Why word learning is not fast

Natalie Munro[1], Elise Baker[1], Karla McGregor[1,2]*, Kimberly Docking[1] and Joanne Arciuli[1]

[1] *Discipline of Speech Pathology, University of Sydney, Lidcombe, NSW, Australia*
[2] *Department of Communication Sciences and Disorders, Delta Center, University of Iowa, Iowa City, IA, USA*

Upon fast mapping, children rarely retain new words even over intervals as short as 5 min. In this study, we asked whether the memory process of encoding or consolidation is the bottleneck to retention. Forty-nine children, mean age 33 months, were exposed to eight 2- or-3-syllable nonce neighbors of words in their existing lexicons. Didactic training consisted of six exposures to each word in the context of its referent, an unfamiliar toy. Productions were elicited four times: immediately following the examiner's model, and at 1-min-, 5-min-, and multiday retention intervals. At the final two intervals, the examiner said the first syllable and provided a beat gesture highlighting target word length in syllables as a cue following any erred production. The children were highly accurate at immediate posttest. Accuracy fell sharply over the 1-min retention interval and again after an additional 5 min. Performance then stabilized such that the 5-min and multiday posttests yielded comparable performance. Given this time course, we conclude that it was not the post-encoding process of consolidation but the process of encoding itself that presented the primary bottleneck to retention. Patterns of errors and responses to cueing upon error suggested that word forms were particularly vulnerable to partial decay during the time course of encoding.

Keywords: fast mapping, word learning, memory, encoding, consolidation, retention

INTRODUCTION

Word learning is an extended process that requires multiple exposures to word forms in meaningful contexts. Via those exposures word forms, word meanings, their receptive linkage (form-to-meaning), and their expressive linkage (meaning-to-form) come to be represented in the lexicon (Gupta, 2005). The initial step in the word learning process, fast mapping, is the focus of this paper.

Fast mapping can be thought of as the brain's initial response to a new word. Much of the literature on fast mapping focuses on the amazing facility of children, even those who are only in their second year of life, to glean relevant information from this first encounter. The term "fast" refers to the fact that children tend to notice a new word and identify its referent immediately. Nevertheless, the child's fast mapped representation tends to be incomplete or inaccurate and highly prone to obliviscence (Horst and Samuelson, 2008). Although fast mapping is fast, word learning is not.

In this paper, we examine the memory processes that limit retention after fast mapping. Specifically, we ask whether encoding or consolidation is the bottleneck to retention. In our two-stage model, e*ncoding* refers to the establishment of a representation in long term memory. Following encoding, that representation can become strengthened via *consolidation*. In the best case scenario, via flawless encoding and consolidation, a new word is stored in the long term lexicon – fast mapping is successful. This does not necessarily happen, however. The memory trace to be encoded or consolidated can decay partially or fully. To examine whether fast mapped words are subject to decay more so during encoding or during consolidation, we trained children on eight new words and their novel referents during a play interaction. Although we

administered two recognition probes to tap retention, our primary concern in this paper is the response to production probes because the mapping of new word forms and their expressive links is thought to be particularly fragile. In numerous studies, it has been demonstrated that children are better at recognizing word–referent pairings in alternative forced choice tasks than they are at producing those new words in response to their referents (Dollaghan, 1985; Gray, 2003, 2004; Gupta, 2005; Booth et al., 2008; Horst and Samuelson, 2008). We also examined the nature of children's production errors for additional insight into the particulars of the retention problem.

ENCODING

Encoding involves establishing a memory trace following exposure to new information, in this case new words and their referents. We tapped encoding via elicited production 1- and 5-min after the final training exposure. To accurately name a referent at the 1- and 5-min retention intervals, a child has to have encoded the word form with sufficient acoustic-phonetic information to support the development of an articulatory-phonetic representation for production (Rvachew and Brosseau-Lapré, 2010), the meaning to a degree sufficient for recognition of the referent, and an expressive link between the two such that recognition of the referent triggers activation of the word form. These encodings have to be robust enough to resist decay over a matter of minutes. In connectionist terms, slow weights between lexical and sublexical levels represent word forms; slow weights between semantic and conceptual levels represent word meaning; and slow weights between lexical and semantic levels represent receptive and expressive links (Martin and Gupta, 2004). These weights must strengthen and

reach saturation if the memory is to be retained. Given robust enough lexical-to-sublexical weights, the children can produce the word. Given robust enough semantic–lexical connection weights, the child can retrieve the new word when the semantic referent is presented.

