Infant Behavior and Development 62 (2021) 101524

Contents lists available at ScienceDirect

Infant Behavior and Development

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

Infant speech perception and cognitive skills as predictors of later vocabulary

Yuanyuan Wang [a,*], Amanda Seidl [b], Alejandrina Cristia [c]

[a] Department of Otolaryngology–Head & Neck Surgery, The Ohio State University, United States
[b] Department of Speech, Language, and Hearing Sciences, Purdue University, United States
[c] Laboratoire de Sciences Cognitives et Psycholinguistique, CNRS, IEC-ENS, EHESS, France

ARTICLE INFO

Keywords:
Multivariate
Speech perception
Cognitive skills
Predictors
Language development

ABSTRACT

Research has identified bivariate correlations between speech perception and cognitive measures gathered during infancy as well as correlations between these individual measures and later language outcomes. However, these correlations have not all been explored together in prospective longitudinal studies. The goal of the current research was to compare how early speech perception and cognitive skills predict later language outcomes using a within-participant design. To achieve this goal, we tested 97 5- to 7-month-olds on two speech perception tasks (stress pattern preference, native vowel discrimination) and two cognitive tasks (visual recognition memory, A-not-B) and later assessed their vocabulary outcomes at 18 and 24 months. Frequentist statistical analyses showed that only native vowel discrimination significantly predicted vocabulary. However, Bayesian analyses suggested that evidence was ambiguous between null and alternative hypotheses for all infant predictors. These results highlight the importance of recognizing and addressing challenges related to infant data collection, interpretation, and replication in the developmental field, a roadblock in our route to understanding the contribution of domain-specific and domain-general skills for language acquisition. Future methodological development and research along similar lines is encouraged to assess individual differences in infant speech perception and cognitive skills and their predictability for language development.

1. Introduction

Language acquisition is one of the major achievements during early child development. How much of language acquisition depends on processes that are specialized for language, as compared to processes that are shared with other areas of cognition? In the past several decades, a growing body of literature has demonstrated a bivariate relationship between speech perception and cognitive measures gathered during infancy and later language outcomes (e.g., Junge, Kooijman, Hagoort, & Cutler, 2012; Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005; Liu, Kuhl, & Tsao, 2003; Marchman & Fernald, 2008; Newman, Rowe, & Bernstein Ratner, 2016; Rose, Feldman, & Jankowski, 2009). However, these speech perception and cognitive measures have been primarily examined separately, thus the exact nature and predictive value of distinct and overlapping skills between these two domains remain largely unclear. In a meta-analysis examining the relationship between infant measures and later language outcomes (Cristia, Seidl, Junge, Soderstrom, &

* Corresponding author at: Department of Otolaryngology-Head and Neck Surgery, The Ohio State University, 915 Olentangy River Rd., Columbus, OH, 43212, United States.
E-mail address: Yuanyuan.Wang@osumc.edu (Y. Wang).

https://doi.org/10.1016/j.infbeh.2020.101524
Received 14 August 2020; Received in revised form 12 December 2020; Accepted 14 December 2020
Available online 26 December 2020
0163-6383/

Hagoort, 2014), both speech perception and general cognitive measures gathered in infancy predicted vocabulary measured concurrently or longitudinally. These findings suggest that language acquisition, or at least vocabulary development, may depend both on early speech perception and cognitive skills.

Despite these findings, there is a lack of research directly comparing how infant speech perception and cognitive skills predict language outcomes using a longitudinal within-subjects design. This constitutes a significant knowledge gap, as without a direct comparison using a within-subjects design, it is difficult to determine whether some of these tasks are better or worse predictors of child language. Moreover, certain infant speech perception and cognitive abilities share common underlying mechanisms and thus should both be correlated with language outcomes (Lalonde & Werker, 1995; Morgan & Saffran, 1995). For example, Lalonde and Werker (1995) tested 40 8- to 10-month-old infants on two speech perception tasks (a native sound discrimination and a nonnative sound discrimination task) and two cognitive tasks (A-not-B and a visual categorization task based on feature cooccurrence). They showed that performance in their nonnative sound discrimination task was correlated with the performance in their two cognitive tasks.

In this research, we sought to contribute to filling this gap by examining *both* speech perception *and* cognitive abilities in infancy and their relationship with later language development together. Because of the range of variability in language development in both typically-developing children and various clinical populations (Geers, Strube, Tobey, Pisoni, & Moog, 2011; Hart & Risley, 1995; Sigman & Norman, 1999), understanding the contribution of infant speech perception and cognitive skills in predicting language development becomes a theoretically and clinically important issue. Such information not only allows for a better understanding of the mechanisms underlying human language, but also provides important insights into early identification of, and intervention for, speech and language disorders.

1.1. *Infant speech perception and language development*

Long before children produce words, they reach important milestones in speech perception that serve to prepare them for later language acquisition (Johnson & Seidl, 2008; Jusczyk, Cutler, & Redanz, 1993; Werker & Tees, 1984; Werker, Pegg, & McLeod, 1994). Both theoretical and empirical evidence posits an important role for infant speech perception in spoken language development. For example, models of child language acquisition, such as the Word Recognition and Phonetic Structure Acquisition (WRAPSA) model (Jusczyk, 1993), the developmental framework for Processing Rich Information from Multidimensional Interactive Representations (PRIMIR; Werker & Curtin, 2005), and the Native Language Magnet theory (NLM; Kuhl et al., 2008) have all suggested that speech perception during the first year of life serves as an emergent source of support for child language development.

Moreover, there is a well-established empirical literature documenting a correlation between both behavioral and neurophysiological measures of infant speech perception (e.g., phonetic, phonological, and lexical perception) and language outcomes, especially vocabulary size (Donnelly & Kidd, 2020; Junge, Cutler, & Hagoort, 2010; Kuhl et al., 2008, 2005; Newman, Ratner, Jusczyk, Jusczyk, & Dow, 2006; Tsao, Liu, & Kuhl, 2004; Vouloumanos & Curtin, 2014; Werker & Curtin, 2005). This body of research has tapped multiple linguistic levels using a variety of methodologies.

Research examining the predictors at the segmental level has been focused on native and nonnative sound discrimination, which has been shown to develop rapidly during the first year of life in the seminal study conducted by Werker and Tees (1983). For example, using the Conditioned Headturn Procedure, Tsao et al. (2004) tested 6-month-old English-learning infants' ability to discriminate vowel contrasts. They found that the number of trials needed to reach criterion was significantly negatively correlated with expressive vocabulary size at 13 and 24 months. In this case, fewer trials indicated better vowel discrimination skills. Recently, Singh (2019) extended this work and examined the relationship between native and nonnative phonetic discrimination abilities at 10–11 months and vocabulary outcomes at 2 and 3 years in English-Mandarin bilingual infants using a standard habituation paradigm. These findings showed that, similar to monolingual infants, native speech perception of /ba/ and /da/ predicted vocabulary size in the bilingual infants, providing additional support for the continuity between infant sound discrimination and later vocabulary growth.

Other research examines infant predictors at the prosodic level. For example, using event-related potentials, Weber, Hahne, Friedrich, and Friederici (2005) examined 5-month-olds' responses to trochaic and iambic stress patterns using the pseudoword /baba/ with either iambic or trochaic stress. Infants were grouped based on their vocabulary production scores collected at 12 and 24 months retrospectively. Infants with low vocabulary production showed reduced brain responses to the trochaic stress pattern. Similarly, Ference and Curtin (2013) tested 5-month-old infants' stress preference using the Headturn Preference Procedure (HPP) and collected vocabulary scores at 12 months in a longitudinal design. They showed that the stress preference score, calculated by subtracting looking time to iambic trials from the looking time to trochaic trials, was positively correlated with receptive vocabulary.

There is also a line of research examining infant predictors at the lexical or proto-lexical level. For example, using the HPP procedure, Newman et al. (2016) demonstrated that 7-month-old infants' ability to segment words from continuous speech predicted vocabulary outcomes at 2 years in a retrospective design. These findings were replicated in a later study which included two-word segmentation tasks varying in complexity using a prospective design (Singh, Steven Reznick, & Xuehua, 2012). Moreover, Junge et al. (2012) assessed word recognition using event-related potentials in 10-month-old month infants and showed that the ability to rapidly recognize words previously presented within an utterance significantly predicted word production at 24 months.

Taken together, the findings from these lines of research, which used a variety of methodologies and tasks, suggest that infant speech perception abilities are correlated with later language development.

