DOI: 10.1111/infa.12471

INFANCY THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES WILEY

RESEARCH ARTICLE

Statistical learning in infancy predicts vocabulary size in toddlerhood

Anton Gerbrand[1] | **Gustaf Gredebäck**[1] | **Martina Hedenius**[2,3,4] | **Linda Forsman**[1] | **Marcus Lindskog**[1]

[1]Department of Psychology, Uppsala University, Uppsala, Sweden

[2]Department of Public Health and Caring Sciences, Speech-Language Pathology, Uppsala University, Uppsala, Sweden

[3]Center of Neurodevelopmental Disorders (KIND), Department of Women's and Children's Health, Division of Neuropsychiatry, Karolinska Institutet, Stockholm, Sweden

[4]Child and Adolescent Psychiatry, Stockholm Health Care Services, Region Stockholm, Stockholm County Council, Stockholm, Sweden

Correspondence
Anton Gerbrand, Department of Psychology, Uppsala University, Uppsala, Sweden.
Email: anton.gerbrand@psyk.uu.se

Funding information
Riksbankens Jubileumsfond, Grant/Award Number: P15-0430:1; Sundhal Disability Foundation; Knut och Alice Wallenbergs Stiftelse, Grant/Award Number: KAW 2012.0120

Abstract

During the first 2 years of life, an infant's vocabulary grows at an impressive rate. In the current study, we investigated the impact of three challenges that infants need to overcome to learn new words and expand the size of their vocabulary. We used longitudinal eye-tracking data ($n = 118$) to assess sequence learning, associative learning, and probability processing abilities at ages 6, 10, and 18 months. Infants' ability to efficiently solve these tasks was used to predict vocabulary size at age 18 months. We demonstrate that the ability to make audio–visual associations and to predict sequences of visual events predicts vocabulary size in toddlers (accounting for 20% of the variance). Our results indicate that statistical learning in some, but not all, domains have a role in vocabulary development.

1 | INTRODUCTION

Learning a first language is an enormous task. Already at 6 months of age, infants seem to have vocabulary knowledge about everyday words (Bergelson & Swingley, 2012; Karmiloff &

Karmiloff-Smith, 2009), suggesting that vocabulary learning starts well before 6 months, and continues at a brisk pace throughout childhood (Kidd & Donnelly, 2020). By age 18 months, children on average have a vocabulary of roughly 200 words (Dale & Fenson, 1996), extending to approximately 2600 words by the time they reach school age (Kidd & Donnelly, 2020). However, the mechanisms behind this rapid increase in infant vocabulary are currently not well understood.

Statistical learning is posited to be one important, early contributor to vocabulary size (Frost et al., 2015; Saffran, 2009; Smith et al., 2018). Often conceptualized as a general ability to detect and learn from statistical regularities and probabilistic relations in the environment, it allows infants to detect and learn from statistical structures present within languages and to segment words in a given language (Arciuli & Conway, 2018; Romberg & Saffran, 2010; Saffran, 2009). Word segmentation, in turn, has been proposed to be an important building block for early vocabulary learning—as it aids in constraining the beginning and end of words (e.g., it is "you and I", not "youa ndi") and the definition of words (Karmiloff & Karmiloff-Smith, 2009; Kidd & Donnelly, 2020). Infants can detect transitional probabilities (the likelihood that one syllable follows another) in a given language as well as tell apart speech sounds by attending to phonetic variation within artificial languages, and then use this information to perform word segmentation (Lany & Saffran, 2013; Maye et al., 2002; Romberg & Saffran, 2010; Saffran et al., 2009). Thus, statistical learning seems to support specific mechanisms of building vocabulary.

Although conceptualized as a general ability to detect and learn from statistical regularities (Arciuli & Conway, 2018; Romberg & Saffran, 2010; Saffran, 2009), statistical learning is often more narrowly operationalized as the ability to extract transitional probabilities in streams of sounds from natural or artificial languages. Furthermore, it has been argued that statistical learning, as traditionally operationalized, might include additional cognitive mechanisms (e.g., short-term memory) to the ability to detect and learn from statistical regularities and probabilistic relations (Arciuli & Conway, 2018; Siegelman et al., 2017). Related to this, there are indications that traditional statistical learning measurements are not very reliable (Cristia et al., 2016). While some have argued that this makes it difficult to capture individual differences in a meaningful way (Arnon, 2020; Siegelman et al., 2017), others have argued for a multi-component view of statistical learning, including separable components of encoding, retention, and abstraction (Arciuli & Conway, 2018). A recent review highlights the need to move from a narrow definition of statistical learning to a more ecologically relevant model that acknowledges that multiple mechanisms and process levels are involved in learning regularities in real life, both within and beyond the linguistic domain (Frost et al., 2019). In other words, infants should need the ability to process statistical information in general in order to detect statistical structures in languages (e.g., transitional probabilities) in the first place (Arciuli & Conway, 2018; Frost et al., 2019).

Here, we take a novel complementary approach to previous research, in line with the suggestion by Frost, et al. (2019), and investigate the contribution of statistical learning to vocabulary size by conceptualizing statistical learning as a broad, multicomponent ability to detect and learn from statistical regularities and probabilistic relations in the environment. To do so, we identified a set of three everyday statistical challenges, outlined below, that infants face when learning a language. Thus, instead of focusing on a narrow set of statistical properties in the linguistic domain, we turned our attention to a broader set of challenges, outside the linguistic domain, that require processing of statistical information in general, and ask how they are related to vocabulary size. Importantly, we suggest that these three challenges are distinct in the sense that solving all of them is potentially important for vocabulary learning and that infants could face any combination of them at the same time.

A first challenge for infants is the need to make one-to-one mappings between cross-modal, co-occurring stimuli, referred to here as "associative learning". That is, infants need to be able to associate, for example, linguistic information with visual information. We propose that this challenge is necessary for infants to learn the relation between labels (e.g., hearing "*dog*") and their referents (e.g., seeing a dog), aiding in learning the meaning of words and, consequently, expanding vocabulary (Kidd & Donnelly, 2020). Further, attending to regularities within the environment should aid in detecting co-occurring stimuli (Saxton, 2010). For example, when an infant sees a dog, it will be followed by the word "*dog*", regardless of context. To detect that this association exists in diverse environments should further help constrain the meaning of the word "*dog*".

A second challenge is the need for infants to detect when, where, and how often stimuli follow each other. That is, to detect, process, and learn patterns from sequential information (Frost et al., 2015; Romberg & Saffran, 2010; Saffran et al., 1996; Saffran et al., 2009), referred to here as "sequence learning". It has previously been suggested that sequence learning is involved in detecting statistical regularities and grammatical structures within a given language, via the mechanism of prediction (Reuter et al., 2018; Saffran, 2009). For example, a child might predict to hear the word "*dogses*", but instead hear "*dogs*"—providing a meaningful feedback loop on grammatical rules (Reuter et al., 2018). Thus, sequence learning might aid in constraining grammatical rules such as singular/plural, which in turn allows for learning even more words. Indeed, previous research has shown that the ability to predict patterns of visually presented stimuli is concurrently related to lexical processing efficiency (Shafto et al., 2012), vocabulary size (Reuter et al., 2018), reading efficiency (von Koss Torkildsen et al., 2019; Hedenius et al., 2021), and syntax comprehension (Kidd & Arciuli, 2016), indicating a relation between predicting patterns in both verbal and nonverbal domains and language ability.

Third, infants face the challenge of evaluating probabilistic information in general, here referred to as "probability processing". We propose that probability processing should aid in detecting regularities within linguistic data (among others) and constrain possible meanings of words. Indeed, in everyday life, infants face an enormous amount of data, containing different kinds of regularities which span from simple to complex events, and between very likely events to very unlikely events (Frost et al., 2019). Therefore, detecting meaningful regularities from the environment is a very difficult task (Frost et al., 2019). It has been proposed that general probabilistic processing allows even young children to form intuitive theories of their world, constraining the possible space of candidate theories (Ullman et al., 2012). For instance, recognizing that the word "*dog"* is used for animals with four legs, that have fur, and a nose, but at the same time is not used for for example, cats, informs children about the definition and meaning of the word "*dog*". It is thus possible that the challenge of word learning requires a more general ability to evaluate and integrate the probability of events (Frank, 2011; Frank & Tenenbaum, 2011; Frost et al., 2019). Without such an ability, making hypothesis about potential patterns and associations in vast amounts of linguistic data would be a very difficult task (Frost et al., 2019). Thus, probability processing might help infants constrain hypothesis about the meaning of words, which in turn should aid in developing vocabulary.

