LANGUAGE LEARNING
A Journal of Research in Language Studies
Language Learning ISSN 0023-8333

Toward Establishing Continuity in Linguistic Skills Within Early Infancy

Amanda Seidl,[a] Brian French,[b] Yuanyuan Wang,[a] and Alejandrina Cristia[c]

[a]Purdue University, [b]Washington State University, and [c]Laboratoire de Sciences Cognitives et Psycholinguistique, CNRS, IEC-ENS, EHESS

A growing research line documents significant bivariate correlations between individual measures of speech perception gathered in infancy and concurrent or later vocabulary size. One interpretation of this correlation is that it reflects language specificity: Both speech perception tasks and the development of the vocabulary recruit the *same* linguistic modules. However, correlations between infant cognitive measures (such as visual recognition memory) and vocabulary are also significant and display comparable strength. Can all of these correlations be reduced to extremely general rather than *specific* factors affecting performance in all laboratory tests? We take a first step in addressing this possibility by estimating the covariance matrix among two speech tasks (preference for the predominant stress pattern and native vowel discrimination) and two cognitive tasks (visual recognition memory and A-not-B), all of them gathered in the same group of infants tested between 5 and 8 months of age. Only the correlation between the two speech tasks was significant, lending little support to the generalist explanation. These data illustrate how a multivariate approach may inform our understanding of how infants build language in the first year of life and beyond. Future multivariate work following up on the same infants longitudinally will be better able to tease apart cognitive and linguistic contributions to vocabulary development.

Keywords infant speech perception; early predictors; individual differences

Introduction

A growing research line suggests that early infant speech perception is correlated with later language (Junge, Cutler, & Hagoort, 2010; Kuhl et al., 2008; Molfese & Molfese, 1985; Newman, Bernstein Ratner, Jusczyk, Jusczyk, &

Correspondence concerning this article should be addressed to Amanda Seidl, 500 Oval Dr., West Lafayette, IN 47906. E-mail: aseidl@purdue.edu

DOI: 10.1111/lang.12059

Dow, 2006; Rivera-Gaxiola, Klarman, Garcia-Sierra, & Kuhl, 2005; Singh, Reznick, & Xuehua, 2012; Tsao, Liu, & Kuhl, 2004; Weber, Hahne, Friedrich, & Friederici, 2005). However, it is not clear what these correlations mean. One extreme interpretation is that speech perception tasks in infancy tap a linguistic module, which interacts minimally with other areas of nonlinguistic cognition and which is the most crucial building block for later linguistic competence. We will call this the *specialist* view. Another extreme interpretation is that infant speech perception tasks tap very general performance skills, and that nearly any measure gathered in infants that reliably indexes individual variation will be a good predictor of language. Put another way, any two measures gathered in infancy may be reduced to the same very general abilities, and thus individual performance in any two tasks will be correlated. We will call this the *generalist* view.[1]

To a certain extent, current data support the latter alternative. For example, performance in tasks thought to tap cognitive skills in infancy correlates with later language (Colombo, Mitchell, Coldren, & Freeseman, 1991; Colombo, McCardle, & Freund, 2008; Kavšek, 2004; Rose, Feldman, & Jankowski, 2009; Tamis-LeMonda & Bornstein, 1989; Thompson, Fagan, & Fulker, 1991) and this correlation strength is not significantly different from that of language tasks (Cristia, Seidl, Junge, Soderstrom, & Hagoort, 2013). In fact, performance in one linguistic and one nonlinguistic task measured concurrently has sometimes been reported to correlate, as follows.

In a landmark study with 40 8- to 10-month-old infants, Lalonde and Werker (1995) showed that infants' performance on two cognitive nonlinguistic tasks (A-not-B and a visual categorization task based on feature cooccurrence) was related to their performance on non native, but not native, speech sound discrimination. Specifically, infants who failed the A-not-B task were more likely to discriminate the non native sounds. Infants who fail in A-not-B are those who continue to reach for an object that had previously been hidden in location A several times, instead of reaching for the object's current hiding location, B. Therefore, one interpretation could have been that infants who *persisted* in or failed to *inhibit* reaching behavior also persisted in or failed to inhibit speech sound discrimination. However, performance on non native sound discrimination was also associated for the other task, which did not clearly rely on inhibition. In this other task, infants saw several examples of two animal categories defined on the basis of correlated attributes (e.g., having round feet and a tail versus having webbed feet and feathers) and were later presented with a new exemplar of the same category, a new animal with uncorrelated attributes (e.g., round feet and feathers) and a completely new animal. Infants who looked as long to the animals with uncorrelated attributes and the novel animals were scored as succeeding at this task. It was also these infants who *failed* at the non native sound discrimination task.

Conboy, Sommerville, and Kuhl (2008) also present a complex relationship between measures of emerging language and measures of nonlinguistic skills. In that study, 11-month-olds ($n = 17$) were tested on their discrimination of a native sound contrast and a non native sound contrast, as well as on two very similar nonlinguistic tasks, both of which relied crucially on infants' reaching behavior (pulling a cloth to get a toy or detouring a reach to get a toy in a clear box, which involved inhibition of directly reaching toward the visible toy). Analyses revealed that infants who succeeded in inhibiting reaching behavior tended to discriminate non native sounds less well than native ones, whereas infants who failed at inhibition were equally sensitive to native and non native sounds (thus showing less native language competence).