1.2. Infant cognition and language development

In the past few decades, a growing literature on infant cognition suggests that infant cognitive skills develop rapidly in the first year of life (Colombo, Shaddy, Richman, Maikranz, & Blaga, 2004; Fagan, 1984). This research also suggests that infant cognitive skills can be measured during infancy (Rose, 1989; Rose, Feldman, & Jankowski, 2004), and variability in performance on infant cognitive tasks has been linked to individual differences in general cognitive outcomes later in life (Fagan & McGrath, 1981; Fagan, 1982; Rose & Feldman, 1995).

The domain-general view of child language development suggests that language learning also relies on a set of cognitive processes that may not be specific to language (Colombo et al., 2009; Fernald, Perfors, & Marchman, 2006; Rose et al., 2009). This body of research examines the associations between cognitive processes, particularly attention, memory, and language skills in infants. In support of this general view, findings have revealed that laboratory measures of infant cognition are significantly correlated with language measures later in toddlerhood and childhood (Choudhury, Leppanen, Leevers, & Benasich, 2007; Colombo et al., 2009; Fernald et al., 2006; Rose et al., 2009; Salley, Panneton, & Colombo, 2013).

In the 1980s, Fagan and colleagues conducted a series of studies examining infant visual recognition memory and its relationship with later vocabulary (Fagan & McGrath, 1981; Fagan & Singer, 1983). For example, in one study, they tested 35 infants for novelty preferences on pairings of face pictures and abstract patterns at 7 months. Vocabulary scores were obtained from these infants when they turned approximately 4 and 7 years. They showed that infants' performance in the visual recognition memory task, measured by a novelty preference score, was correlated with vocabulary outcomes measured at both ages (Fagan & McGrath, 1981). This constitutes one of the first pieces of evidence showing an association between infant cognitive skills and later vocabulary development.

More recently, Rose et al. (2009) examined a battery of cognitive measures, including memory, representational competence, processing speed, and attention at 12 months from 56 preterm and 126 full-term infants; they also gathered vocabulary measures from these infants at 12 and 36 months. Results showed that visual recognition memory (delayed recall) and representational competency (cross-modal transfer and object performance) at 12 months of age were associated with both concurrent and later vocabulary size for both full-term and preterm children. Similarly, Salley et al. (2013) conducted a longitudinal study and examined the associations among several infant cognitive skills as well as language measures at multiple points in 53 infants. Specifically, they measured infant attention on a distractibility task at 11 months, joint attention at 14 months, word-object association at 14 months, and vocabulary size and multi-word production at 18 months. Their analyses showed that infant attention/distractability and joint-attention were significantly associated with vocabulary size; infant attention also significantly predicted multi-word production.

In sum, this work suggests that language emergence and growth may also be related to basic cognitive processes that affect all aspects of cognitive competence, thus supporting a domain-general contribution to the emergence of language.

1.3. Multivariate approach to language development

Taken together, these two bodies of research suggest that both infant speech perception and early cognitive abilities are associated with concurrent and later language outcomes. However, less is known about the *relative* importance of early speech perception and cognitive skills in predicting language outcomes. This question could be tested through a prospective longitudinal study examining both speech perception and cognitive abilities in infancy and their relationship with later language development in the same individuals. Unfortunately, there is very little empirical evidence on this topic.

To the best of our knowledge, the only study that has examined multiple speech perception and cognitive tasks and their relationship with language measures was conducted by Conboy and colleagues, and this study measured concurrent, rather than prospective vocabulary (Conboy, Sommerville, & Kuhl, 2008). Specifically, they examined 17 11-month-olds' performance on both a native sound discrimination and a nonnative sound discrimination task, two cognitive tasks requiring attention and inhibitory control, and *concurrent* vocabulary size measured via the MacArthur-Bates Communicative Development Inventory. Findings showed that infants who succeeded in inhibiting reaching behavior tended to discriminate nonnative sounds less well than native ones. Moreover, they divided the infants into two equal groups based on the median split of the vocabulary size and found that infants in the higher vocabulary group showed greater sensitivity to the native contrast. They also divided the infants into two groups based on their behavior in cognitive tasks, and found that these two groups did not differ in their vocabulary size. These findings provide preliminary evidence for a specific relation between native speech perception and vocabulary. However, due to the small sample size ($N = 17$), the nature of tasks involved (both cognitive tasks tapped a similar mechanism and relied on reaching behavior), and the statistical approach (they examined group differences due to small sample size), the question of the weighting of speech perception and cognitive skills for language outcomes remains open.

1.4. The challenges of infant data

Before we move on, it is worth mentioning the challenges related to infant studies. Since the 1950s, the field has noticed that the overwhelming majority of published studies in psychology seemed to focus on evidence against the null hypothesis (Sterling, 1959), and this bias remains strong decades later (Sterling, Rosenbaum, & Weinkam, 1995). One step to recovery involves pre-registering analyses and adopting other techniques to avoid picking analysis schemes and data selection strategies that only leave the author with the choice of rejecting the null hypothesis (Scheel, Schijen, & Lakens, 2020). A related approach involves using Bayesian statistics, in which a support for the null (and against the alternative) can be quantified (Lakens, McLatchie, Isager, Scheel, & Dienes, 2020).

This methodology is particularly important for infant studies because there is some indication that infant data are particularly noisy. Collecting infant data is challenging and expensive, which often leads to small sample sizes (Oakes, 2017). In this context, allowing more results, especially the null results, to become public is necessary so that they can be combined using a meta-analytic or mega-analytic approach. In fact, the publication of such studies may also provide the context necessary for others who have data in their file drawers coming forward and contributing it to the field (Tsuji, Cristia, Frank, & Bergmann, 2020).

1.5. Current study

Our main goal was a conceptual one, to deepen our understanding of the relative contributions of domain-specific and domain-general skills for infants' language acquisition. To achieve this goal, in the current study, we investigate relationships between speech perception skills during infancy, general cognitive abilities during infancy, and later language development as indexed by vocabulary size. We overcome previous methodological limitations evident in the existing literature by (1) reporting on a larger group of children ($n = 97$), (2) using four measures that tap different aspects of speech perception and cognitive skills, (3) using a prospective longitudinal design, (4) relying on a frequentist statistical approach selected prior to seeing the results, and (5) complementing these statistics with a Bayesian analysis.

Specifically, the study consists of two infant speech perception tasks: a stress pattern preference task (henceforth Stress) and a native vowel discrimination (henceforth Vowel); and two infant cognitive tasks: a visual recognition memory task (henceforth VRM) and an A-not-B task. All speech perception and cognitive measures were gathered when infants were between 5 and 7 months. We chose these infant measures for the following reasons: (1) All the measures are commonly used in previous research assessing infant speech perception and cognitive skills (Conboy, Sommerville et al., 2008; Fagan & McGrath, 1981; Ference & Curtin, 2013; Jusczyk et al., 1993; Lalonde & Werker, 1995; Rose et al., 2009; Singh, 2019; Tsao et al., 2004). (2) They have previously been shown to be correlated with later vocabulary (Conboy, Sommerville et al., 2008; Rose, Feldman, Jankowski, & Van Rossem, 2005, 2009; Tsao et al., 2004). (3) They tap different aspects of infant foundational skills and thus would allow us to assess the predictability of different infant measures for language outcomes. Specifically, the Stress task assesses sensitivity to prosodic information; the Vowel task assesses sound discrimination ability; the VRM task assesses infant recognition memory; and the A-not-B task assesses infant inhibitory control. We tested infants on these tasks at the youngest ages possible for each task. In addition, early identification of language predictors would allow for the most effective early intervention. We additionally measured receptive and expressive vocabulary at 18 months and expressive vocabulary at 24 months.

The overarching goal of this study was to examine the relative contributions of speech perception and cognitive skills to vocabulary outcomes from the same infants. Based on previous work, we predicted that all tasks correlate with later vocabulary; however, some infant measures will have better predictability for vocabulary outcomes compared with others. Specifically, (1) if vocabulary development recruits both domain-specific and domain-general processing mechanisms, then both sets of tasks may be equally good predictors; (2) if general cognitive skills are central, then cognitive tasks should be better predictors than speech perception skills; and (3) if vocabulary development draws most heavily on domain-specific processes, then speech perception skills should be better predictors than general cognitive skills.

2. Material and methods

Data and scripts are available from online supplementary materials https://osf.io/qd8as/ (Cristia, Wang, & Seidl, 2020).