Here, we targeted the three challenges described above, and explored how they might contribute to building vocabulary in infancy. To do so, we used three eye-tracking tasks that tapped into one of the challenges from a longitudinal data set following children from 6 to 18 month of age. This data set was used because there were eye-tracking tasks already implemented and data collected that allowed us to assess the three challenges noted above. However, the larger longitudinal project was not designed with statistical learning specifically in mind (see method for further detail). Using eye-tracking and a longitudinal data set, we were able to broaden the scope of previous research by taking an individual difference approach and measuring all infants on a set of three tasks at multiple time points.

We measured the ability to solve the first challenge, related to associative learning, with an associative learning task, involving an assessment of infants' ability to detect an association between a visual stimulus and an auditory, nonlinguistic sound cue. This task is a proxy for being able to keep track of how stimuli in different domains co-occur (Richardson & Kirkham, 2004). A relation between individual differences in this task and vocabulary size would indicate that the ability to form associations across modalities in general influence vocabulary learning. To measure a capacity to solve the second challenge, related to sequence learning, we used a visual sequence task—an assessment of infants' ability to extract and predict patterns of reoccurring visual events. This task is a proxy for being able to keep track of when, where, and how often stimuli follow each other (Sheese et al., 2008). If infants' efficiency in solving this task predicts later vocabulary, it would indicate that sequence learning in general influence vocabulary learning. Finally, we used a probabilities task to measure the ability to solve the third challenge, related to general probability processing. This task is an assessment of infants' ability to respond to the likelihood of samples being drawn from a larger population (Kayhan et al., 2018). Similar tasks have previously been used to argue that infants are intuitive statisticians (Xu & Garcia, 2008). Accordingly, we consider our task to be a proxy for a general sensitivity toward probabilities. If this task is related to vocabulary size, it would indicate that a general ability to evaluate the likelihood of events influence vocabulary learning. Finally, we collected data on vocabulary size at 18 months of age (Eriksson et al., 2002). Our goal was to predict vocabulary size from individual differences in the ability to process regularities (patterns and associations) at 6–18 months.

Before any analysis, we identified the three tasks that captured different aspects of the three challenges from tasks available in the longitudinal project. For analyzing our longitudinal data, we used path modeling. As stated above, we argue that infants' abilities to solve the three challenges could potentially be related, but we made no hypotheses concerning the possible directions of such relations. One reason for this is that infants' performance on a specific task might not always fall on a continuum in an expected way. Hence, it is difficult to hypothesize how individual differences might be related to each other. However, we hypothesized that there are individual differences in how infants solve the three challenges that in turn can predict individual differences in vocabulary size, which would indicate that processing statistical information, in a broad sense, plays a role for vocabulary size.

2 | METHOD

2.1 | Participants

The study sample participated in the BASIC Child project, a longitudinal study aiming to assess 120 infants from 6 to 30 months of age (see Supporting Information 1). Exclusion criteria included serious physical health problems, such as premature birth and neurological issues. The initial sample consisted of 118 six-month-old infants and their families (50% females, *M* age at the sixth month lab visit = 185 days; SD = 7 days, min = 165 days, and max = 203 days). Of this initial sample, 110 (93%) returned for the second lab visit at 10 months of age (*M* age = 302 days; SD = 9 days, min = 289 days, and max = 326 days). Within the longitudinal project, participants also visited the lab at 12 months of age. At this visit, they participated in a "strange situation" and an eye-tracking task unrelated to the current project. At the fourth lab visit, at the age of 18 months (*M* age = 544 days; SD = 12 days, min = 524 days, and max = 583 days), 104 participants (88%) returned.

The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Etikprövningsmyndigheten (2013/423) at the department of psychology, Uppsala University.

Swedish was spoken in the home of 99.7% of the participating families, and in 14% of families, one or more additional language was spoken. Furthermore, all participants attended a Swedish-speaking preschool at 18 months of age.

Approximately 28% of mothers reported having less than 12 years of education (min = 11 years of education, max = 15 years or more, and M = 13–15 years). In Sweden, 29% of women have less than 12 years of education, meaning our sample was close to the national average (Statistiska Centralbyrån, 2019).

The target sample size of the project (n = 120) was set before the recruitment of participants, and data collection began. It was based on both practical (e.g., feasibility to carry out the data collection within the project time, see SOM for more details) and statistical (e.g., feasibility to test models with measurements taken from several tasks at different time points) considerations. Notably, our sample size is larger than those found in several previous studies aimed at predicting children's language development (Kidd & Arciuli, 2016; Reuter et al., 2018; Singh et al., 2012; von Koss Torkildsen et al., 2019). To date, three published papers have used data from the BASIC Child project, focusing on internal models, action prediction, and action evaluation at 6 and 10 months of age (Gredebäck et al., 2018), the social components of executive functions (Marciszko et al., 2020), and how gaze following is affected by social and emotional contexts (Astor et al., 2020).

2.2 | Tasks and procedures

All eye-tracking tasks were conducted using a corneal reflection eye tracker (60 Hz, Tobii TX300; Tobii Technology). During these tasks, the infants sat on a parent's lap at approximately 60 cm viewing distance in front of a 23-inch monitor. Prior to the start of the data collection, we ran a standard 5-point calibration procedure (Gredebäck et al., 2009), which was repeated until calibration was satisfactory. All infants (at 6, 10, and 18 months of age) observed the same order of stimuli in the tasks used. See Table 1 for an overview of at what visit the different tasks were administered. The eye-tracking tasks were selected from an existing longitudinal data set. While other tasks might, in principle, have been selected, these three were deemed the most relevant for the current research question. Importantly, we decided on the tasks before any analyses of associations or paths. A full list of tasks used in the longitudinal study and a video recording of each eye-tracking session can be found at Databary. We adopted a similar rationale for which ages to include. The exact ages were decided by the larger longitudinal study (i.e., 6–30 months of age) from which this data is taken. Within the confines of this data set, we focused on analyzing longitudinal data from 6- to 18 months of age. This allows us to assess the early potential correlates of language development specified above (from 6 to 18 months) and for us to relate these to individual differences in early stages of vocabulary at 18 months. Note that all tasks are not available at all ages and that this is an effect of earlier design considerations for the longitudinal project from which the data is taken (for detailed info see Table 1).

TABLE 1 Overview of at what visit the tasks were administered, and number of excluded participants and proportion of excluded participants for each task, at each lab visit

	Visit		
Task	**6 months**	**10 months**	**18 months**
Associative learning	14 (0.12)	22 (0.20)	
Visual sequence		45 (0.41)	22 (0.21)
Probabilities task		51 (0.46)	47 (0.45)

2.2.1 | Associative learning task

This task assessed the ability to detect associations between visual and auditory stimuli. The paradigm was adopted from Experiment 2 in Richardson and Kirkham (2004). In a small sample of 10 participants, they demonstrated that 6 month-old infants could detect and learn an association between visual and auditory stimuli. However, in contrast to Richardson and Kirkham (2004), we focused on infants' individual differences rather than group performance. Within the longitudinal project, pilot data indicated that this task worked well at 6 and 10 months of age, which is why the task was implemented at these ages. However, the task was not implemented at 18 months of age, as to not tax the participants with a too large test battery.