Taken together, the simplest explanation for results of both of these studies may fit the generalist explanation. Indeed, one may hypothesize that infants were at ceiling in the native sound discrimination tasks, whereas the non native sound discrimination task was more sensitive to individual variation. In this context, then, a correlation in one case (between performance on non native sound discrimination and reaching behavior) and not the other (between performance on native sound discrimination and reaching behavior) would simply indicate that individual infants' general abilities were accurately measured by the two correlated measures, and not by the measure that does not correlate with the others. In other words, correlations would have been found in any and all sensitive tasks.

Given that the matter is far from settled, the present study sought to adjudicate between the specialist and generalist interpretations by inspecting individual variation across two linguistic and two cognitive tasks, all four of which were gathered in infancy. We reasoned that the specialist and generalist views would be best teased apart as follows: Following the generalist interpretation, the infant speech perception measures should be as closely related to the cognitive measures as they are to each other, because any and all sensitive tasks should pick up on the same individual variance. In contrast, the specialist interpretation predicts the infant speech perception measures to be more highly correlated with each other than with the cognitive measures gathered at the same time. It should be noted that our key specialist prediction applied only to the linguistic tasks, because the two nonlinguistic tasks were in part selected because they were fundamentally different from each other, as we explain next.

As in previous work (Conboy et al., 2008; Lalonde & Werker, 1995), we included an age-appropriate version of A-not-B. While this past research focused on 11-month-olds, A-not-B errors are most likely at about 8 months, an age at which this task draws on inhibition as well as several other cognitive processes (Clearfield, Diedrich, Smith, & Thelen, 2006). Interestingly, A-not-B performance in infancy has not, to our knowledge, been linked to later language development. The second cognitive task was Visual Recognition Memory (VRM), which had three desirable features. First, success in the task does not depend on a single cognitive skill, but on several basic building blocks of cognition (Rose, Feldman, & Jankowski, 2003). Second, its predictive value for both cognition and language has been repeatedly demonstrated (Bornstein & Sigman, 1986; Thompson et al., 1991; Rose et al., 2009). Finally, it has been studied extensively and the measure of individual differences it yields is reliable and relatively stable (Rose & Feldman, 1990). For all these reasons, VRM provides an excellent control for individual variation both of specific cognitive skills and in general laboratory-based task performance.

Both of the speech perception tasks were chosen to tap infants' acquired knowledge, albeit at somewhat different linguistic levels (segmental and surpasegmental). In line with previous work, one of the speech tasks focused on speech sound categories. Infants have been reported to home in on their native language's categories at about 6 to 12 months of age (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984). Previous research shows that both increased sensitivity to contrasts present in the native language and decreased sensitivity to non native languages account for some variance in language skill at 2 to 3 years of age (Kuhl et al., 2008). Because non native discrimination abilities were most closely tied to cognitive abilities in the work cited above, we put the generalist hypothesis to the strongest test by focusing on a native contrast. At the same time, as pointed out above, it is possible that infants were at ceiling in the native contrast perception in previous work. Therefore, we selected a pair of vowel sounds that are particularly difficult, to ensure that it would induce some individual variability. The second task tested infants' preference for stress patterns that are common, as opposed to rare, in the infants' native language. Tuning to the native stress patterns is evident already at 4 to 9 months (Friederici, Friedrich, & Christophe, 2007; Skoruppa et al., 2009). Furthermore, differences in processing of native stress have been found to differ in populations at risk of language delays (Herold, Hoehle, Walch, Weber, & Obladen, 2008) and stress preferences have been found to predict later language growth (Weber et al., 2005).

Infants were tested on one cognitive and one linguistic task at each of two ages: 5–6 months and 6–8 months. These two ages are close in time, yet they encompass crucially different periods of language and cognitive development. For language, the former period is widely associated with rapid development in the perception of prosody, whereas the clearest evidence for native sound attunement comes from the second half of the first year (Tsuji & Cristia, 2014). Thus, infants are likely on the cusp of stress pattern acquisition at 5–6 months (Friederici et al., 2007; Skoruppa et al., 2009) and on the cusp of language-specific phoneme tuning at 6.5–8 months (a process that certainly continues well into the second year; Werker, Fennell, Corcoran, & Stager, 2002). As for cognition, while VRM is relatively stable throughout early infancy, it appears that A-not-B is most challenging at around 8 months of age. Indeed, while 5–6-month-olds do not show a general tendency for A-not-B errors, 7–8-month-olds do display such patterns before outgrowing them in certain A-not-B versions (Clearfield et al., 2006). Note that our goal was that, in order to best show individual differences, our tasks should be challenging at each age. Given this goal we did not require behavior to globally be above chance on any given task. Rather, infants as a group should overall hover around chance levels, with only a subset performing above chance.

To clarify our key predictions, following the specialist view, we predict that the strongest correlations exist between stress and vowel measures. In contrast, following the generalist view, we predict correlations to be strong regardless of the linguistic versus cognitive divide, for example, between the two habituation-dishabituation tasks vowel perception and VRM.