CONSOLIDATION

Unlike encoding, consolidation is not driven by exposure. That is, memories are consolidated in the absence of new experiences with, or active rehearsal of, the word or referent. Instead, the passage of time and, in some cases, the occurrence of sleep, drives consolidation (McClelland et al., 1995; Walker, 2005). During consolidation, the fragile newly encoded memory is stabilized, enhanced, and integrated with other memories. With stabilization, the memory becomes less prone to interference and forgetting. The behavioral manifestation of consolidation is the maintenance of performance levels measured immediately following encoding. For example, among adults, gains in comprehension of synthesized speech fell over a 12-h period without sleep but returned to immediate post-training levels after sleep (Fenn et al., 2003). Enhancement is manifested as improved performance relative to encoding levels. It has often been reported that children who were taught novel words and referents recognized them better days or weeks later than immediately after training (Rice et al., 1994; Storkel, 2001; Booth, 2009; McGregor et al., 2009; Norbury et al., 2010). In studies of word learning among adults, the behavioral signature of integration is the emergence of competition and priming effects between the new information and old related information (Dumay and Gaskell, 2007). Again to illustrate with a connectionist metaphor, during consolidation the network relaxes into an attractor state (an existing memory) and the weights between nodes belonging to that memory pattern are updated. At the physiological level, there is evidence that memories are replayed during sleep. For example, among adults, sleep spindle density recorded after an associative word learning task correlated with the number of newly learned word pairs correctly recalled after sleep (Gais et al., 2002).

CURRENT STUDY

In the current study, we set out to explore the basis for memory limitations associated with fast mapping. To do so, we asked children to produce names for eight objects that the examiner had previously presented and labeled with two- and three-syllable nonce neighbors of words in the children's existing lexicons. The labels were elicited four times, immediately following the examiner's model, and 1-min, 5-min, and multiple days later. Expecting declining performance over time, we also cued the children after failed productions by providing the first syllable and a beat gesture indicating the word length in the target form at the 5-min and multiday retention intervals to better determine partial representations[1].

---

[1] There were four training conditions that varied by gestural support such that hand gestures that conveyed semantic information about the trained referents (i.e., shape) or phonological information about the trained words (i.e., word length) were or were not presented. Although the original intent was to determine whether gestural support during training influenced encoding and consolidation as measured

Whereas we took correct naming responses as indicative of full representations of word form, we mined all responses for further information. Indicators of partial form representations were approximations of the target word and substitutions of familiar lexical neighbors before or after cueing. Responses that provided no evidence of a form representation involved substitution of a training neighbor (i.e., a trained word associated with a different referent than the one the child was asked to name), substitution of a semantic neighbor, no response or don't know responses, and non-compliant behavior. In the case of cued naming, exact repetitions of the cue were also considered to provide no evidence of form representation.

Given that the children heard the target word five times during training and a sixth time immediately before they repeated the word, our immediate repetition task likely tapped both previous encoding and the child's phonological short term memory for the examiner's model. We therefore did not use the children's immediate repetitions to inform our conclusions about encoding *per se*; however, their repetitions were useful baseline data in that they allowed us to determine whether the children had perceived the word and could articulate it.

HYPOTHESIS AND PREDICTIONS
Encoding
If encoding is not sufficient to resist decay, performance should decline between immediate repetition and the 1-min interval and might continue to decline over the 5-min interval. Encoding of either the expressive link to the word form, or the word form itself, could be problematic. If the former, we would expect numerous training neighbor substitutions. Also, production should improve following cues that convey information about the word form because that information specifies the expressive link. If the latter, the decay of the word form could be complete, in which case we would expect errors that evince no knowledge of the target form, or partial, in which case we would obtain many approximations or lexical neighbor substitutions.