2.1. Participants

Ninety-seven full-term typically-developing infants (43 female, 54 male) were included in the final analyses. They participated in the study between 2011 and 2015. Infants were from monolingual English-speaking (English spoken more than 90% of the time) homes in a Midwestern city in the United States. Experimenters informed caregivers of the broad interests and potential (minimal) risks of the study during the consenting process. The recruitment strategy and materials, and the study protocol were approved by an Institutional Review Board. During the first visit, infants were tested first on their perception of their native stress pattern (Stress) and

Table 1
Means (standard deviations) of infant ages (in months) and scores for the continuous measures gathered. No. indicates the number of infants who contributed data to that measure. Cohen's d indicates the Cohen's d against no preference (.5) for the three continuous metrics. Estimate (standard error) indicates the strength of association with vocabulary, from a Bayesian model in which only that measure (in interaction with age) is entered as predictor of vocabulary outcomes.

Measure	No.	Age (SD)	Score (SD)	Cohen's d	Estimate (SE)
Stress	97	5.68 (0.47)	0.51 (0.14)	0.053	−0.269 (1.37)
VRM	95	5.68 (0.47)	0.59 (0.07)	1.392	0.114 (1.38)
Vowel	95	6.92 (0.36)	0.52 (0.1)	0.173	0.043 (1.39)
18 m receptive	65	18.09 (0.53)	231.09 (95.42)	–	–
18 m expressive	65	18.09 (0.53)	70.17 (66.24)	–	–
24 m expressive	62	24.1 (0.56)	288.92 (188.34)	–	–

Note: Stress: stress pattern preference; VRM: visual recognition memory; Vowel: native vowel discrimination; 18 m receptive: receptive vocabulary size at 18 months; 18 m expressive: expressive vocabulary size at 18 months; 24 m expressive: expressive vocabulary size at 24 months.

then on the VRM task. During the second visit, they were tested first on the vowel discrimination task (Vowel) and then on the A-not-B task. For all the tasks, the caregivers were instructed not to intervene with their infants' behavior and wore opaque glasses during the VRM and Vowel task and headphones with masking music during the Stress task. The data were rejected if the caregivers interfered with the infants. Infants were given a short break (approximately 10 min) between the two paired tasks. Fifteen infants were unable to complete both the Vowel and the A-not-B tasks during the second visit. These children came into the lab for a third visit to complete A-not-B, which always took place fewer than 10 days after the second visit. Infants had to complete at least 3 of the 4 infant tasks (Stress, Vowel, VRM, A-not-B) in order for their data to be included in analyses. Data were missing for 2 children for vowel discrimination, 2 for VRM, and 3 for A-not-B. In all, 90 infants completed all four infant tasks. An additional 52 children were tested whose results were not reported because they fussed or cried, failed to show up due to illness or moving, or had experimenter errors during testing, resulting in fewer than 3 measures per infant. The average age and standard deviation of infants who were included in each task are reported in Table 1. Infants were compensated with a book or a toy for their participation at each visit.

2.2. Stress task

2.2.1. Stimuli

The stimuli for this task were based on previous research (Herold, Höhle, Walch, Weber, & Obladen, 2008). Specifically, a female speaker produced two segmentally identical consonant-vowel-consonant-vowel sequences (e.g., gába vs. gabá). The stimuli consisted of six trochaic (stress on the first syllable) and six iambic lists (stress on the second syllable), among which one trochaic and one iambic list was used during the familiarization phase and the rest of the lists were used during the test phase. Each list consisted of 15–16 tokens created from the recordings. The computer presented the lists alternating between trochaic and iambic lists with the type of the first stimulus and sides of presentation varied between participants just as in the original study.

2.2.2. Procedure

The Headturn Preference Procedure was used (Jusczyk & Aslin, 1995). Each infant was seated on a caregiver's lap in the middle of a 3-sided booth. The booth was quiet and comfortable and consisted of three panels: a center panel with a green light and two side panels each with a red light. Each trial began with the blinking of the green light on the center panel. When the infant looked at the green light, the light was extinguished and one of the two red lights began to blink. A computer program randomly chose which side light began to blink. When the infant oriented at least 30° in the direction of the red light, the stimuli for that trial began to play at approximately 70 dB SPL. The stimuli played until either the infant looked more than 30° away from the red light for 2 consecutive seconds or the stimuli file was complete. At this point, the red light was extinguished and the sound was stopped. Then the green light at the center began to blink in preparation for the next trial. Each child's head orientation was coded online by an experimenter who observed the child through the display of a video camera that also recorded the infant during the experiment. The computer recorded the amount of time the infant was oriented to the red light while the stimuli played. Orientation time was defined as the amount of time the infant spent looking at the red light. If the infant turned away from the target by 30° for less than 2 s, that time was not included in the orientation time, although the light did not extinguish and the sounds continued to play. Each experimental session began with a familiarization phase and was followed by a test phase. During the familiarization phase, the infants received a trochaic list and an iambic list to acquaint them with the task. Infants listened to the lists until they looked away, after which the test trial began. Both the experimenter and the caregiver wore headphones (Peltor Aviation headset 7050) that played continuous music to mask the stimuli the infant heard. Similar tasks have been administered on infants as young as 2 months (Jusczyk & Thompson, 1978).

The dependent measure for this task was the *Stress score*, calculated by dividing the orientation time to the trochaic lists by the total amount of orientation time to both the trochaic and iambic lists for each child. If infants showed a preference for the trochaic over the iambic pattern, then this score was higher than .5. Infants who prefer trochaic over iambic stress patterns (with the Stress score higher than 0.5) are expected to show larger vocabulary scores, because this preference is in line with the predominant pattern in their native language.

2.3. Vowel task

2.3.1. Stimuli

A female speaker produced many tokens of /ʃɪp/ "ship" and /ʃip/ "sheep" with a tense-lax contrast in the vowels using an infant-directed register. Four tokens of /ʃɪp/ that varied in pitch and intensity were selected for use in the habituation phase. Two additional tokens of /ʃɪp/ as well as a token of /ʃip/ were selected for the test phase. The /ʃip/ token was matched in intensity and pitch (minimum, maximum, and average) to one of the two test /ʃɪp/ tokens.

2.3.2. Procedure

Following Houston, Horn, Qi, Ting, and Gao (2007) this task used a hybrid version of visual habituation. Each infant sat on a caregiver's lap in a sound-attenuated room. The visual stimulus was a colorful image of a bull's eye. Speech stimuli were presented through a central speaker beneath the display at approximately 70 dB SPL. The image was shown first on the display to initiate the infant's fixation on the display, and then the auditory stimuli began to play. The auditory and visual stimuli continued to play until infants turned away. Caregivers wore darkened sunglasses to avoid looking at the display, and were asked to avoid interfering with their child's performance. Infants were habituated to the selected four different tokens of /ʃɪp/ until their average looking times in the preceding three trials dropped below 50 % of the longest looking trials. When the habituation criterion was reached, infants heard two types of test trials: same and alternating trials. During same trials, two new /ʃɪp/ tokens were presented, whereas in the alternating trials, one new /ʃip/ token alternated with a new /ʃɪp/ token. Similar tasks have been administered on 6-month-old infants (Warner-Czyz, Houston, & Hynan, 2014).

The dependent measure was the *Vowel score* calculated by dividing looking time to the alternating trials by the total looking time to the alternating and the same trials. We expected infants to show a novelty preference by looking longer to alternating trials than to same trials. If infants showed a novelty preference, then this score would be higher than .5. Infants who show a novelty preference (with the Vowel score higher than 0.5) are expected to show larger vocabulary scores, as a novelty preference suggests more sufficient stimuli processing.

2.4. VRM task

2.4.1. Stimuli

Stimuli were slightly modified digital versions of Rose et al. (2005); the stimuli consisted of black-and-white images of faces and colorful geometric patterns.

2.4.2. Procedure

The infant was seated on a caregiver's lap in front of a large display. The stimuli were projected onto this display. Infants were shown up to nine problems (average = 7; range = 2–9), each consisting of a familiarization and a test phase and had to complete at least 2 trials in order for their data to be included in the analyses. 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 non-speech sound tracks which varied across problems). When the infant looked at the display, visual stimuli were presented on the left and right side of the display until infant looked away from the screen for more than 2 s. During familiarization, the images to the left and right were identical: the same (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). The familiarization phase ended when the infant accumulated a fixed exposure time by looking at either side for 20 s for the faces, and 10 s for the geometrical patterns. During the test phase, the infant saw two images, one that had been shown in the familiarization phase, and another that was a new face (for the face problems) or a new geometric pattern (for the pattern problems). There were two test trials in each problem, with the side of the new image counterbalanced across them. The duration of each test trial was 10 s.