Method

The study consisted of two to three visits to the lab for each participant. For the first visit, infants were tested first on the stress/trochaic preference task and then on the VRM task, with a short break in between. For the second visit, infants were tested first on vowel discrimination and then on the A-not-B task. If the infant was unable to do both tasks on this second visit, the A-not-B task was occasionally done at a third visit, which always took place fewer than 10 days after the second visit.

Participants

We included data from 95 healthy full-term monolingual English-learning infants (44 female). Infants were on average 5.62 months old at the first visit (min = 4.87 months, max = 6.55 months) and 6.92 months old at the second visit (min = 6.35, max = 8.39). Twelve infants required a third visit, at which they were 7.25 months old (min = 6.61, max = 8.06). An additional 41 infants were excluded from the study because they fussed or cried, failed to show up (because they moved away or were ill for an extended period) or showed a lack of parental compliance, or had experimental errors in more than two experimental tasks.

Children sometimes failed to complete one or two tasks. Data were missing for 1 child for trochaic preference, 2 for VRM, 7 for A-not-B, and 6 for vowel discrimination. In all, 84 infants completed all four experimental tasks. To maximize the power in each analysis, all infants with data in the relevant tasks were included.

Procedures and Measures

Trochaic Preference

The stimuli and procedure were based on Herold et al. (2008). Specifically, a female speaker spoke five trochaic and five iambic lists of segmentally identical words (gába vs. gabá) repeated 15–16 times (for more details, see Herold et al.). The order of the stimuli lists was randomly chosen by the computer with the stipulation that no two identical stress pattern lists occurred adjacently.

Infants were tested using the Headturn Preference Procedure (Jusczyk & Aslin, 1995). Each infant was seated on the caregiver's lap on a chair in the middle of a small three-sided booth within a sound-attenuated testing room. The experimenter was situated behind the testing booth and observed the infant through a monitor. During the experiment, the orientation of the infant's gaze was recorded on the computer by means of a button box. All choices regarding the side light and specific auditory stimulus list were made randomly via computer program. Both the experimenter and the caregiver wore tight-fitting headphones that played a mixture of continuous music and white noise to mask the auditory stimuli the infant heard. The overhead light was dimmed to make the panel lights more salient. Each trial began with the central green light flashing to attract the infant's attention to the center. When the infant looked to the center light, this light would extinguish, and one of the two side red lights (chosen randomly by the computer program used to run the experiment) would begin to flash. When the infant oriented to the side light, one of the auditory test lists (iambic or trochaic) would play repeatedly. This continued until the infant looked more than 30 degrees away from the light for 2 consecutive seconds. At this point, the side light would extinguish, the sound would stop, and the front green light would begin to flash in preparation for the next trial. The computer recorded the amount of time the infant was looking toward the light while each stimulus list was playing. If the infant looked away for less than 2 seconds and then looked back again, the trial continued, but the amount of time spent looking away was not counted in the overall tally.

The dependent measure for this task was a score of "Trochaic Preference" calculated as the orientation time to the trochaic pattern divided by the total looking time to both stress patterns for each child. When significant preferences are found, infants in this general paradigm show a familiarity preference, leading us to predict that this ratio will be higher than .5.

VRM

The stimuli and procedure were slightly modified versions of Thompson et al. (1991), the only modifications being that the stimuli were digital (rather than digitized from paper stimuli) and that the length of individual familiarization trials was infant controlled.

The infant was seated on a caregiver's lap in front of a large screen, onto which images were projected. Infants were shown up to nine problems (average = 7; range = 2–9), each consisting of a familiarization and a test phase. Both familiarization and test trials began with an attention-getter (a green square with a black spot that appeared and disappeared at regular intervals and nonspeech sound tracks which varied across problems). When the infant fixated on the screen, visual stimuli were projected to the left and right of the infant's visual field until she looked away for more than 2 seconds. During familiarization, the images to the right and left were identical (the same black-and-white photograph of a face in five of the problems or the same colorful geometric stimuli in the other four problems). Familiarization ended when the infant accumulated a fixed exposure time by looking at either of the sides (20 seconds for the faces, 10 seconds for the geometric stimuli). During testing, the familiarization image continued to be projected on one of the sides, while a new stimulus appeared on the other side (a new face in the face problems and a new geometric shape in the geometric shape problems). There were two test trials in each problem, with side of the new image counterbalanced across them, and their duration was fixed to 10 seconds.

After the experiment was completed, looking times during the test phase were coded offline from a version digitized at 30 frames per second. A novelty score was calculated for each testing phase. This novelty score was calculated by dividing the looking time to the novel image by the total looking time to either image. The average was then computed over all the problems the infant had completed. Infants tested in this paradigm are typically expected to show a novelty score above .5.

Native vowel discrimination

A female speaker recorded many exemplars of /ʃɪp/ and /ʃip/ in an infant-directed register. Four tokens of /ʃɪp/ that varied in pitch and intensity were selected for the habituation phase. Two additional tokens of the same type were selected for the test phase, as well as a token of /ʃip/, which was a perfect match in intensity and pitch characteristics (minimum, maximum, and average pitch) to one of the test /ʃɪp/ tokens.