In each trial, the participants saw two white frames on a black background, one on the right and one on the left side of the screen (Figure 1a). Participants were shown two blocks of trials, each containing six familiarization trials and two test trials. In total, the two blocks of trials lasted 112 s. Each familiarization trial lasted 7 s, and participants were presented with a visual stimulus within one frame (leaving the other frame blank), while a unique auditory stimulus played simultaneously. The first block of familiarization trials started with a visual stimulus on the left side of the screen (and then continued right, left, right, etc.). The second block of familiarization trials started with a visual stimulus on the right side of the screen. In each block, participants saw three visual stimuli being presented on the left side of the screen and three visual stimuli on the right side. Each block of trials had two unique visual stimuli paired with two unique auditory stimuli. For instance, in one familiarization trial, a dog presented on the right side of the screen was paired with a trumpet sound. In the following familiarization trial, a chicken presented on the left side of the screen was paired with a whistling sound. In the test trials, only the two blank frames and auditory stimulus were presented; for example, only a whistling sound was presented.

The outcome variable was defined as the proportional looking time toward the "critical frame" in test trials—i.e., the frame where the visual stimuli had previously been presented. The outcome was calculated such that the looking time toward the critical frame was in proportion to the total looking time toward both frames (thus ranging from 0.0 to 1.0; see SOM for further details). We required

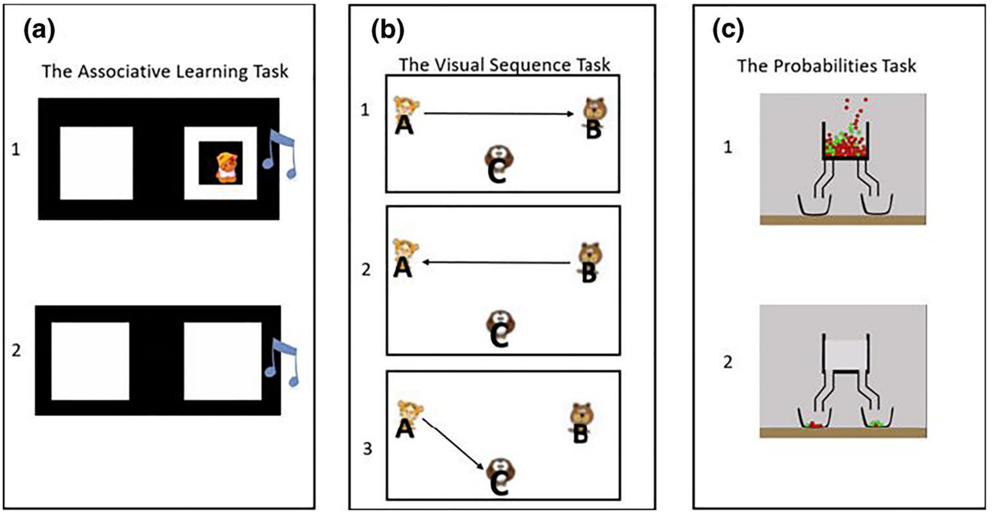


FIGURE 1 Snapshots of the stimuli used in the three eye-tracking tasks. Panel (a) Illustration of the associative learning task. Panel (b) Illustration of the visual sequence task. Panel (c) Illustration of the probabilities task

participants to have a minimum of two valid trials to be included in the analyses, as to not present ambiguous proportional looking time data. The final sample size at 6 months of age was 106 participants, and at 10 months of age it was 98 participants (see Table 1 for an overview).

2.2.2 | Visual sequence task

This task assessed the ability to extract patterns from a regular presentation order of visual events. The paradigm was adopted from Sheese, Rotbart, Posner, White, and Fraundorf (Sheese et al., 2008), who assessed infants' ability to predict events within a visual pattern. In a group of 50 participants at 6–7 months of age, they demonstrated that infants could, on average, correctly predict 4% of possible events. Importantly, even though the amount of correct predictions were low, they were related to individual self-regulation (Sheese et al., 2008). For the current study, we made slight modifications to the paradigm. Within the longitudinal project, this task was conducted at 10 and 18 months of age but not at age 6 months. The reasons for this exclusion of this task from the 6 month battery were twofold. First, in the pilot phase of the longitudinal project, pilot data suggested that this task worked best at 10 and 18 months. Second, there was a general need to keep the duration of the 6 month testing as brief as possible, as to not introduce high levels of fatigue.

Participants were shown visual stimuli (cartoon animals, e.g., cartoon tiger, cartoon dog, or cartoon owl), accompanied by a unique sound, presented in a spatial pattern on the screen. The sound was played only to keep the attention of the participants. The visual stimuli were presented at three different locations, A, B, and C (Figure 1b). The location and presentation order of the visual stimuli were kept constant for the entire task, without familiarization trials. The participants were presented one visual stimulus at a time, starting with a stimulus appearing for one second at location A. After a one-second blank screen, another stimulus appeared at location B, and so on. The stimuli were presented in the specific order of A → B → A → C, repeated for 180 s (Figure 1b). After 35 s, the visual content of each stimulus was changed (e.g., from an image of a tiger to that of a bee) to keep the attention of the participants. When the visual stimulus changed, so did the auditory stimuli. Thus, no auditory stimuli stayed constant during the task, effectively making it impossible to learn the pattern based on auditory information. In all, the participants saw five different combinations of cartoon animals.

The outcome measure was defined as the number of correct predictions, in line with Sheese et al. (2008). A correct prediction was defined as one where the participant looked at the correct location (i.e., where the next visual stimuli appeared) within a time window extending 200 milliseconds prior to, and following, the onset of the stimuli (see SOM for further details). The Visual Sequence task was not implemented in the longitudinal project for the first 28 participants at the 10-month-old visit, as the task struggled with implementation issues at the launch of the study. The final sample size at 10 months of age was 75, and at 18 months it was 98 participants (see Table 1 for an overview).

2.2.3 | The probabilities task

This task assessed the ability to respond to the likelihood of samples being drawn from a larger population, adapted from Kayhan, Gredebäck, and Lindskog (Kayhan et al., 2018). The original study demonstrated that 6, 12, and 18 month-old infants changed their looking pattern as a function of the relative likelihood between likely and unlikely outcomes (being 625, 81, 25, and 9, respectively). When the relative likelihood was low (e.g., nine, where the likely outcome is nine times more likely to be randomly drawn from a larger population, compared to the unlikely outcome), infants at all ages looked at the likely outcome more than the unlikely outcome. As the relative likelihood increased, they shifted their gaze and looked more at the unlikely outcome than the likely. While the original study investigated performance on a group level, across different levels of relative likelihoods, we investigated individual performance at one level of relative likelihoods instead. This task was presented to infants at ages 10 and 18 months in the longitudinal project because effects in the original study (Kayhan et al., 2018) were smaller at 6 months of age than in the older infants.

The task started with a set of colored balls (red or yellow) falling from the top of the screen into a rectangle-shaped container. For each trial, one color was more common in a ratio of 3:1. In total, participants saw four trials, totaling 68 s. Once all balls had landed in the container (taking 8 s), the container was covered with a light gray occluder so that balls were no longer visible (taking 2 s). Next, participant attention was drawn to the bottom of the container with a flashing light, accompanied by the sound of sirens (taking 2 s). Then, two doors at the bottom of the container were opened, and two samples of balls fell into two smaller separate containers at the bottom of the screen (2 s). For the remaining trial time (3 s), the two samples lay static in their separate containers, allowing the participants to compare the two outcomes. As shown in Figure 1c, one sample was of a likely outcome, while the other was of an unlikely outcome. The relative likelihood of the containers was 81, meaning that the likely outcome was 81 times more likely to be drawn randomly from the population than the unlikely outcome. The location of the unlikely and likely outcomes varied over trials.

The probabilities task was intended to tap into how efficiently infants could respond to likely and unlikely events. The outcome variable was defined as the proportional looking time toward the unlikely outcome, per Kayhan, Gredebäck, and Lindskog (Kayhan et al., 2018). The outcome was calculated such that the looking time toward the unlikely outcome was in proportion to the total looking time toward both the likely and unlikely outcome (thus ranging from 0.0 to 1.0, see SOM for further details). For a valid trial, we required participants to have looked at both the likely and unlikely sample, as participants should need to look at both samples to detect the unlikely one. We required participants to have a minimum of two valid trials to be included in the analyses. As such, the final sample size at 10 months of age was 69 and at 18 months of age it was 73 (see Table 1 for an overview).

