words for the Voicing Rule were *diva*, *daeva*, *daeba*, *taepa* and therefore had the same local spurious generalizations as the last four words of the POA Rule.

The current study, in combination with the preceding work by Gerken and Knight (2015), suggests that researchers might consider taking a more active role in ordering their stimuli in infant experiments. If the stimuli are generated on a number of dimensions (like those used here), it might be wise to draw each subsequent stimulus from a part of the stimulus space that is relatively far from the last stimulus drawn. This approach is related to the notion of ‘pedagogical sampling’ in which a teacher knows the relevant generalization (hypothesis) and produces data for the learner that demonstrate the boundaries of the hypothesis space (Shafto, Goodman & Griffiths, 2014). When a small set of input data is provided, the teacher/experimenter can easily select examples that are representative of the entire hypothesis space (e.g. the Non-Conflicting Condition in Table 1). However, with a large set of input data, the current study suggests that local sub-samples of the input set should also be representative of the hypothesis space. That is, the infant learner cannot be counted on to perform a batch analysis of all or most of the input. Rather, generalizations about the input appear to be influenced by local distributional properties.

## Acknowledgements

This research was supported by NSF 0950601 to LAG and NIH F32 HD065382 and K99 DC013795 to CMQ. We thank two anonymous reviewers for very helpful comments on previous drafts of the ms.

## References

- Aslin, R.N. (2007). What’s in a look? *Developmental Science*, **10** (1), 48–53.
- Chambers, K.E., Onishi, K.H., & Fisher, C.L. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, **87**, B69–B77.
- Chambers, K.E., Onishi, K.H., & Fisher, C.L. (2011). Representations for phonotactic learning in infancy. *Language Learning and Development*, **7** (4), 287–308.
- Gerken, L.A., Dawson, C., Chatila, R., & Tenenbaum, J. (2015). Surprise! Infants consider possible bases of generalization for a single input example. *Developmental Science*, **18**, 80–89.
- Gerken, L.A., & Knight, S. (2015). Infants generalize from just (the right) four words. *Cognition*, **143**, 187–192.
- 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 learning. *Infancy*, **7** (2), 183–206.
- 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. *et al.* (1995). The headturn preference procedure for testing auditory perception. *Infant Behavior and Development*, **18**, 111–116.
- Lany, J.A., & Gómez, R.L. (2008). Twelve-month-olds benefit from prior experience in statistical learning. *Psychological Science*, **19**, 1247–1252.
- Onnis, L., Waterfall, H.R., & Edelman, S. (2008). Learn locally, act globally: learning language from variation set cues. *Cognition*, **109** (3), 423–430.
- Roy, D.K., & Pentland, A.P. (2002). Learning words from sights and sounds: a computational model. *Cognitive Science*, **26** (1), 113–146.
- Saffran, J.R., & Thiessen, E.D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, **39**, 484–494.
- Seidl, A., Onishi, K.H., & Cristia, A. (2014). Talker variation aids young infants’ phonotactic learning. *Language Learning and Development*, **10** (4), 297–307.
- Shafto, P., Goodman, N.D., & Griffiths, T.L. (2014). A rational account of pedagogical reasoning: teaching by, and learning from, examples. *Cognitive Psychology*, **71**, 55–89.
- Thiessen, E.D., & Pavlik, P.I. Jr (2013). iMinerva: a mathematical model of distributional statistical learning. *Cognitive Science*, **37** (2), 310–343.
- Wang, Y., & Seidl, A. (2014). The learnability of phonotactic patterns in onset and Coda positions. *Language Learning and Development*, **11** (1), 1–17.
- Waterfall, H.R. (2006). A little change is a good thing: feature theory, language acquisition and variation sets. PhD thesis, University of Chicago. ProQuest Information & Learning, US.
- Yu, C., & Smith, L.B. (2012). Modeling cross-situational word-referent learning: prior questions. *Psychological Review*, **119** (1), 21–39.

Received: 4 August 2015

Accepted: 24 December 2015