The procedure was a replication of Houston, Horn, Qi, Ting, and Gao's (2007) hybrid method, which has led to high test–retest reliability (see Houston et al. for more details). The infant was seated on a caregiver's lap in front of a large screen. The image of a bull's eye was projected onto the screen during the playing of the stimuli. Infants in this study were habituated to four different tokens of the syllable /ʃɪp/, separated by 600 milliseconds of silence, until their average looking times in the preceding three trials dropped below 50% of the longest looking trials. At this point, same and alternating trials were introduced. During same trials, two new /ʃɪp/ tokens were presented, whereas in alternating trials one of those new /ʃɪp/ tokens alternated with the /ʃip/ token.

The dependent measure for this task was also a novelty preference (looking time to the alternating trials divided by the sum of looking to the alternating and the same trials). Infants in this habituation-type task, as in VRM, are expected to show a novelty preference, looking longer to the alternating pattern.

A-not-B

Like all others, this measure was conducted in a quiet experimental room; here, we used the infant version of the A-not-B task described in Clearfield et al. (2006). In this measure the infant was seated on the caregiver's lap across a small table from the experimenter. Within the reach of the experimenter and the infant was a low yellow box (32 centimeters [cm] x 27 cm x 7 cm), in the center of which sat two identical black lightweight metal objects with two bunny-ear-like bumps on each. These were placed approximately 8 cm apart: object A (set on infant's left) and object B (set on infant's right; see Figure 1).

At the beginning of the experiment, the experimenter made eye contact with the infant and greeted him/her by calling his/her name in infant-directed speech. At the same time, the experimenter touched the box and objects so as to attract the infant's attention until the infant was aware of, and was comfortable with, the box and the objects. In Trial 1, the experimenter waved object A for several seconds until the infant's attention was directed to this object. Object A then was placed at the edge of the box (Figure 2, Trial 1), which was closer to the infant than object B. The experimenter then moved the box forward so the

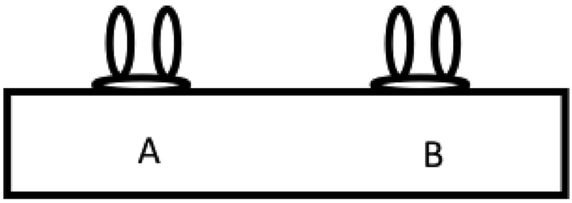

Figure 1 Objects in A-not-B task.

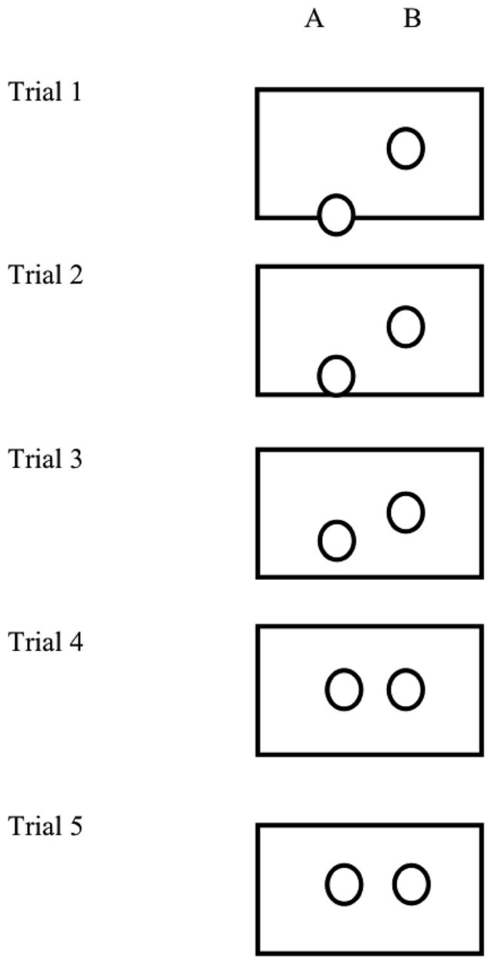

Figure 2 Order of trials for A-not-B task.

Table 1 Means and standard deviations for all continuous measures gathered

Measure	Mean	*SD*
Trochaic preference	0.51	0.14
VRM	0.59	0.07
Vowels	0.51	0.10

infant was able to reach both object A and object B. The experimenter observed the infant's behavior and made a note as to which object the infant grabbed (object A or object B). After the infant grabbed the object, the experimenter placed the object back onto the box in the place where this object was originally placed. This procedure was repeated for the next three trials with the location of object A progressively moving toward the back of the box (Figure 2, Trials 2, 3, and 4) until it was in line with object B (Trial 4). In Trial 5 (see Figure 2, Trial 5), the experimenter waved object B and placed it back in the original position after the infant's attention was captured. The box was then again moved forward in the reaching space of the infant. We recorded on this test trial whether the infant lifted object B (hit) or Object A (error). If in the course of the proceeding trials infants consistently randomly grabbed the B object when the A object was indicated this was coded as random (indicating a lack of infant attention or ability). Thus, there were three possible scores for this measure: hit (indicating a success in inhibiting a reach to the A object), error (indicating a failure to inhibit a reach to the A object), or random (random behavior).