2.2.4 | The CDI questionnaire

At the 18th month lab visit, mothers filled out a Swedish version of the CDI, designed by Eriksson, Westerlund, and Berglund (Eriksson et al., 2002). For each item, the mothers were instructed to answer if the child understood or understood and said the word. If the child in question did not yet comprehend the word, the mothers were instructed to leave the response blank. Thus, the CDI consisted of two subscales: receptive vocabulary (the number of words the participants understood) and expressive vocabulary (the number of words the participants understood and could say; see SOM for more details). Eriksson et al. (2002) demonstrated high internal consistency for both the receptive (Cronbach's alpha = 0.96) and the expressive subscale (Cronbach's alpha = 0.97) subscales of the CDI. Raw scores, rather than age corrected values, were used because all infants were measured at the same age (in months). Within the longitudinal project, it was reasoned that assessing vocabulary at 18 months of age was the most appropriate, as this allowed for an assessment of early vocabulary, while still allowing for a good level of variation.

2.3 | Data reduction

The raw eye-tracking data were exported from Tobii Studio (Tobii Technology, Sweden) and pre-processed using an open source analysis program, TimeStudio (Nyström et al., 2016), operating in MATLAB (Version 9.3.0.7; The Mathworks, Natick, MA). For all eye-tracking tasks, we required a minimum of 25% looking time to the screen for individual trials to be included. For more information on data reduction, see SOM. Data and settings for the eye-tracking data analysis can be downloaded from: https://osf.io/9zmvr/?view_only=2940a3f831be4330a97d57965418cd48.

Table 1 summarizes which task participants did at what visits as well as the number of participants that were excluded from each of the eye-tracking tasks due to not providing enough data.

2.4 | Statistical analysis

To address the main hypotheses of the study, we conducted our analysis in three steps. For all tasks, we relied on standardized measures previously used in the literature. Notably, the measures were decided upon before conducting the main analysis listed below (Pearson zero-order correlations, *t*-tests, and path model). There are always additional ways in which variables of interest could be analyzed (e.g., calculating multiple outcome variables for each predictor). However, we refrained from doing so, as to not introduce a plethora of possible variables and analyses that could quickly exceed what the available data could handle.

For descriptive analyses, we first calculated zero-order correlations between all predictors and outcome variables. We modeled the two subscales of the CDI separately to gain a rich understanding of early predictors in vocabulary size. For zero-order correlations, we refrained from correcting for multiple comparisons since the goal of this step was to understand the structure of the data before conducting the path model that accounts for all data in one analysis (thus circumventing the issue of multiple comparisons), rather than to probe for statistical significance. That is, this step was only conducted in order to get a sense of the feasibility of our planned path model and for descriptive purpose. Any and all conclusions made in this paper is based on the results of the path model and not the zero-order correlations. Second, we also performed single-sample *t*-test for the associative learning task and the probabilities task in order to get a sense of group level performance. That is, we tested if the proportional looking time toward target (the critical frame and the unlikely outcome, respectively) was significantly different from chance (i.e., 0.5). As it was not possible to conduct a meaningful singe-sample *t*-test on the visual sequence task, here we report proportion of correct predictions.

As our main analysis, we created a path model using the *lavann* package in R (Rosseel, 2012). The model was set up so that individual differences on earlier tasks could influence individual differences on all subsequent tasks. That is, the model was structured so that we could investigate if individual differences on the early outcome measurements could predict later outcome, and if any task was predictive of the CDI subscales. Thus, all variables used in the zero-order correlation were used in the path model.

Some of the predictors had missing data, particularly the visual sequence task at 10 months and evaluation of probabilities at both 10 and 18 months. We chose to handle the missing data by fitting our path models using full information maximum likelihood (FIML) estimation. With this method, parameters in the model were estimated using all available data. Previous research has demonstrated that this is a robust method for handling a relatively large amount of missing data. It produces relatively small bias and sampling variability when estimating the model parameters. Furthermore, the

FIML method can also handle different types of missing data reasonably well, making it quite robust against data that is missing not at random (Enders, 2001).

The model-to-data fit was evaluated using the chi-square goodness of fit as well as several alternative fit indices, including the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and the standardized root mean square residual. Of note, and contrary to traditional inferential statistics, the chi-square goodness of fit is testing if the model is *not* different from the null hypothesis. As such, if the chi-square shows a significant result, this would suggest that the structure of the model does *not* add any explained variance in the dependent measure (i.e., the CDI subscales; Hu & Bentler, 1999). For a model to be considered as having a good fit, the chi-square test should not be significant, the CFI and TLI should be equal to or greater than 0.95, the RMSEA should be less than 0.05, and the standardized root mean square residual should be less than 0.08 (Hu & Bentler, 1999; Maccallum et al., 1996). We investigated possible multivariate outliers by calculating a generalized Cook's distance for the full model, and cases with a Cook's distance greater than 1 were excluded. This procedure resulted in 15 cases (i.e., individuals) being excluded from all tasks, and a final sample of 103 participants being used to fit the model.

3 | RESULTS

Descriptive statistics. On the associative learning task at 6 months, the mean proportional looking time toward the target was 0.51, SD = 0.14, min = 0.14, and max = 0.89, which was not significantly different from chance, $t(105) = 0.68$ and $p = 0.50$. At 10 months of age, the mean proportional looking time was 0.49, $SD = 0.14$, min = 0.13, and max = 0.83, which was not significantly different from chance, $t(92) = -0.53$ and $p = 0.60$.

On the visual sequence task, participants made 4.2 ($SD = 4.5$, min = 0, and max = 25) and 5.6 ($SD = 4.6$, min = 0, and max = 20) correct predictions at 10 and 18 months, respectively. That is, at 10 months of age, participants made, on average, 7.5% of possible correct predictions—while at 18 months of age, participants made 9.3% of possible correct predictions. These numbers are in line with Sheese et al. (2008), who reported that participants made 4% of possible correct predictions at 6–7 months of age.

Finally, on the probabilities task at 10 months, the mean proportional looking time toward the target was 0.60, $SD = 0.20$, min = 0.21, and max = 0.98, which was significantly different from chance, $t(68) = 4.1$ and $p < 0.01$. At 18 months of age, the mean proportional looking time was 0.51, $SD = 0.18$, min = 0.06, and max = 0.83, which was not significantly different from chance, $t(72) = 0.26$ and $p = 0.80$.

Zero-order correlation. For the associations between predictors, an initial analysis showed a longitudinal relation between individual differences in the visual sequence task at 10 and 18 months, $r = 0.44$ and $p < 0.001$. Furthermore, individual differences in the associative learning task at 6 months correlated with individual differences in the probabilities task at 18 months, $r = -0.30$ and $p = 0.01$. We also observed a nonsignificant correlation between individual differences in the probabilities task at 10 months and individual differences in the visual sequence at 10 months, $r = -0.26$ and $p = 0.07$.

For the outcome variables, scores on the two subscales of the CDI were significantly correlated, $r = 0.56$ and $p < 0.001$. For relations between predictors and outcome variables, we noted significant correlations between scores on both CDI subscales and individual differences in the visual sequence task at 18 months (receptive vocabulary: $r = 0.22$ and $p = 0.04$; expressive vocabulary: $r = 0.30$ and $p < 0.001$), respectively. There were no significant longitudinal relations within the associative learning task or within the probabilities task. All zero-order (Pearson) correlations and descriptive statistics for all predictors and outcomes are summarized in Table 2.

3.1 | Path analysis

The pattern of zero-order correlations in Table 2 indicated a possible predictive relationship between individual differences in the visual sequence task and scores on the CDI subscales. To investigate this relationship in a more parsimonious way, we conducted a path analysis (see Figure 2 and statistical analysis for the structure of the model).