Videos of the infants' looking patterns to objects on the display were digitized at 30 frames per second and coded offline frame by frame by two highly trained coders using the SuperCoder (Hollich, 2008). Participants' looking directions to the left, right, or away in the test trials were coded. This procedure allowed us to determine the total amount of time infants spent looking at the familiar and novel images. A novelty score was calculated for each test trial by dividing the looking time to the novel image by the total looking time to either image. Similar tasks have been administered to infants at least as young as 6 months old (Oakes & Kovack-Lesh, 2013).

Infant behavior in this task was measured as a *VRM score* which was computed overall for all the problems the infant completed. Similar to the habituation paradigm used for native vowel discrimination, in this paradigm infants who show a VRM score above .5 are expected to show larger vocabulary scores, as the novelty preference suggests better recognition memory (Rose et al., 2005; Thompson, Fagan, & Fulker, 1991).

2.5. A-not-B task

2.5.1. Stimuli and procedure

In this task, we used the infant version of the A-not-B task described in Clearfield, Diedrich, Smith, and Thelen (2006). The experiment was conducted in a quiet room. The infant sat on a caregiver's lap across a table from an experimenter. A shallow yellow box (32 cm x 27cm × 7cm) was placed on the table within the reach of the experimenter and the infant. Two identical black lightweight metal objects with two bunny-ear-like bumps were placed in this box (object A and object B) approximately 8 cm apart. Object A was placed on the infant's left and object B on the infant's right, see Fig. 1.

At the beginning of the experiment, the experimenter made eye contact with the infant and greeted him/her by calling his/her name using infant-directed speech. At the same time, the experimenter touched the box and objects to engage the infant's attention on the box and the objects, until the infant was comfortable with them. Then the test, with five trials in total, began. On Trial 1, the experimenter lifted and waved object A for several seconds until the infant gazed at this object. Object A then was placed at the edge of the box (Fig. 2, Trial 1), which was closer to the infant than object B. The experimenter then moved the box forward so the 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

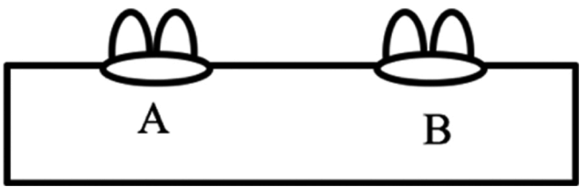

Fig. 1. Objects used in A-not-B task.

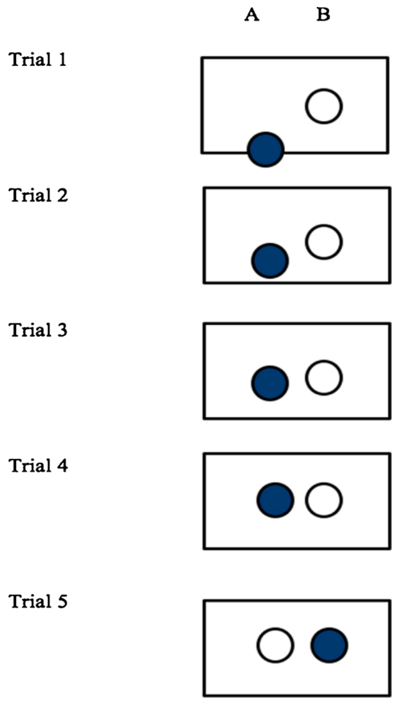

Fig. 2. Orders of trials for A-not-B task. The shaded blue object was the target. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

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 away from the infant (Fig. 2, Trials 2, 3, and 4) until it was in line with object B (Trial 4). On Trial 5 (see Fig. 2, Trial 5), the experimenter lifted and waved object B and placed it back to 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 whether the infant lifted object A (random) or Object B (hit). If in the course of the proceeding trials the infant consistently randomly grabbed the B object when the A object was indicated this was coded as error (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 and reaching for B), random (indicating a failure to inhibit a reach to the A object), or error (indicating a failure to complete the study due to inattention or inability). This task has been administered on infants as young as 5 months old (Clearfield et al., 2006). We expect that infants who lift Object B (hit), an indication of better inhibitory control, will show better vocabulary scores.

2.6. *Outcome measure: vocabulary*

The language outcome of children in this longitudinal study was assessed using the MacArthur-Bates Communicative Development Inventory, a reliable and stable parent survey for assessing language development in children from 8 to 30 months (Fenson, 2007). At the time when the infant who met the inclusion criteria reached 18 and 24 months, parents were sent the MacArthur-Bates Communicative Development Inventory or invited to the lab to complete the forms. At 18 months, the parent completed the Words and Gestures form. Vocabulary size was the number of words reported to be understood (18 m receptive) and produced (18 m expressive). At 24 months, the parent completed the Words and Sentences form. Vocabulary size was the number of words reported to be produced (24 m expressive). We report vocabulary size in Table 1.

3. Results

Analyses were implemented in the R environment (R Core Team, 2014), and are available in supplementary materials. Table 1 presents the summary statistics for each continuous measure. Prior to any analysis, we inspected the distributions of each measure using histograms and QQ plots. The distributions of the 3 continuous infant measures, namely, Stress, Vowel, and VRM, were approximately normal. However, the 3 vocabulary measures were not normally distributed. Therefore, we standardized the three outcome measures by applying z-scoring, which yielded an approximately normal distribution of the outcome measures.

3.1. Bivariate relationship among predictors

The first question one may ask is to what extent our different infant measures (Stress, Vowel, VRM, A-not-B) cluster together, with infants who score highly on one also doing so on the others. To answer this question, we examined the relationships among the four infant measures. Table 2 presents the bivariate correlations among the infant measures that were continuous (Stress, Vowel, VRM). Given that the A-not-B measure was categorical, we performed 3 separate one-way analyses of variance (ANOVAs) to examine whether there were differences on the other tasks with continuous data as a function of A-not-B outcome. Among the 94 infants with an A-not-B score, 23 were scored as “hit”, 35 as “random”, and 36 as “error”.

Our analyses showed that these associations were very weak, which does not seem in principle to be compatible with the idea that all of these variables are picking up the same underlying skills. The one exception was the significant positive correlation between infants’ performance in the Stress and the Vowel tasks, $r = .231$, $p = .024$, which was consistent with Seidl et al. (2014), which included a subset of our participants.

3.2. Prediction of vocabulary from each infant measure

We next explored the extent to which each infant measure predicted vocabulary scores. We did this by first examining the bivariate correlations between continuous infant measures and language outcome measures, shown in Table 2 (for continuous measures); and ANOVA results for A-not-B, shown in Table 3. These analyses were then complemented with a more complex Bayesian model to assess whether there was support for the null (i.e., no relationship between the infant measure and vocabulary), the alternative, or neither hypothesis.

Regarding bivariate correlations, we observed that vowel scores were marginally positively correlated with 18 m expressive vocabulary, $r\ (62) = .21$, $p = .089$, and 24m expressive vocabulary, $r\ (58) = .24$, $p = .070$ (see Table 2). Surprisingly, the other two continuous measures (Stress, VRM) showed very small correlations with vocabulary (maximally $r = .14$ for Stress and expressive 24 month vocabulary).

We complemented this frequentist statistical approach with a Bayesian mixed model, which has several benefits. First, since there was some variation in the exact age at which infants participated in the task, we declared this age in interaction with the measure, to allow for the possibility that a measure was more predictive at a younger or older age. Second, this allowed us to integrate all vocabulary outcomes together, declaring this as a repeated within-infant measure. Finally, Bayesian models provide evidence on the distribution of results via resampling, information that can be interrogated to assess to what extent evidence is consistent with the null (i.e., the infant measure does not predict vocabulary), the alternative (it does), or ambiguous between the two.

Using the brms packages (Bürkner, 2017), we fit *brm(VocabularyScore ~ Predictor*Age_at_predictor + Vocab.type + (1/Subject))* for each one of our four predictors separately, with weakly informative priors (student_t(3,0,1) for both intercepts and estimates), fitting 4 chains with 4000 iterations (500 warm-up) each. Full results of this analysis are in the online supplementary materials https://osf.io/hpw9f/-, and the information that is key to our discussion is provided in Table 4. This Table shows that none of our infant measures, declared in isolation, was an important predictor of vocabulary. Their estimates’ 95 % CI overlap with zero, with considerable mass of the distribution on both the positive and negative side. Although there was somewhat more support for the null than the alternative, all of the Bayes Factors were most consistent with the data being ambiguous between the two, as they were all very close to 1 (equal support for null and alternative).