Results

Table 1 shows means and standard deviations for all continuous measures gathered. Histograms of Vowels and Trochaic preference are included in Figure 3, and a histogram of VRM is shown in Figure 4. Aside from the 7 infants without an A-not-B score, there were 22 classified as "hit," 33 as "error," and 33 as "random." (The complete data set is available for download from sites.google.com/site/invarinf/documents.)

To address our research questions, we examined the correlations between the three tasks that had continuous outcomes. As seen in Table 2, the only correlation that was significant was between the two speech perception tasks assessed at different points in time, indicating that approximately 6 percent of the variance at time 2 can be explained by the infant's score at time 1. Table 3

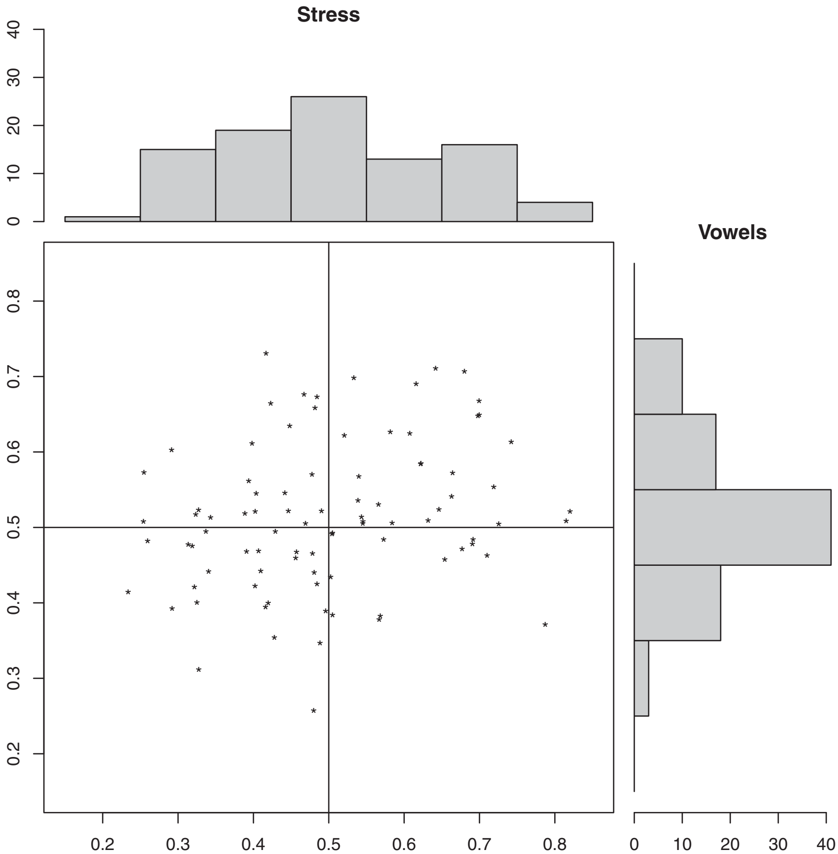

Figure 3 Plot of vowel scores as a function of trochaic preference, and histograms of these two variables.

shows that performance in the two linguistic tasks cooccurs significantly above chance, $\chi^2(1) = 4.53$; $p = .03$. Notice, furthermore, that there is no indication of a nonlinear or U-shaped relationship between scores in the stress and vowel tasks in Figure 3.

The two correlations with VRM were not significant; additionally, they were very small in magnitude, indicating a lack of significance was not due simply to low statistical power.

As noted above, the A-not-B task outcome was a classification (hit, error, or random). Therefore, we investigated whether there were differences on the other three tasks as a function of performance in the A-not-B task through

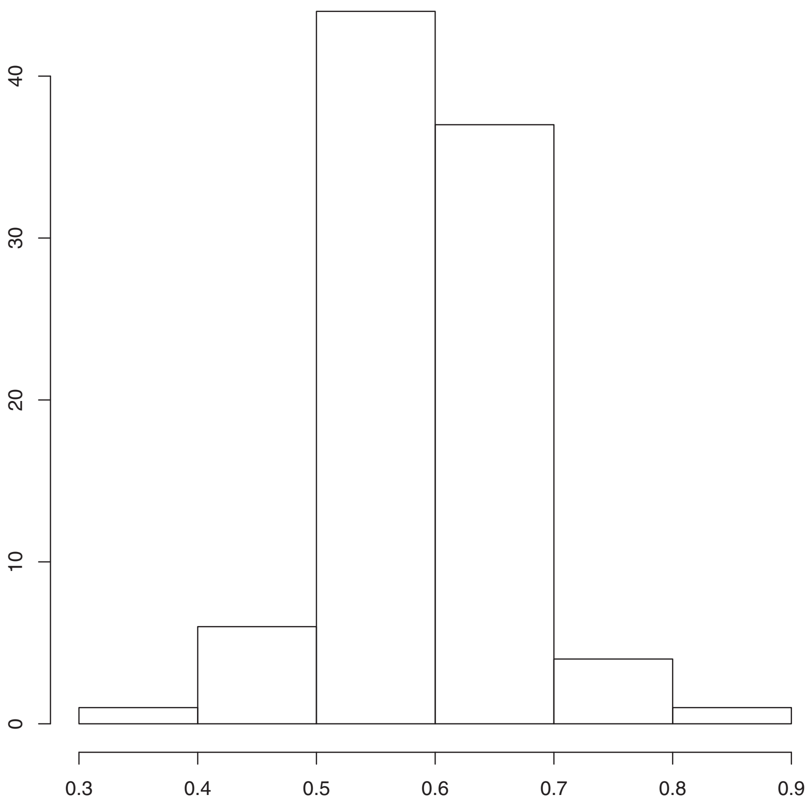

Figure 4 Histogram of Visual Recognition Memory scores.