The analysis revealed a model with a good fit, χ^2 (4) = 3.742, p = 0.442; CFI = 1.0; TLI = 1.04; RMSEA <0.01; and SRMSR = 0.058, allowing for the investigation of individual paths within it. As illustrated in Figure 2, our model had six significant paths. Starting with the visual sequence task, individual differences at 10 months predicted both individual differences in the visual sequence at 18 months, $\beta = 0.30$ and $p = 0.017$, and scores on the receptive vocabulary subscale, $\beta = -0.28$ and $p = 0.044$. This indicates that the more correct predictions participants made at 10 months of age, the more correct predictions they made at 18 months of age—implying that we measured a somewhat stable individual ability. At the same time, the more correct predictions participants made at 10 months, the fewer words they understood at 18 months of age. Furthermore, individual differences in the visual sequence task at 18 months predicted scores on both the receptive, $\beta = 0.35$ and $p = 0.003$, and expressive vocabulary subscales, $\beta = 0.42$ and $p < 0.001$, indicating that the more correct predictions infants made in the visual sequence task at 18 months, the greater vocabulary they had.

Individual differences in the associative learning task at 6 months predicted individual differences in the probabilities task at 18 months, $\beta = -0.36$ and $p = 0.008$. The more participants looked at

TABLE 2 Zero-order correlations between predictor variables and vocabulary size (expressive and receptive) together with means, standard deviations, skewness, and kurtosis for all variables

Variable	1	2	3	4	5	6	7	8
1. AL – 6 months	-							
2. AL – 6 months	−0.08	-						
3. VS – 10 months	−0.18	−0.04	-					
4. VS – 18 months	−0.11	−0.07	0.44***	-				
5. Prob. – 10 months	0.06	−0.14	−0.26*	−0.16	-			
6. Prob. – 18 months	−0.30***	0.03	−0.01	−0.06	0.21	-		
7. CDI – Rec.	0.06	0.06	−0.05	0.22**	−0.07	0.07	-	
8. CDI – Exp.	−0.08	0.15	−0.07	0.30****	−0.14	0.03	0.56****	-
Mean	0.509	0.492	4.20	5.57	0.595	0.505	59.9	23.4
SD	0.142	0.136	4.45	4.57	0.195	0.176	16.0	20.8
Skewness	0.117	0.015	2.34	1.20	−0.043	−0.416	−0.779	1.01
Kurtosis	0.557	−0.048	6.41	1.06	−0.757	−0.335	0.404	0.141

Note: n in the table varies between 67 and 106.

Abbreviations: AL, associative learning; CDI–Exp., expressive vocabulary; CDI–Rec., receptive vocabulary; Prob., probabilities: VS, visual sequence.

*p = 0.07, **p = 0.04, ***p = 0.01, ****p < 0.001.

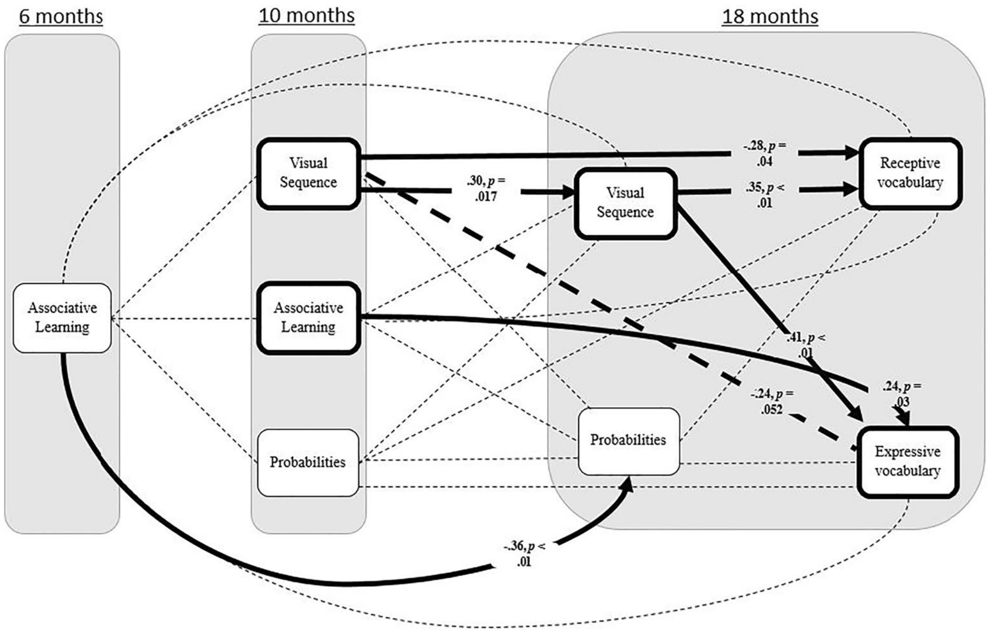


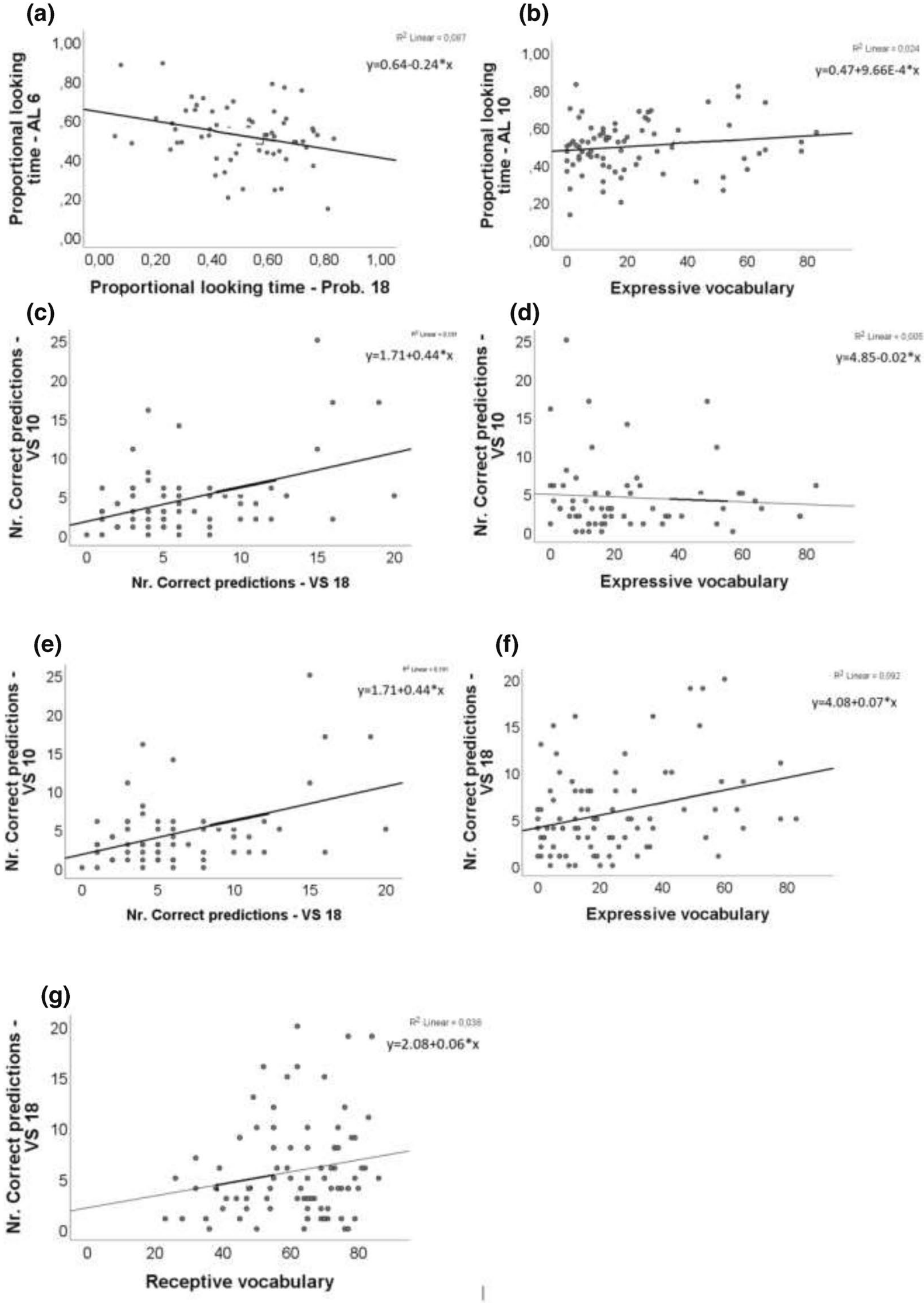
FIGURE 2 Illustration of the path model assessing early predictors of vocabulary size. Arrows denote significant paths ($p < 0.05$), and numbers on arrows are standardized path coefficients with corresponding p values. Dotted bold line denotes a marginally significant path ($p < 0.10$). Dotted lines denote nonsignificant paths ($p \geq 0.12$). See Table S1 for path coefficients and p values for all paths. Note that the model includes forward associations from individual differences at 6 months to 10 and 18 months, from individual differences at 10–18 months, and concurrent associations at 18 months

the critical frame in the associative learning task, the less they looked at the unlikely outcome in the probabilities task. Additionally, individual differences in the associative learning task at 10 months predicted scores on the expressive vocabulary subscale, $\beta = 0.24$ and $p = 0.037$, indicating that the more participants looked at the critical frame, the more words they could express. This implies that the ability to detect cross modal associations is related to later vocabulary. No other path coefficient was significant. See Figure 3 for scatterplots on each significant, or near significant, paths.