3.3. Evaluating the relative importance of different infant predictors for vocabulary

To evaluate the relative importance of infant speech perception and cognitive measures in predicting vocabulary, we had decided to use a multiple linear regression mixed model. Note that to avoid potential biases, all analytical decisions on this frequentist model were made before we saw any statistical results. Specifically, we used lme4 (Bates, Mächler, Bolker, & Walker, 2015). We visually inspected univariate distributions, and found that the two data points from the Vowel task seemed to be univariate outliers. We therefore removed these two data points before fitting the model. The predictors were the individual performance scores in each of the four infant tasks (predictors: Stress, Vowel, VRM, A-not-B); we declared participant as a repeated measure, and controlled for gender and type of vocabulary measure (18 m receptive, 18 m expressive, 24 m expressive). The full model, fitted with the complete structure,

Table 2
Pearson correlation coefficients (and degrees of freedom) between Stress, Vowels, VRM, 18 m receptive, 18 m expressive, and 24 m expressive measures (z-scored).

	VRM	Vowel	18 m receptive	18 m expressive	24 m expressive
Stress	0.00 (93)	0.23 (93)*	0.02 (63)	0.03 (63)	0.14 (60)
VRM		0.04 (91)	−0.02 (62)	−0.03 (62)	0.09 (59)
Vowel			0.17 (62)	0.21 (62)~	0.24 (58)~
18 m receptive				0.60 (63)**	0.58 (54)**
18 m expressive					0.70 (54)**

Note: *: Correlation is significant at the .05 level (2-tailed); **: Correlation is significant at the .01 level (2-tailed); ~: Correlation is marginal $.05 < p < .1$ (2-tailed). Stress: stress pattern preference; VRM: visual recognition memory; Vowel: native vowel discrimination; 18m receptive: receptive vocabulary size at 18 months; 18m expressive: expressive vocabulary size at 18 months; 24m expressive: expressive vocabulary size at 24 months.

Table 3
Means (SD, N) for each of the 6 continuous measures, as a function of A-not-B score.

Task	Hit	Error	Random	*F*(df), *p*
Stress	0.47 (0.12; 23)	0.52 (0.16; 35)	0.52 (0.13; 36)	$F(91) = 1.29$, $p = 0.28$
VRM	0.61 (0.05; 23)	0.59 (0.06; 34)	0.59 (0.08; 35)	$F(89) = 0.93$, $p = 0.398$
Vowel	0.5 (0.09; 22)	0.52 (0.1; 35)	0.53 (0.1; 35)	$F(89) = 0.63$, $p=0.537$
18 m receptive	−0.01 (1.05; 19)	−0.12 (0.88; 21)	0.11 (1.08; 25)	$F(62) = 0.3$, $p=0.745$
18 m expressive	0.09 (0.91; 19)	−0.25 (0.69; 21)	0.15 (1.25; 25)	$F(62) = 1.03$, $p = 0.363$
24 m expressive	0.4 (1.05; 18)	−0.2 (0.89; 21)	−0.13 (1; 23)	$F(59) = 2.15$, $p = 0.125$

Note: *: *p* is significant at the .05 level (2-tailed); **: *p* is significant at the .01 level (2-tailed); † : *p* is marginal $.05 < p < .1$ (2-tailed). Stress: stress preference; VRM: visual recognition memory; Vowel: native vowel discrimination; 18m receptive: receptive vocabulary size at 18 months; 18m expressive: expressive vocabulary size at 18 months; 24m expressive: expressive vocabulary size at 24 months.

Table 4
Estimate (standard error), 95 percent credibility interval, and Bayes Factor for each measure as predictor of vocabulary (in a Bayesian mixed model declaring all vocabulary measures together), as well as the same metrics for this predictor's interaction with age in that model. Null > Alt indicates that the Bayes Factor shows more support for the null than the alternative equal to the ratio provided. No predictor or interaction showed more support for the alternative than the null.

	Predictor			Interaction with age		
Task	Estimate (SE)	95 % CI	Null > Alt BF	Estimate (SE)	95 % CI	Null > Alt BF
Stress	−0.27 (1.37)	[-3.34, 2.29]	1.08	0.16 (0.28)	[-0.36, 0.74]	3.85
VRM	0.11 (1.38)	[-2.57, 3.09]	1.1	0.46 (0.29)	[-0.11, 1.04]	1.07
Vowel	0.04 (1.39)	[-2.77, 2.82]	1.05	0.28 (0.25)	[-0.23, 0.78]	1.92
anotBHit	0.28 (1.35)	[-2.18, 3.23]	1.01	−0.02 (0.2)	[-0.45, 0.35]	6.47
anotBRandom	−0.06 (1.28)	[-2.65, 2.48]	1.00	−0.01 (0.19)	[-0.38, 0.37]	6.77

was *lm(VocabularyScore ~ Stress + Vowel + VRM + A-not-B + Gender + Vocab.type + (1/Subject))*. Model assumptions were assessed using gvlma() function in the gvlma package (Pena & Slate, 2012), which performs a global validation of linear model assumptions as well as separate evaluations of skewness, kurtosis, and heteroscedasticity. No evidence of assumption violation was found.

The model had an R^2 of .076, $F(8, 176) = 1.818$, $p = .076$. Vowel was the only significant predictor of vocabulary, $\beta = 2.12$, SE = .75, $p = .005$, see Fig. 3. Other infant predictors were not significant, $ps > .24$ (see Table 5 for regression results). These results suggested that vocabulary outcomes were higher in children who had achieved better discrimination scores in the Vowel task, and that this measure explained more variance than the others.

We complemented this model with a comparable Bayesian mixed model using the brms packages (Bürkner, 2017). The precise model was *brm(VocabularyScore ~ Stress + Vowel + VRM + A-not-B + Gender + Vocab.type + (1/Subject))*, with weakly informative priors (student_t(3,0,1) for both intercepts and estimates), fitting 4 chains with 4000 iterations (500 warm-up) each. We did not declare age here to simplify the models, because the single predictor analyses above consistently showed no interaction between the predictors and the age at which they were gathered. The full results of this analysis are in the online supplementary materials, and the information that is key to our discussion is provided in Table 6. This Table shows that none of our infant measures, declared in

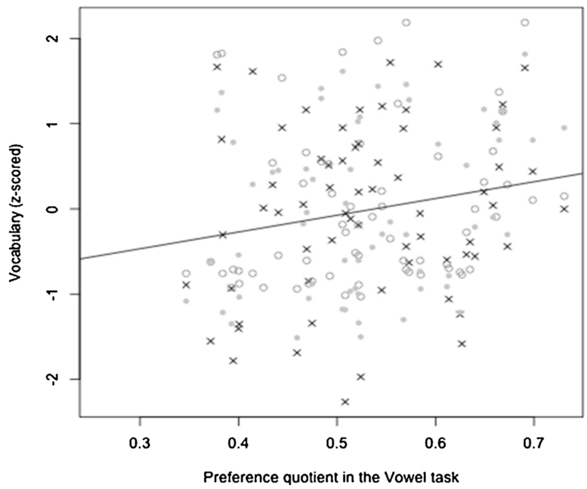

Fig. 3. Vocabulary size measured at 18 and 24 months as a function of infant performance in the Vowel task. The black crosses refer to 18 months comprehension, the dark gray open circles to 18 months expressive, and the light grey filled circles refer to 24 months expressive.

Table 5
Mixed linear regression model with fixed predictors of Stress, Vowel, VRM, and A-not-B scores.

Predictor	*b*	*b* 95 % *CI [LL, UL]*	*sr2*	*sr2* 95 % *CI [LL, UL]*	Fit
(Intercept)	−0.93	[-2.67, 0.80]			
Stress	0.31	[-0.87, 1.48]	0	[-.01, .01]	
Vowel	2.12**	[0.64, 3.61]	0.04	[-.01, .10]	
VRM	−0.4	[-2.68, 1.88]	0	[-.01, .01]	
A-not-B – Hit	0.19	[-0.16, 0.54]	0.01	[-.02, .03]	
A-not-B – Random	−0.2	[-0.53, 0.14]	0.01	[-.02, .03]	
Gender –M	−0.21	[-0.49, 0.07]	0.01	[-.02, .04]	
18 m expressive	0.01	[-0.33, 0.34]	0	[-.00, .00]	
24 m expressive	0	[-0.34, 0.34]	0	[-.00, .00]	
					R2 = *.076* 95 % CI[.00,.12]

Note: **: *p* value is significant at the .01 level (2-tailed); †: *p* is marginal .05 < *p* < .1 (2-tailed). Stress: stress pattern preference; VRM: visual recognition memory; Vowel: native vowel discrimination.