Table 2 Pearson correlation coefficients (and degrees of freedom) between the three tasks with continuous outcomes

Variables	Vowels	VRM
Trochaic pref	0.24 (87)*	0.02 (90)
Vowels		−0.02 (85)

*$p < .05$

three one-way analyses of variance. The independent variable was formed by creating groups based on the score (hit, error, or random). Means for each of the three outcome measures, as a function of A-not-B performance, are provided in Table 4. For all three dependent variables, there are no significant differences:

Table 3 Cooccurrence of predicted good performance (novelty preference in vowels, trochaic preference in stress) across the linguistic tasks

		Stress	
		Not	Good
Vowels	Not	26	22
	Good	13	28

Table 4 Means (*SD*; *N*) for each of the three outcome measures, as a function of A-not-B performance

A-not-B group	Stress	Vowels	VRM
Hit	.46 (.12; 22)	.49 (.08; 21)	.61 (.05; 22)
Error	.52 (.16; 33)	.51 (.10; 33)	.59 (.06; 32)
Random	.52 (.14; 33)	.53 (.10; 32)	.58 (.08; 32)

Trochaic preference, $F(2,85) = 1.17$, $p = 0.32$; Vowels, $F(2,83) = 1.07$, $p = 0.35$; VRM, $F(2,83) = 1.08$, $p = 0.34$.

Discussion

While it is clearly the case that measures of infant cognition predict concurrent and later language (Conboy et al., 2008; Colombo et al., 1991; Lalonde & Werker, 1995; Rose et al., 2009; Kavšek, 2004; Thompson et al., 1991; Tamis-LeMonda & Bornstein, 1989), the present study suggests that not every and any performance measure will correlate with linguistic skills, contrary to the generalist hypothesis. Specifically, the early linguistic task (trochaic preference) accounted for a greater percentage of the variance in the later linguistic measure (native vowel discrimination) than either a cognitive measure administered at the same time (A-not-B) or a highly sensitive cognitive measure administered earlier on (VRM). This not only suggests that there is a certain degree of specialization, but also provides some support for the validity of a linguistic construct that even holds over short time spans. That is, the linguistic tasks seem to be tapping a different construct than the cognitive tasks, rather than simply reflecting individual variation in performance on laboratory tasks.

In addition, the present data replicate previous failures to correlate A-not-B performance and *native* sound discrimination (Conboy et al., 2008; Lalonde

& Werker, 1995). There are two possible interpretations for this result. First, it seems likely that, as suggested in previous work (Conboy et al., 2008), native speech perception tasks do not recruit inhibition, but non native speech perception tasks do, thus leading to correlations with A-not-B. Unfortunately, this (specialist) interpretation seems unlikely because in previous work infants' performance in non native discrimination correlated also with an unrelated cognitive task, whereas native discrimination correlated with neither task (Lalonde & Werker, 1995). The present study breaks this deadlock through the inclusion of VRM and trochaic preference. Despite its proven sensitivity to individual variation, VRM failed to correlate with native sound discrimination. Moreover, both of the speech tasks used here depended on infants' reliance on positive evidence in the ambient speech, rather than on the *absence* of evidence. As a result, we can now extend the lack of correlation to other linguistic developments that do not rely on inhibition. Finally, one previous longitudinal study had included VRM, in addition to a prosodic preference measure, as predictor of childhood vocabulary (Cristia & Seidl, 2011). In that work, the predictive value of the prosodic measure was not mediated by VRM performance. Putting all results together, non native sound discrimination appears to load primarily on inhibition and may only secondarily be an index of linguistic development. It is not entirely clear how non native sounds are processed by the native linguistic system and thus whether they even tap the same construct as native sounds. Indeed, recent neuroimaging work suggests that, by the end of the first year, processing sound contrasts that are non native recruits different neural networks than first language contrasts (a recent review in Minagawa-Kawai, Cristia, & Dupoux, 2011).

Nonetheless, there remains one conceptual explanation for the lack of relationship between native speech perception tasks and cognitive tasks in the present and previous work. One could argue that native sound discrimination is simply easier and less resource intensive than non native sound discrimination (after all, native sounds are more frequently experienced), and this fact alone may account for the more powerful non native correlations with cognitive tasks. In other words, native vowel discrimination in particular, and native speech perception tasks in general, may be less sensitive to individual variation.

At this point, it is relevant to inspect Table 1 and Figures 3 and 4, where it is clear that the two linguistic tasks were much more challenging than VRM. While in the two language tasks averages are very close to .5, very few children scored .5 or less in VRM. In view of this, one may wonder whether a floor effect present in the language tasks (and absent in VRM) may have affected our results. It should be noted that, if anything, the idea of a floor effect for the linguistic tasks plays against finding a correlation: If all infants tend to fail in a task, individual variation should be minimal and it should be more difficult to measure it reliably. Put another way, if the two linguistic tasks elicited too little variance or were too noisy, their intercorrelation should have been affected as much as, or more than, their respective associations with the two cognitive tasks.