In sum, the model suggests that individual differences in the associative learning task at 10 months and individual differences in the visual sequence task at both 10 and 18 months predicts later vocabulary (albeit differently), while individual differences in the associative learning task at 6 months and the probabilities task did not. The model could explain 17% of the variance in the receptive vocabulary subscale and 21% of the variance in the expressive vocabulary subscale.

4 | DISCUSSION

Infants face a range of challenges in their day-to-day lives that they need to overcome to develop an extensive vocabulary. In the current study, we investigated the impact of three such challenges: sequence learning, associative learning, and probability processing. Consistent with previous findings (Kidd & Arciuli, 2016; Reuter et al., 2018; Shafto et al., 2012; von Koss Torkildsen et al., 2019), our results suggest that children's ability to solve the challenges of associative learning and sequence learning in infancy plays a part in vocabulary size in toddlerhood. This further highlight that learning


(a)
y=0.64-0.24*x
Proportional looking time - AL 6
Proportional looking time - Prob. 18
(b)
y=0.47+9.66E-4*x
Proportional looking time - AL 10
Expressive vocabulary
(c)
y=1.71+0.44*x
Nr. Correct predictions - VS 10
Nr. Correct predictions - VS 18
(d)
y=4.85-0.02*x
Nr. Correct predictions - VS 10
Expressive vocabulary
(e)
y=1.71+0.44*x
Nr. Correct predictions - VS 10
Nr. Correct predictions - VS 18
(f)
y=4.08+0.07*x
Nr. Correct predictions - VS 18
Expressive vocabulary
(g)
y=2.08+0.06*x
Nr. Correct predictions - VS 18
Receptive vocabulary

a language is not merely the result of a narrow statistical learning mechanism (Saffran et al., 1996). Instead, it depends on multiple mechanisms that develop during the first 2 years of life. Specifically, the results suggest that a broader set of abilities used to detect and learn from regularities, beyond transitional probabilities in the linguistic domain, is related to language (Frost et al., 2019).

Our model could explain 17% and 21% of the variance in the CDI subscales, respectively. Given the complexity of vocabulary development and the variability in family contexts (Hindman et al., 2016), these values are large. We are not, however, claiming that associative learning and sequence learning are the only important challenges for vocabulary development or suggesting a strong causal pathway. Indeed, the current model leaves a large portion of the variance unexplained and other factors not present in the model could potentially increase the variance explained, account for a part of one or more of the observed paths, or function as a mediating factor for one or more of the observed paths. For example, previous research suggests that pedagogical context, social interaction, and socioeconomic status, to name a few, are all important contributors to language development (Duff et al., 2015; Hindman et al., 2016; Karmiloff & Karmiloff-Smith, 2009; Kidd & Donnelly, 2020; Kuhl et al., 2003). Further, it is likely that specific linguistic challenges (in addition to the nonlinguistic challenges assessed here), such as attending to transitional probabilities to perform word segmentation, have a large contribution to later vocabulary. Thus, extending the current model with variables that measure an infant's ability to perceive and encode social information, the quality of their home environment, and their ability to process linguistic information should provide a more integrative model of vocabulary development. However, although we are not in a position to make strong claims about the mechanism(s) underlying the relation between associative learning and sequence learning and vocabulary development, we argue that the model is a promising and noteworthy step, demonstrating that broad nonlinguistic abilities are related to vocabulary growth. In turn, this might promote further investigations into how infants might detect and learn from regularities in everyday life, outside the lab (Frost et al., 2019).

Recent research has indicated that young children are intuitive statisticians (Xu & Garcia, 2008) that might use a general ability to evaluate and integrate probabilities to constrain hypothesis during word learning (Frank, 2011; Frank & Tenenbaum, 2011; Ullman et al., 2012). In the current study, individual differences in the probabilities task could not predict vocabulary size. Thus, our findings do not support the idea of a more general statistical ability playing a central role in word learning. However, this does not call into question the relation between learning from regularities and vocabulary size, but rather suggests that a general sensitivity to probabilities is not related to vocabulary size. Of course, a general sensitivity to probabilities might play a role in language development that does not become evident in a relation between our probabilities task and vocabulary size. For example, it might be important for generating hypothesis about the surrounding environment rather than words per se. Further investigating the role of such general statistical abilities will be an interesting venue for future research.

FIGURE 3 Scatterplots for all significant, or near significant, paths based on available raw data. VS—Visual sequence, AL—Associative learning, Prob—Probabilities. (a) Individual differences in the associative learning task at 6 months and individual differences in the probabilities task at 18 months. (b) Individual differences in the associative learning task at 10 months and CDI expressive vocabulary. (c) Individual differences in the visual sequence task at 10 months, and individual differences in the visual sequence task at 18 months. (d) Individual differences in the visual sequence task at 10 months and CDI receptive vocabulary. (e) Individual differences in the visual sequence task at 10 months and CDI expressive vocabulary. (f) Individual differences in the visual sequence task at 18 months and CDI expressive vocabulary. (g) Individual differences in the visual sequence task at 18 months and CDI receptive vocabulary. Note. Nr. Correct predictions = number of correct predictions. Proportional looking time AL = proportional looking time toward the critical frame. Proportional looking time Prob. = proportional looking time toward the unlikely outcome

Some methodological choices made here needs to be discussed. In this study, we refrained from making predictions about the direction of effects in the model, did not assess group-level performance, and did not test against chance performance on a group level as a main analysis. Instead, we assessed if associations exist between individual differences and related constructs (see Starr et al., 2013, for a similar approach). Indeed, there are strong theoretical reasons to believe that the tasks used should be related to later vocabulary, as outlined in the introduction. The aim of this study was to operationalize statistical learning as a broad, multi-component ability and investigate how that relates to vocabulary—rather than to investigate the group performance of each specific task. This is closely related to the fact that in general, interpreting individual differences in infants using eye tracking is challenging. The response of a child will undoubtedly reflect a combination of ability (e.g., group-level performance), attention strategy (e.g., novelty/familiarity preference), and exploration (focusing on where events are expected to unfold), to name a few. Thus, infants' responses, and individual differences, may not always fall on a continuum expected from the group-level analysis. Indeed, our results suggest that at a group-level, the participants did not seem to learn associations, nor did they seem to differentiate between a likely and unlikely outcome (except for at 10 months of age), nor did all participants make correct predictions in the visual sequence. At the same time, individual performance related to later vocabulary. As such, it seems that there is a meaningful difference between the participants who could solve the challenge of associative learning and sequence learning, compared to those who could not. While it can be difficult to interpret how and why individual differences occur, that is not to say that the differences are meaningless.

Indeed, individual differences in both the associative learning task and the sequence learning task was associated with later vocabulary, explaining up to 20% of the CDI measurement. First, this would suggest that the efficiency with which infants can solve the challenge of detecting cross-modal, co-occurring stimuli is related to later vocabulary. It has been proposed that associative learning helps infants constrain possible meanings of words, by detecting that a certain object is denoted by a certain, specific word—regardless of context (Kidd & Donnelly, 2020; Saxton, 2010). It is less clear what an efficient strategy might be, however. To name a few examples, it could be that detecting an association from few pairings allows for faster vocabulary learning or it could be that observing multiple pairings before making an association allows for a more correct, robust hypothesis on the meaning of a word, allowing vocabulary to grow slowly but surely. It could also be that the most efficient strategy changes over time, such that infants need to observe co-occurring, cross-modal stimuli many more times in order to learn it, compared to children at an older age.