Table 6
Results from a multivariate Bayesian mixed model analysis. ß indicates the fitted estimate, LCI and HCI the lower and higher end of the 95 % credibility interval respectively, and BF10 the strength of the evidence supporting the alternative over the null estimated via a Bayesian Factor (values under .3 indicate support for the null, above 1 support for the alternative).

	ß	LCI	HCI	BF10
Intercept	−0.55	−2.28	1.135	0.46
Stress	0.384	−0.944	1.818	0.646
Vowel	1.055	−0.605	3.088	1.488
VRM	−0.177	−2.199	1.789	0.827
A-not-B – Hit	−0.119	−0.609	0.376	0.253
A-not-B – Random	0.158	−0.345	0.671	0.288
Gender –M	−0.192	−0.614	0.221	0.294
18 m expressive	0.035	−0.172	0.246	0.104
24 m expressive	−0.019	−0.241	0.201	0.101

isolation, was an important predictor of vocabulary: Their estimates' 95 % CI overlap with zero, with considerable mass of the distribution on both the positive and negative side. Unlike in the univariate Bayesian analysis, Vowel showed slightly more evidence for the alternative than the null, whereas all other metrics showed more support for the null than the alternative. Nonetheless, Bayes Factors between 1 and 3 are typically considered to be "anecdotal evidence". In this view, the result that was strongest pertains to A-not-B, where analysis of results' distribution revealed moderate evidence for the null hypothesis.

4. Discussion

Although it has long been clear that language learning recruits both domain-specific skills and domain-general abilities, less is known about their relative contributions in predicting language outcomes because there are few studies including both sets of infant measures and language outcomes in a longitudinal design. To contribute to closing this gap, we tested a large group of infants on two distinct speech perception (Stress, Vowel) and two cognitive (VRM, A-not-B) tasks and examined the predictive power of these skills for child vocabulary taken at 18 and 24 months.

Despite selecting tasks that had previously been found to predict vocabulary, we did not find strong bivariate correlations between our infant measures and childhood vocabulary. The strongest correlation was only $r \sim .2$, and it pertained to our Vowel measure. One interpretation of these results is that speech perception skills, and in particular phoneme discrimination, are more important for later vocabulary than the other skills we measured. If this hypothesis holds, these findings would be consistent with Conboy, Rivera-Gaxiola, Silva-Pereyraand, and Kuhl (2008) and Conboy, Sommerville et al. (2008), who also report that infant native vowel discrimination is a better predictor of concurrent vocabulary than cognitive skills. However, there are other interpretations of these results. Saliently, notice that the Vowel measure correlated with our other speech perception measure, preference for stress, that had been gathered a month earlier, whereas none of the other tasks correlated with each other. These results could thus indicate simply that our Vowel measure was a better measure of individual variation than the other measures.

In analyses that combined all of the infant measures, only the Vowel measure was a significant predictor. However, the model as a whole was not significant, and a complementary Bayesian analysis revealed that evidence was in the "anecdotal" region for most of the measures. The only case of clear support involved A-not-B, where moderate support for the null obtained. This ambiguity is born in mind for the rest of the discussion. Nonetheless, a main goal we had, after seeing these results, was to encourage the larger community to conduct more work along these lines to obtain less ambivalent data through larger samples and/or data aggregation. Therefore, in addition to discussing implications for each measure, we provide some recommendations for readers interested in undertaking similar research.

4.1. Vowel discrimination as a predictor of vocabulary

Some of our analyses were consistent with the view that performance in the Vowel task predicts later vocabulary, which is in line with a large literature demonstrating an association between infant sound discrimination skills and later language outcomes (Kuhl et al., 2005; Singh, 2019; Tsao et al., 2004). For example, both Tsao et al. (2004) and Kuhl et al. (2005) demonstrated an association between native vowel discrimination skills and later vocabulary. In addition, Singh (2019) showed that native sound discrimination was correlated with language outcomes in bilingual infants. Our results provide an important extension of this work. Such findings could lend support to language acquisition theories (e.g., PRIMIR, WRAPSA, NLM), which suggest that attention to language-specific properties leads to successful word recognition and representation.

The correlations we obtained, however, were weak, and evidence for the alternative hypothesis was underwhelming. We therefore wondered whether changes to our procedure and/or stimuli could lead to stronger predictive power. We chose the procedure inspired by Houston et al. (2007) who reported strong sensitivity to individual variation based on test-retest. However, since our data were collected, Cristia, Seidl, Singh, and Houston (2016) published a re-analysis of those and other test-retest data using the same procedure, which yielded an overall test-retest correlation that was close to zero. Those authors suggest that the few test trials found in such habituation-dishabituation designs may provide an overly noisy signal of infants' discrimination, and suggest testing infants with multiple contrasts in shorter, chained experiments – an idea that has not been explored, to our knowledge. Improvements may also be obtained via the choice of stimuli. A thorough inspection of the prediction literature reveals that there are actually few studies using vowel discrimination, and more employing consonant discrimination, and thus future studies should also explore the association between various types of contrasts, including both vowel and consonant discrimination.

4.2. Stress preference as a predictor of vocabulary

Not only did we not find a significant correlation between our Stress task and vocabulary development, but also estimates of this correlation were sometimes negative. This is surprising, given findings from previous studies seem to suggest that infants can discriminate between trochaic and iambic stress patterns as early as 2 months (Jusczyk & Thompson, 1978), and sensitivity to stress patterns is related to vocabulary development (Ference & Curtin, 2013). For example, Ference and Curtin (2013) tested 5-month-old infants' stress preference using a similar procedure, and collected vocabulary measures at 12 months. They showed that the trochaic stress preference score was positively correlated with receptive vocabulary. These seemingly contrasting findings may be due to differences in methodology, test-retest reliability, or the nature of the task.

Our Stress task followed previous research and used HPP procedure, in which the direction of preference can be either way. Previous studies examining the mechanisms leading to different preference directions suggest that infants may show novelty preferences when familiar stimuli have been completely processed and infants can thus devote their attentional resources to the unfamiliar stimuli; however, when the task is challenging and the representation and processing of the familiar stimuli are still difficult, infants would show familiarity preference (Hunter & Ames, 1988). This poses significant challenges for studies examining individual differences, as it is difficult to differentiate mature learners from immature listeners using the HPP. Our findings that infants as a group did not show either preference might be related to individual differences in the direction of preference, such that the advanced infants showed a novelty/familiarity preference, and less advanced infants showed no preference, results in a lack of preference at the group level.

The direction of preference has also led to mixed findings in the infant preference literature, which use preference scores as predictors for outcomes. For example, while Newman et al. (2006) showed that the size of familiarity preference in a speech segmentation study on infants aged 7.5–12 months was linked to later language development, they found the opposite pattern in a later speech segmentation study: Specifically, the size of the *novelty* preference from 7-month-old infants was related to vocabulary development (Newman et al., 2016). Given the nature of the direction of preference using HPP, differences in multiple factors, including infant age, developmental stage, stimuli complexity, and familiarization duration, etc. could all affect results. We return to this issue when discussing all predictors together in Section 4.5.

4.3. VRM as a predictor of vocabulary

We were particularly surprised by the fact that our VRM measure did not predict vocabulary outcomes in the second year. These findings differ from previously reported results showing an association between infant cognitive abilities and later language (Conboy, Rivera-Gaxiola et al., 2008; Rose et al., 2005, 2009), even at ages close to the ones tested here. For example, Fagan and Detterman (1992) and Fagan (1984) find that dishabituation in a visual paired comparison task (conceptually identical to Rose's VRM) predicts Peabody Picture Vocabulary Test results at about 36 and 60 months respectively. We do not have a clear explanation for these results, but we speculate that this might also relate to the direction of preference. Similar to HPP, previous studies have demonstrated both a familiarity and a novelty preference using VRM. For example, Rose, Gottfried, Carminar, and Bridger (1982) tested 3.5-, 4.5-, and 6.5-month-olds infants for VRM using shapes in Study 1. Only the 3.5-month-olds showed a strong familiarity preference, whereas the 4.5- and especially the 6.5-month-ods showed a novelty preference. In Study 2, they manipulated familiarization duration and tested 3.5- and 6.5-month-olds. Infants of both ages showed a familiarity preference when the familiarization duration was shorter (less than 10 s), but a novelty preference when the familiarization duration was longer (more than 15 s). These findings suggest that familiarity with the stimuli seems to play a role in the direction of preference using VRM, which could potentially introduce noise when examining individual differences.