Quite to the contrary, it was precisely between the two native speech perception tasks that a significant correlation was found, and not with either of the cognitive tasks. The correlational results between the two speech tasks are strengthened through an association that is notable in Figure 3 and Table 3. The higher counts are for the two cells in the diagonal (with the same performance in both tasks), and it was relatively rare to find infants who failed at the stress task (i.e., look longer to the iambs than the trochees) but succeed in the vowels task (i.e., dishabituated to a change in vowel). This pattern reinforces the idea that both tasks are tapping, to a certain extent, a common construct.

Despite this clear pattern, some readers may suggest that the correlation coefficient reported is small. In infant work, however, higher correlations have more frequently been observed with very small sample sizes (e.g., $n = 10$, $r = .653$ in Houston et al., 2007; $n = 16$, $r = .481$ in Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005). A recent meta-analysis on correlations between performance in a range of infant speech perception and nonlinguistic tasks, on the one hand, and concurrent or longitudinal measures of vocabulary, on the other hand, shows that such high values do not hold for the population of effect sizes reported (Cristia et al., 2013): The weighted median correlations were about $r = .3$. Similar median correlations have been reported even for test–retest of the exact same task at two ages; for example, for visual recognition memory, the stability in performance over a 3-month interval (similar to the one used here) was $r = .34$ with $n = 38$ (Colombo, Mitchell, & Horowitz, 1988). While the r found here is somewhat lower, it pertained to performance gathered in two different tasks (tested in two different rooms, etc.), rather than in the exact same task tested twice.

Nonetheless, we should make it clear that an interpretation of whether the two linguistic tasks tap the exact same construct or simply an overlapping set of abilities goes beyond the specific question we set out to answer. Our null hypothesis was that all lab-based tasks (that were sensitive to individual variation) were tapping a very general construct of lab performance. The data clearly do not support this hypothesis. We observe *some* relationship between the two linguistic tasks, in the face of *no* relationship between them and our two cognitive tasks. Thus, our departure from the previous tradition, focused on inhibition and sound discrimination, has allowed us to demonstrate that individual variation in infant speech perception measured in the laboratory does *not* reflect solely broad performance factors affecting each and every lab measure.

Although this is the most conservative interpretation that can be made, we would like to take one step further and discuss some potential implications regarding the *divergent* constructs that our linguistic and our cognitive tasks recruited. Specifically, previous work has demonstrated that performance in A-not-B at the ages tested reflects variation in inhibition skills, whereas performance in VRM is associated with memory and selective attention. Insofar as the absence of a correlation may be interpreted, one could say that (unlike non native sound discrimination) neither of the native sound/stress tasks relied on these skills to any great extent. A particularly intriguing possibility is that different memory systems are recruited by VRM and the speech tasks, which rely more crucially on long-term acquired representations. Whereas a novelty preference in VRM can emerge through memory of specific episodes, the speech tasks used here may recruit long-term memory of more abstract categories to a greater extent.

While inhibition has not been described as playing a key role in the development of native vowel categories or prosodic templates, it has been invoked as the way in which the more common prosodic template (trochaic) comes to be preferred over the less common one (Herold et al., 2008). That is, it has been proposed that attention to non native categories is detrimental because it indicates that infants are not effectively allocating their attentional resources (Kuhl et al., 2008), and the same has been said of attending to rare stress patterns as much as (or even more than) common stress patterns. Perhaps, on the basis of this repeated lack of correlation between measures of inhibition and native language skills, experts should revise their description of this process so that it does not seem to rely on allocation of attention. Instead, tuning to sounds and prosodic patterns that are present could be described as the development of abstract categorizations for the types of sounds frequently encountered. We believe it will be difficult to tease these two visions apart behaviorally, but they do make very different neuroimaging predictions. A network where inhibition and selective attention is involved relies crucially on frontal and striatal structures (Booth et al., 2003); in contrast, the second description predicts increased engagement of middle to posterior temporal cortices, without any long-distance connections being necessary (Scott & Johnsrude, 2003).

In sum, the present study contributes to a growing body of work investigating meaningful individual variation through multiple infant measures and/or longitudinal approaches. Together with previous work, present results suggest that performance in speech perception tasks captures some aspects of infant behavior and language development that are not only stable over a few months, but further are somewhat specialized. The current work is the first to demonstrate that individual variation in infant performance in speech perception tasks can be both reliable and separable from that in nonlinguistic tasks. We hope these conclusions pave the way for a renewed exploration of the linguistic and nonlinguistic contributions to infants' discovery of language.

Note

1 This description coincides with the dichotomy between domain specificity and domain generality, but not with that between nativist and emergentist. For example, infants could be born with learning mechanisms that are used in both speech and visual processing—a nativist, generalist view that is espoused in Perfors, Tenenbaum, and Regier (2011), to give just one example. However, the present work is not concerned with and cannot disambiguate between nativist and emergentist views of acquisition.

References

Booth, J. R., Burman, D. D., Meyer, J. R., Lei, Z., Trommer, B. L., Davenport, N. D., et al. (2003). Neural development of selective attention and response inhibition. *Neuroimage*, *20*, 737–751.