Second, the results indicate that infants' efficiency in detecting and processing patterns in sequential information is related to later vocabulary. It has been proposed that processing patterns aid in detecting other statistical structures within a given language, which in turn can be used to perform word segmentation (among others; Saffran, 2009). Again, exactly what an efficient strategy might look like is less clear. For example, it could be that a reactive strategy (i.e., attending to events as they occur) might allow infants to observe patterns unfold in many different contexts, providing a rich environment for learning complex events. On the other hand, a predictive strategy (i.e., predicting where events will occur) might allow infants to test hypotheses about the surrounding environment, providing a positive or negative feedback loop (Gredebäck et al., 2018; Marciszko et al., 2020; Reuter et al., 2018). In turn, this feedback could aid in constraining a working hypothesis. It is also possible that what the most efficient strategy might be changes with age. For example, research demonstrate that the ocular-motor system is relatively slow at 10 months of age (Gredebäck et al., 2018), which makes incorrect predictions particularly costly as corrective saccades also take a long time to execute, ensuring that infants miss a larger sequence of events than older infants or adults (Gredebäck et al., 2018). Indeed, our results suggest that a predictive strategy at 10 months of age is later related to a smaller vocabulary, whereas a predictive strategy at 18 months of age is related to a larger vocabulary—suggesting age-dependent differences in strategy efficiency. However, future research is needed to investigate the developmental trajectory of both associative learning and sequence learning further.

In addition, we did not find any longitudinal stability within the associative learning task nor the probabilities task. At the same time, we observed a longitudinal relationship between individual differences in the associative learning task at 6 months and individual differences in the probabilities task at 18 months. At this point, it is difficult to assess the reason for this relation. Although previous research has demonstrated group-level effects for these tasks during infancy, individual stability, as we can measure it, might appear somewhat later. Important to note here is also that Richardson and Kirkham (2004) conducted the associative learning task on only 10 participants. As such, it might be that only a few 6- to 10-month-old infants can learn an association with this paradigm—which could serve as a partial explanation as to why we cannot observe a stable group or individual performance. Future research is needed to explore this venture further. In a similar vein, it is important to note that Kayhan et al. (2018) demonstrated that their participants (aged 6-, 12, and 18 months) changed their looking behavior as a function of likelihood differences. The larger the difference in relative likelihood between the likely and unlikely outcome, the more participants looked toward the unlikely outcome. As such, testing group performance against chance (i.e., 0.5) is not necessarily the most informative way to investigate performance within this task. At the same time, we kept the relative likelihood constant over trials in this study, which means that we cannot do the same analyses as the original article (Kayhan et al., 2018). Additionally, the probabilities task was presented at the end of the experiment procedure (see design). It is thus possible that participants were tired by the time they conducted this task, and that we, consequently, lost meaningful variance here. While we cannot exclude this possibility entirely, we note that the performance at 10 months are similar to performance reported in previous research (Kayhan et al., 2018), which lends some support to the notion that we have captured individual statistical abilities with the task. However, the performance at 18 months of age were generally lower than at 10 months, suggesting that we could have an issue of fatigue at this age point. Future research is needed to investigate this potential issue further.

We also need to acknowledge the possibility that simple frequency learning could underlie performance in the visual sequence task. That is, simpler events could be easier to predict simply because they involved a gaze shift to a location where stimuli frequently occurred (50% of all stimuli appeared in location A) than more complex events (25% of stimuli appeared in B and C, respectively). However, we do not find this alternative explanation likely because it would imply that simply focusing on the frequency of occurrences (e.g., how frequent is a particular syllable) without attention to the pattern of occurrences (how likely is it that one syllable appears after another) would be related to vocabulary size. Instead, we argue that the visual sequence task assesses the ability to detect patterns of events in general and suggest tentatively that this ability is involved in vocabulary development.

It should also be noted that the number of correct predictions is quite low in the visual sequence task, with an average of 7.5% and 9.3% made of possible correct predictions, at 10 and 18 months, respectively. However, similar response pattern has been reported in past research, using a highly similar task—where participants only performed 4% of possible correct predictions (Sheese et al., 2008). In this task, infants seldom predict the reappearance of the next stimuli, but the predictions they do make seem important and likely captures meaningful variance (Sheese et al., 2008).

There are also general limitations to both our eye-tracking and language measures, which may constrain the scope of our conclusions. Although typical for infant research, our eye-tracking tasks contain few trials per participant, which makes it difficult to estimate their reliability. We acknowledge that this is far from optimal, but note that it primarily puts an upper limit on the size of the relation that we can detect. A further limitation, illustrated in Table 1, is the high proportion of missing data on some of the tasks for some ages. In particular, just over half of the infants contributed with data on some of the tasks at the 10-month visit. Although we chose to handle our missing data with FIML estimation, which can reliably handle a relatively large amount of missing data (Enders, 2001), it would have been better to have a complete data matrix. At the same time, data loss might be an unavoidable side effect of conducting a longitudinal study with infants. Prior studies have argued that stable individual differences are hard to find in infancy and childhood, particularly concerning statistical learning (Arnon, 2020; Christiansen, 2019). The fact that we found some stability across infancy is promising and noteworthy. Even so, an important direction for future studies will be to further evaluate both the reliability and validity of these tasks. Further, we did not counterbalance the order of the tasks—thus, all participants saw the same tasks in the same order at each lab visit. This might have introduced the issue of fatigue. Indeed, this could serve as an (at least partial) explanation of why we did not observe any stable group performances. At the same time, all participants should have been equally affected by this potential order effect. As such, fatigue cannot explain the observed associations and stable individual differences within our model. Furthermore, our language measure is based on the parental report rather than direct observations. Although this approach is consistent with much previous research on language development and statistical learning (Reuter et al., 2018; Shafto et al., 2012; Singh et al., 2012), we acknowledge that it might give rise to a biased language measure.

Finally, we recognize that the eye-tracking tasks used also tap into general cognitive mechanisms, not controlled for directly here. Indeed, several factors are common to all eye-tracking tasks, such that they all rely on selective attention, oculomotor control, and visual short-term memory. These factors are surely involved in the process of detecting information in the first place. However, this does not diminish the finding that the current tasks are related to vocabulary growth. Instead, it suggests that sequence learning and associative learning contribute to vocabulary size, beyond common eye-tracking factors (otherwise, there should be no unique contribution of any of the tasks). As we see it, the most parsimonious explanation is that the paths illustrate different challenges that infants face when learning a language and expanding their vocabulary.

5 | CONCLUSION

In the current study, we set out to move beyond a narrow operationalization of statistical learning often used in previous research toward the idea of a broad, multi-component ability. We demonstrate that individual differences in associative learning and sequence learning in infancy predict vocabulary size in toddlers. Combined with previous research (Evans et al., 2009; Hedenius et al., 2011; Reuter et al., 2018; Romberg & Saffran, 2010; Saffran et al., 2009; Shafto et al., 2012; von Koss Torkildsen et al., 2019), compelling evidence now suggests that statistical learning abilities, broadly defined, have a role in vocabulary development. One potential extension of this work is to assess the relation between low performance on statistical tasks in early infancy and later language difficulties. The earlier such difficulties can be identified, the greater the possibility of providing support for children in need.

ACKNOWLEDGEMENTS

The project was funded by grants to GG (Wallenberg Fellowship: KAW 2012.0120) and ML (Riksbankens Jubileumsfond: P15-0430:1), and MH was funded by Royal Swedish Academy of Sciences (Lennart "Aktiestinsen" Israelssons stipendiefond), the Promobilia Foundation, the Sunnerdahl Disability Foundation, and the Sven Jerring Foundation. We thank members of the BASIC team (Mari Fransson, Carin Marciszko, and Ben Kenward) for valuable discussions and input in the different phases of this project. We thank also Maria Rogell, Irma Leijlic, Claire Almer, and Anna Sewall for help with the data collection, the members of the Uppsala Child and Baby lab for helpful comments on the manuscript, and the participating families. The authors declare no conflicts of interest with regard to the funding source for this study.