4.4. A-not-B as a predictor of vocabulary

We included a measure of executive function via an adaptation of A-not-B, for which we found moderate evidence against this measure being related to vocabulary outcomes. This is also in contrast to other results, which link executive function to language development in childhood. For example, Gooch, Thompson, Nash, Snowling, and Hulme (2016) assessed children's executive function and language skills at ages 4, 5, and 6 years. They found a significant correlation between concurrent executive function and language skills; however, they did not find such relationship when the two measures were taken longitudinally. Notice that this study measured the relationship between executive function and vocabulary concurrently, whereas we are studying longitudinal prediction. Additionally, we are aware of the fact that most previous work using A-not-B has focused on older children, which may mean that measuring executive function at this age with this procedure is far from trivial. We return to this issue when discussing all predictors together in the next section.

4.5. Evaluating the relative importance of speech-specific and domain-general skills for language acquisition

We were excited to contribute to this important research question, but in view of our data, we want to first discuss some potential methodological concerns. To begin with, we noticed that the effect sizes indexing group performance in our speech-specific tasks (Vowel, Stress) are very small: Comparing against no preference (.5), we find $ds < .2$. One may wonder whether this means infants' performance is hovering around the lack of preference mostly due to noise in the measure. We provide two arguments against this interpretation. First, our data suggested that Vowel performance was (weakly) associated both with Stress performance gathered in infancy and with vocabulary scores at the end of the second year, incidentally showing the measure to be sensitive to meaningful individual variation. Second, we noted that the effect size for the Vowel task was clearly positive and larger than that for Stress, which was close to zero, and highlighted the fact that the two measures were gathered with different procedures. Vowel discrimination was measured in a habituation-dishabituation procedure and stress preference was measured using the HPP. It is possible that different procedures may yield different effect sizes. Indeed, a recent study comparing an HPP and a central-fixation with either eye tracking or manual coding on a word segmentation task showed different effect sizes in infant preference among the three procedures (Junge et al., 2020). Notice additionally that in one of our analyses (controlling for age at test) the estimate for the association between Stress and vocabulary scores was negative, which may suggest that infants who showed a preference for iambic stress patterns in this task tended to have larger vocabularies later.

In any case, group performance may not always be an index of how good a task is at detecting individual variation – a point that was made on the basis of results from 13 test-retest studies in Cristia et al. (2016). In the present study, we add one data point to that argument, since the strongest effect size was observed for VRM, which was not found to be a significant predictor of vocabulary (and which did not covary with any other infant measure). The fact that this effect size is large meshes well with this measure's long history in the field of infant cognition, as a robust and replicable phenomenon. Notice finally that our Bayesian results suggest undecided evidence for the association between VRM and vocabulary, with Bayes Factors very close to 1 (i.e., a similar level of support for the null and the alternative) in both the analysis looking exclusively at VRM and that looking at all predictors together.

It is also possible that the predictable strength of infant measures changes as a function of child age (Singh, 2019). We believe it is important to continue gathering data on these questions. The best way to adjudicate between these possibilities would be to assess a range of speech perception and cognitive skills at multiple points in development for the same infants, and determine which variables best predict language outcomes. To better understand the respective roles of speech perception and cognitive skills in language development, the combination of large-scale, multivariate, longitudinal studies with controlled experiments and modeling is necessary.

4.6. Limitations

Despite its relatively large sample size, our study obtained ambiguous results that could not have been predicted given the design of the tasks. Importantly, we employed the tasks where similar procedure and stimuli had led to significant correlations between the ensuing infant measures and subsequent vocabulary scores. Bayesian analyses suggest that data are ambiguous between null and alternative hypotheses, but it is unclear that greater sample sizes would have sufficed, given that bivariate correlations between several of the measures and vocabulary were very small. This leads us to recommend that data be gathered with different implementations of the same tasks. One difficulty facing those who follow our suggestion is that the only way to know whether a task correlates with vocabulary longitudinally is to gather it and wait – by which point one may find that it does not. Such a result may be hard to publish given current preferences for statistically significant results. Therefore, these authors should consider submitting their work as a registered report, and thus ensuring publication prior to data collection. In the meantime, we hope the present study serves as an inspiration to colleagues who have similar unpublished data to come forth and submit it to public scrutiny, so we can better learn what kinds of procedures and stimuli seem to be most robustly sensitive to individual variation.

5. Conclusion

We have reported on what may be the first prospective longitudinal study combining both speech perception and general cognitive measures as potential predictors of individual differences in vocabulary outcomes later in the development. We used frequentist statistical models and Bayesian models to test whether speech perception or cognitive measures are better predictors of child vocabulary. Despite selecting tasks that had previously been found to predict vocabulary, we did not find strong bivariate correlations between our infant measures and childhood vocabulary. We discussed several possible explanations including test sensitivity and test-retest reliability of infant measures. Future methodological development and research to assess individual differences in infant measures and their predictability for language development would provide valuable information for the developmental field.

Author statement

The research was conducted in accordance with APA ethical standards. The recruitment strategy and materials, and the study protocol were approved by an Institutional Review Board. Experimenters informed caregivers of the broad interests and potential (minimal) risks of the study during the consenting process.

CRediT authorship contribution statement

Yuanyuan Wang: Conceptualization, Methodology, Data curation, Writing - original draft, Project administration, Investigation. **Amanda Seidl:** Conceptualization, Methodology, Data curation, Resources, Supervision, Writing - review & editing, Funding acquisition, Project administration, Investigation. **Alejandrina Cristia:** Conceptualization, Methodology, Data curation, Software, Visualization, Writing - review & editing, Project administration, Investigation.

Declaration of Competing Interest

There are no conflicts of interest, financial, or otherwise.

Acknowledgements

We are grateful to the parents and infants who took time to participant in these studies. We thank many research assistants working at Purdue Infant Speech Lab, especially Rana Abu-Zhaya, Katie Courtaney, Lisa Helderman, and Carrie Wade for help in testing and recruiting; Rebecca Henaghan, Rachel Morrow, and Madeline Vulanich, for coding experimental data. The research was supported by the Purdue Research Foundation Grant from Purdue University. AC acknowledges funding from the Agence Nationale de la Recherche (ANR-17-CE28-0007 LangAge,ANR-16-DATA-0004 ACLEW, ANR-14-CE30-0003 MechELex, ANR-17-EURE-0017); and the James S. McDonnell Foundation Understanding Human Cognition Scholar Award.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.
Bürkner, P. (2017). An R package for bayesian multilevel models using Stan. *Journal of Statistical Software, 80*, 1–28.
Choudhury, N., Leppanen, P. H., Leevers, H. J., & Benasich, A. A. (2007). Infant information processing and family history of specific language impairment: Converging evidence for RAP deficits from two paradigms. *Developmental Science, 10*(2), 213–236.
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*(3), 435–444.
Colombo, J., Shaddy, D. J., Blaga, O. M., Anderson, C. J., Kannas, K. N., & Richman, W. A. (2009). Early attentional predictors of vocabulary in childhood. In J. Colombo, P. McCardle, & L. Freund (Eds.), *Infant pathways to language: Methods, models, and research directions* (pp. 143–167). New York, NY: Psychology Press/Taylor and Francis.
Colombo, J., Shaddy, D. J., Richman, W. A., Maikranz, J. M., & Blaga, O. M. (2004). The developmental course of habituation in infancy and preschool outcome. *Infancy, 5*(1), 1–38.
Conboy, B., Rivera-Gaxiola, M., Silva-Pereyraand, J., & Kuhl, P. K. (2008). Event-related potential studies of early language processing at the phoneme, word, and sentence levels. *Early Language Development: Bridging Brain and Behaviour*, 23. Retrieved from: syncii:///Event-related potential studies of early language.pdf.
Conboy, B., Sommerville, J. A., & Kuhl, P. K. (2008). Cognitive control factors in speech perception at 11 months. *Developmental Psychology, 44*(5), 1505–1512.
Cristia, A., Seidl, A., Junge, C., Soderstrom, M., & Hagoort, P. (2014). Predicting individual variation in language from infant speech perception measures. *Child Development, 85*(4), 1330–1345.
Cristia, A., Seidl, A., Singh, L., & Houston, D. M. (2016). Test–Retest reliability in infant speech perception tasks. *Infancy, 21*(5), 648–667.
Donnelly, S., & Kidd, E. (2020). Individual differences in lexical processing efficiency and vocabulary in toddlers: A longitudinal investigation. *Journal of Experimental Child Psychology, 192*, Article 104781.
Fagan, J. F. (1982). New Evidence for the predicton of intelligence from infancy. *Infant Mental Health Journal, 3*, 219–228.
Fagan, J. F. (1984). Recognition memory and intelligence. *Intelligence, 8*(1), 31–36.
Fagan, J. F., & Detterman, D. K. (1992). The Fagan test of infant intelligence: A technical summary. *Journal of Applied Developmental Psychology, 13*(2), 173–193.
Fagan, J. F., & McGrath, S. K. (1981). Infant recognition memory and later intelligence. *Intelligence, 5*, 121–130.
Fagan, J. F., & Singer, L. T. (1983). Infant recognition memory as a measure of intelligence. *Advances in Infancy Research, 2*, 31–78.
Fenson, L. (2007). *MacArthur-Bates communicative development inventories: User's guide and technical manual*. Baltimore, MD: Paul H. Brookes Publishing Company.
Ference, J., & Curtin, S. (2013). Attention to lexical stress and early vocabulary growth in 5-month-olds at risk for autism spectrum disorder. *Journal of Experimental Child Psychology, 116*(4), 891–903.
Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42*(1), 98–116. Retrieved from: syncii:///Picking Up Speed in Understanding Speech Processi.pdf.
Geers, A. E., Strube, M. J., Tobey, E. A., Pisoni, D. B., & Moog, J. S. (2011). Epilogue: Factors contributing to long-term outcomes of cochlear implantation in early childhood. *Ear and Hearing, 32*(1 Suppl), 84S–92S.
Gooch, D., Thompson, P., Nash, H. M., Snowling, M. J., & Hulme, C. (2016). The development of executive function and language skills in the early school years. *Journal of Child Psychology and Psychiatry, 57*(2), 180–187.
Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H Brookes Publishing.