Bornstein, M. H., & Sigman, M. D. (1986). Continuity in mental development. *Child Development*, *57*, 251–274.

Clearfield, M. W., Diedrich, F. J., Smith, L. B., & Thelen, E. (2006). Young infants reach correctly in A-not-B tasks: On the development of stability and perseveration. *Infant Behavior and Development*, *29*, 435–444.

Colombo, J., McCardle, P., & Freund, L. (2008). *Infant pathways to language*. Hove, UK: Psychology Press.

Colombo, J., Mitchell, D. W., Coldren, J., & Freeseman, L. (1991). Individual differences in infant visual attention: Are short lookers faster processors or feature processors? *Child Development*, *62*, 1247–1257.

Colombo, J., Mitchell, D. W., & Horowitz, F. D. (1988). Infant visual attention in the paired-comparison paradigm: Test-retest and attention-performance relations. *Child Development*, 1198–1210.

Conboy, B. T., Sommerville, J. A., & Kuhl, P. K. (2008). Cognitive control factors in speech perception at 11 months. *Developmental Psychology*, *44*, 1505–1512.

Cristia, A., & Seidl, A. (2011). Attention to prosodic units in infancy predicts later language acquisition. In I. K. N. Denis (Ed.), *Proceedings of the 35th Annual*

Boston University Conference on Language Development (pp. 145–156). Somerville, MA: Boston University Press.

Cristia, A., Seidl, A., Junge, C., Soderstrom, M., & Hagoort, P. (2013). Predicting individual variation in language from infant speech perception measures. *Child Development*, doi: 10.1111/cdev.12193.

Friederici, A., Friedrich, M., & Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Current Biology*, *17*, 1–4.

Herold, B., Hoehle, B., Walch, E., Weber, T., & Obladen, M. (2008). Impaired stress pattern discrimination in very low birth weight infants during the first six months of life. *Developmental Medicine and Child Neurology*, *50*, 678–683.

Houston, D. M., Horn, D. L., Qi, R., Ting, J. Y., & Gao, S. (2007). Assessing speech discrimination in individual infants. *Infancy*, *12*, 119–145.

Junge, C., Cutler, A., & Hagoort, P. (2010). Ability to segment words from speech as a precursor of later language development: Insights from electrophysiological responses in the infant brain. *Proceedings of 20th International Congress on Acoustics*.

Jusczyk, P. W., & Aslin, R. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology*, *29*, 1–23.

Kavšek, M. (2004). Predicting later IQ from infant visual habituation and dishabituation: A meta-analysis. *Applied Developmental Psychology*, *25*, 381–392.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B*, *363*, 979–1000.

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the critical period. *Language Learning and Development*, *1*, 237–264.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606–608.

Lalonde, C. E., & Werker, J. F. (1995). Cognitive influences on cross-language speech perception in infancy. *Infant Behavior and Development*, *18*, 459–475.

Minagawa-Kawai, Y., Cristia, A., & Dupoux, E. (2011). Cerebral lateralization and early speech acquisition: A developmental scenario. *Developmental Cognitive Neuroscience*, *1*, 217–232.

Molfese, D. L., & Molfese, V. J. (1985). Electrophysiological indices of auditory discrimination in newborn infants: The bases for predicting later language development? *Infant Behavior and Development*, *8*, 197–211.

Newman, R. S., Bernstein Ratner, N., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology*, *42*, 337–367.

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, *118*, 306–338.

Rivera-Gaxiola, M., Klarman, L., Garcia-Sierra, A., & Kuhl, P. K. (2005). Neural patterns to speech and vocabulary growth in American infants. *Developmental Neuroscience*, *8*, 162–172.

Rose, S. A., & Feldman, J. F. (1990). Infant cognition: Individual differences and developmental continuities. In J. Colombo & J. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, prediction* (pp. 229–246). Hillsdale, NJ: Erlbaum.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2003). The building blocks of cognition. *Journal of Pediatrics*, *143*, S54–S61.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2009). A cognitive approach to the development of early language. *Child Development*, *80*, 134–150.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in neurosciences*, *26*, 100–107.

Singh, L., Reznick, J., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: A longitudinal analysis. *Developmental Science*, *15*, 482–495.

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., et al. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, *12*, 914–919.

Tamis-LeMonda, C. S., & Bornstein, M. H. (1989). Habituation and maternal encouragement of attention in infancy as predictors of toddler language, play, and representational competence. *Child Development*, *60*, 738–751.

Thompson, L. A., Fagan, J. F., & Fulker, D. W. (1991). Longitudinal prediction of specific cognitive abilities from infant novelty preference. *Child Development*, *62*, 530–538.

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*, *75*, 1067–1084.

Tsuji, S., & Cristia, A. (2014). Perceptual attunement in vowels: A meta-analysis. *Developmental psychobiology*, *56*, 179–191.

Weber, C., Hahne, A., Friedrich, M., & Friederici, A. (2005). Reduced stress pattern discrimination in 5-month-olds as a marker of risk for later language impairment: Neuropshysiological evidence. *Cognitive Brain Research*, *25*, 180–187.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*, 1–30.

Werker, J. F., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.