ORCID

Anton Gerbrand https://orcid.org/0000-0002-9696-1919
Linda Forsman https://orcid.org/0000-0003-0208-4212
Marcus Lindskog https://orcid.org/0000-0003-1326-6177

REFERENCES

Arciuli, J., & Conway, C. M. (2018). The promise—and challenge—of statistical learning for elucidating atypical language development. *Current Directions in Psychological Science*, *27*(6), 492–500.

Arnon, I. (2020). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behavior Research Methods*, *52*(1), 68–81. https://doi.org/10.3758/s13428-019-01205-5

Astor, K., Lindskog, M., Forssman, L., Kenward, B., Fransson, M., Skalkidou, A., Tharner, A., Cassé, J., & Gredebäck, G. (2020). Social and emotional contexts predict the development of gaze following in early infancy. *Royal Society Open Science*, *7*, 201178. https://doi.org/10.1098/rsos.201178

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literature. *Topics in Cognitive Science*, *11*(3), 468–481. https://doi.org/10.1111/tops.12332

Cristia, A., Seidl, A., Singh, L., & Houston, D. (2016). Test–retest reliability in infant speech perception tasks. *Infancy*, *21*(5), 648–667.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 125–127. https://doi.org/10.3758/BF03203646

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry*, *56*(8), 848–856. https://doi.org/10.1111/jcpp.12378

Enders, C. K. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement*, *61*(5), 713–740. https://doi.org/10.1177/0013164401615001

Eriksson, M., Westerlund, M., & Berglund, E. (2002). A screening version of the Swedish communicative development inventories designed for use with 18-month-old children. *Journal of Speech, Language, and Hearing Research*, *45*(5), 948–960. https://doi.org/10.1044/1092-4388(2002/077

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Hearing Research*, *52*(2), 321–335. https://doi.org/10.1515/COGL.2011.010.Input

Frank, M. C. (2011). Computational models of early language acquisition. *Current Opinion in Neurobiology*, *21*(3), 381–386.

Frank, M. C., & Tenenbaum, J. B. (2011). Three ideal observer models for rule learning in simple languages. *Cognition*, *120*(3), 360–371.

Frost, R., Armstrong, B., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, *145*(12), 1128–1153. https://doi.org/10.1037/bul0000210

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010

Gredebäck, G., Johnson, S., & Hofstenvon, C. (2009). Eye tracking in infancy research. *Developmental Neuropsychology*, *35*(1), 1–19. https://doi.org/10.1080/87565640903325758

Gredebäck, G., Lindskog, M., Juvrud, J. C., Green, D., & Marciszko, C. (2018). Action prediction allows hypothesis testing via internal forward models at 6 Months of age. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00290

Hedenius, M., Lum, J., & Bölte, S. (2021). Alterations of procedural memory consolidation in children with developmental dyslexia. *Neuropsychology*, *35*(2), 185–196. https://doi.org/10.1037/neu0000708

Hedenius, M., Persson, J., Tremblay, A., Adi-Japha, E., Veríssimo, J., Dye, C. D., Alm, P., Jennische, M., Bruce Tomblin, J., & Ullman, M. T. (2011). Grammar predicts procedural learning and consolidation deficits in children with Specific Language Impairment. *Research in Developmental Disabilities*, *32*(6), 2362–2375. https://doi-org.ezproxy.its.uu.se/10.1016/j.ridd.2011.07.026

Hindman, A. H., Wasik, B. A., & Snell, E. K. (2016). Closing the 30 million word gap: Next steps in designing research to inform practice. *Closing the 30 Million Word Gap: Next Steps in Designing Research to Inform Practice*, *10*(2), 134–139. https://doi.org/10.1111/cdep.12177

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Karmiloff, K., & Karmiloff-Smith, A. (2009). *Pathways to language: From fetus to adolescent*. Harvard University Press.

Kayhan, E., Gredebäck, G., & Lindskog, M. (2018). Infants distinguish between two events based on their relative likelihood. *Child Development*, *89*(6), e507–e519.

Kidd, E., & Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development*, *87*(1), 184–193. https://doi.org/10.1111/cdev.12461

Kidd, E. & Donnelly, S. (2020). Individual differences in language acquisition. *Annual review of linguistics*, *6*, 319, 340. https://doi.org/10.1146/annurev-linguistics-011619-030326

Kuhl, P. K., Tsao, F., & Liu, H. (2003). *Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning*.

Lany, J., & Saffran, J. R. (2013). Statistical learning mechanisms in infancy. In J. Rubenstein, & P. Rakic (Eds.), *Neural Circuit Development and Function in the Healthy and Diseased Brain: Comprehensive Developmental Neuroscience*. Academic Press.

Maccallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149.

Marciszko, C., Forssman, L., Kenward, B., Lindskog, M., Fransson, M., & Gredebäck, G. (2020). The social foundation of executive function. *Developmental Science*, *23*(3), e12924.

Maye, J., Werker, J. F., & Gerken, L. (2002). *Infant sensitivity to distributional information can affect phonetic discrimination*, *82*, 101–111.

Nyström, P., Falck-Ytter, T., & Gredebäck, G. (2016). The TimeStudio Project: An open source scientific workflow system for the behavioral and brain sciences. *Behavior Research Methods*, *48*(2), 542–552. https://doi.org/10.3758/s13428-015-0616-x

Reuter, T., Emberson, L., Romberg, A., & Lew-Williams, C. (2018). Individual differences in nonverbal prediction and vocabulary size in infancy. *Cognition*, *176*, 215–219. https://doi.org/10.1016/j.cognition.2018.03.006

Richardson, D. C., & Kirkham, N. Z. (2004). Multimodal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General*, *133*(1), 46–62. https://doi.org/10.1037/0096-3445.133.1.46

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906–914. https://doi.org/10.1002/wcs.78

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Saffran, J. R. (2009). What is statistical learning, and what statistical learning is not. In S. P. Johnson (Ed.), *Neoconstructivism: The New Science of Cognitive Development*.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Pelucchi, B., & Hay, J. F. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–685.

Saxton, M. (2010). *Child language: Acquisition and development*. Sage.

Shafto, C. L., Conway, C. M., Field, S. L., & Houston, D. M. (2012). Visual sequence learning in infancy: Domain-general and domain-specific associations with language. *Infancy*, *17*(3), 247–271. https://doi.org/10.1111/j.1532-7078.2011.00085.x

Sheese, B. E., Rothbart, M. K., Posner, M. I., White, L. K., & Fraundorf, S. H. (2008). Executive attention and self-regulation in infancy. *Infant Behavior and Development*, *31*, 501–510. https://doi.org/10.1016/j.infbeh.2008.02.001

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods*, *49*, 418–432. https://doi.org/10.3758/s13428-016-0719-z

Singh, L., Reznick, J. S., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: A longitudinal analysis. *Developmental Science*, *4*, 482–495. https://doi.org/10.1111/j.1467-7687.2012.01141.x

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, *22*(4), 325–336. https://doi.org/10.1016/j.tics.2018.02.004

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences*, *110*(45), 18116–18120.

Statistiska Centralbyrån. (2019). *Utbildningsnivån i svergie*.

Ullman, T. D., Goodman, N. D., & Tenenbaum, J. B. (2012). Theory learning as stochastic search in the language of thought. *Cognitive Development*, *27*, 455–480. https://doi.org/10.1016/j.cogdev.2012.07.005

vonKoss Torkildsen, J., Arciuli, J., & Bø Wie, O. (2019). Individual differences in statistical learning predict children's reading ability in a semi-transparent orthography. *Learning and Individual Differences*, *69*, 60–68. https://doi.org/10.1016/j.lindif.2018.11.003

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, *105*(13), 5012–5015.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Gerbrand, A., Gredebäck, G., Hedenius, M., Forsman, L., & Lindskog, M. (2022). Statistical learning in infancy predicts vocabulary size in toddlerhood. *Infancy*, 1–20. https://doi.org/10.1111/infa.12471