Herold, B., Höhle, B., Walch, E., Weber, T., & Obladen, M. (2008). Impaired word stress pattern discrimination in very-low-birthweight infants during the first 6 months of life. *Developmental Medicine and Child Neurology, 50*(9), 678–683.
Hollich, G. (2008). *Supercoder: A program for coding preferential looking (Version 1.7.1). [Computer Software]*. West Lafayette: Purdue University.
Houston, D. M., Horn, D. L., Qi, R., Ting, J. Y., & Gao, S. (2007). Assessing speech discrimination in individual infants. *Infancy, 12*(2), 119–145.
Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In C. R.-C. L. P. Lipsitt (Ed.), *Advances in infancy research* (Vol. 5, pp. 69–95). Westport, CT, US: Ablex Publishing.
Johnson, E. K., & Seidl, A. (2008). Clause segmentation by 6-month-old infants: A crosslinguistic perspective. *Infancy, 13*(5), 440–455.
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. *Paper presented at the the 20th international congress on acoustics (ICA 2010)*.
Junge, C., Everaert, E., Porto, L., Fikkert, P., de Klerk, M., Keij, B., et al. (2020). Contrasting behavioral looking procedures: A case study on infant speech segmentation. *Infant Behavior and Development, 60*, Article 101448.
Junge, C., Kooijman, V., Hagoort, P., & Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Developmental Science, 15*(4), 463–473.
Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics, 21*, 3–28.
Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*(1), 1–23.
Jusczyk, P. W., & Thompson, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Perception and Psychophysics, 23*(2), 105–109. https://doi.org/10.3758/BF03208289.
Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*(3), 675–687.
Kuhl, P. K., Conboy, B., 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: Biological Sciences, 363*(1493), 979–1000.
Kuhl, P. K., Conboy, B., Padden, D., Nelson, T., & Pruitt, J. (2005). Early Speech Perception and Later Language Development: Implications for the "Critical Period". *Language Learning and Development, 1*(3 & 4), 237–264.
Lakens, D., McLatchie, N., Isager, P. M., Scheel, A. M., & Dienes, Z. (2020). Improving inferences about null effects with Bayes factors and equivalence tests. *The Journals of Gerontology: Series B, 75*(1), 45–57.
Lalonde, C. E., & Werker, J. F. (1995). Cognitive influences on cross-language speech perception in infancy. *Infant Behavior and Development, 18*, 459–475.
Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6*(3), F1–F10. https://doi.org/10.1111/1467-7687.00275.
Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowlege in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11*(3), F9–F16.
Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development, 66*(4), 911–936. https://doi.org/10.1111/j.1467-8624.1995.tb00913.x.
Newman, R. S., Ratner, N. B., 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*(4), 643–655.
Newman, R. S., Rowe, M. L., & Bernstein Ratner, N. (2016). Input and uptake at 7 months predicts toddler vocabulary: The role of child-directed speech and infant processing skills in language development. *Journal of Child Language, 43*(5), 1158–1173. https://doi.org/10.1017/S0305000915000446.
Oakes, L. M. (2017). Sample size, statistical power, and false conclusions in infant looking-time research. *Infancy, 22*(4), 436–469.
Oakes, L. M., & Kovack-Lesh, K. A. (2013). Infants' visual recognition memory for a series of categorically related items. *Journal of Cognition and Development, 14*(1), 63–86.
Pena, E. A., & Slate, E. H. (2012). *Gvlma: Global validation of linear models assumptions. R package version 1.0. 0.1*.
R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
Rose, S. A. (1989). Measuring infant intelligence: New perspectives. In M. H. Bornstein, & N. A. Krasnegor (Eds.), *Stability and continuity in mental development: Behavioral and biological perspectives* (pp. 171–188). Lawrence Erlbaum Associates, Inc.
Rose, S. A., & Feldman, J. F. (1995). Prediction of IQ and specific cognitive abilities at 11 years from infancy measures. *Developmental Psychology, 31*, 685–696.
Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2004). Dimensions of cognition in infancy. *Intelligence, 32*(3), 245–262.
Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2009). A cognitive approach to the development of early language. *Child Development, 80*(1), 134–150. https://doi.org/10.1111/j.1467-8624.2008.01250.x.
Rose, S. A., Feldman, J. F., Jankowski, J. J., & Van Rossem, R. (2005). Pathways from prematurity and infant abilities to later cognition. *Chlid Development, 76*(6), 1172–1184.
Rose, S. A., Gottfried, A. W., Carminar, P. M., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology, 18*, 703–713.
Salley, B., Panneton, R. K., & Colombo, J. (2013). Separable attentional predictors of language outcome. *Infancy, 18*(4), 462–489.
Scheel, A. M., Schijen, M., & Lakens, D. (2020). *An excess of positive results: Comparing the standard Psychology literature with Registered Reports*.
Seidl, A., French, B., Wang, Y., & Cristia, A. (2014). Towards establishing continuity in infant language. *Language Learning - Cognitive Neuroscience Supplement, 64*(2), 165–183. https://doi.org/10.1111/cdev.12193.
Sigman, M., & Norman, K. (1999). Continuity and change in the development of children with autism. *The changing nervous system: Neurobehavioral consequences of early brain disorders* (pp. 274–291).
Singh, L. (2019). Does infant speech perception predict later vocabulary development in bilingual infants? *Journal of Phonetics, 76*, Article 100914.
Singh, L., Steven Reznick, J., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: A longitudinal analysis. *Developmental Science, 15*(4), 482–495.
Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—Or vice versa. *Journal of the American Statistical Association, 54*(285), 30–34.
Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician, 49*(1), 108–112.
Thompson, L. A., Fagan, J. F., & Fulker, D. W. (1991). Longitudinal prediction of specific cognitive abilities from infant novelty preference. *Child Development, 62*(3), 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*(4), 1067–1084. https://doi.org/10.1111/j.1467-8624.2004.00726.x.
Tsuji, S., Cristia, A., Frank, M. C., & Bergmann, C. (2020). *Addressing publication bias in meta-analysis*. Zeitschrift für Psychologie.
Vouloumanos, A., & Curtin, S. (2014). Foundational tuning: How infants' attention to speech predicts language development. *Cognitive Science, 38*(8), 1675–1686.
Warner-Czyz, A. D., Houston, D. M., & Hynan, L. S. (2014). Vowel discrimination by hearing infants as a function of number of spectral channels. *The Journal of the Acoustical Society of America, 135*(5), 3017–3024.
Weber, C., Hahne, A., Friedrich, M., & Friederici, A. D. (2005). Reduced stress pattern discrimination in 5-month-olds as a marker of risk for later language impairment: Neurophysiological evidence. *Cognitive Brain Research, 25*(1), 180–187.
Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*(2), 197–234.

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology, 37*, 278–286.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3.

Werker, J. F., Pegg, J. E., & McLeod, P. J. (1994). A cross-language investigation of infant preference for infant-directed communication. *Infant Behavior & Development, 17*(3), 323–333.