Consolidation
If, alternatively, consolidation of newly encoded words is problematic, then immediate repetition would be intact and performance would be relatively steady from immediate repetition to retention at the 1- and 5-min intervals but performance should decline when measured at the multiday retention interval. The multiday interval ranged from 1 to 7 days across children, thereby allowing time and sleep in support of consolidation. If performance declines, we could examine response to cueing and error types to determine whether consolidation of word forms, expressive links, or both were vulnerable.

MATERIALS AND METHODS
PARTICIPANTS
Ethical approval was obtained from the University of Sydney's Human Research Ethics Committee (Approval number:

---

by production, it did not, and the effect of gesture during training will not be considered further. We will, however, discuss the effect of gestural cues provided during the test probes.

11941). Forty-nine children (aged between 29 and 36 months, mean = 32.65, SD = 2.28; 19 boys) were recruited via advertisements in community-based parenting newspapers. Inclusionary criteria included: unremarkable birth, medical and developmental history, and normal hearing as determined by parent report as well as normal receptive vocabulary based on >16th percentile Peabody Picture Vocabulary Test – Fourth edition (PPVT-IV: Dunn and Dunn, 2007; mean percentile = 84.67, SD = 13.81), and typical speech production skills, based on >16th percentile on the Goldman Fristoe Test of Articulation – Second edition (GFTA-2: Goldman and Fristoe, 2000; mean percentile = 76.48, SD = 13.89). Although not used as an inclusionary criterion, data were also gathered on the participants' phonological short term memory abilities as measured by the test of early non-word repetition (TENR; Stokes and Klee, 2009; mean percent phonemes correct = 81%, SD = 8).

Assessments to determine eligibility for the study were conducted during the first of two data collection sessions. The TENR was also administered during the first session. All 49 children participated in the first session; 48 returned for the second.

WORDS AND REFERENTS

Eight 2- or 3-syllable nonce words were used in the word learning protocol. These were based on polysyllabic real words found within the lexicon of Australian English speaking toddlers (MARCS Auditory Laboratories, 2004) and contained early-developing phonemes and syllable shapes. To create the nonce words, one consonant in each two syllable word and two consonants in each three syllable word were altered (e.g., *vegemite–bekemite*). Thereby, the nonce words constituted lexical neighbors for familiar English words. This decision was purposeful because the nonce words that occur repeatedly across fast mapping studies (e.g., *koob, dax, blicket*) have neighbors (e.g., *cool, ducks, blanket*) and, more importantly, because words that toddlers learn everyday have neighbors as well (e.g., *ball–fall, book–look, nose–no, bubble–buckle*).

The words were arbitrarily paired with eight real but unusual toys representing one of two play themes: sand or music toys. The order of presentation of the themes was counterbalanced across participants.

Although our focus was on production, we also wished to confirm that the children accomplished fast mapping as measured by recognition because this task is more often reported in the literature. For this purpose, we designed a 4-AFC task in which we presented each of the target referents alongside three foils (a fourth foil was a highly familiar object that was used for practice and removed before the target was requested, see **Figure 1** for 4-AFC with sand-play stimuli and **Figure 2** for 4-AFC with music-play stimuli). One was a familiar object that was a semantic neighbor of the trained referent (e.g., *maraca* for *bekemite*, both being musical instruments), one was a familiar object whose name was a lexical neighbor of the trained referent, the name being the one on which the novel word was based (e.g., *vegemite* for *bekemite*), and one novel object that was introduced and named once during training (see below). The names of all of the familiar foil objects are common to the lexicons of Australian English speaking toddlers (MARCS Auditory Laboratories, 2004). To further ensure

Play theme	Highly familiar object	Target object	Novel/ unrelated object	Lexical neighbor foil	Semantic neighbor foil
Sand play	cow /kaʊ/	teggydair /tɛgidɛə/	paum /paʊm/	teddybear /tɛdibɛə/	mould 'can press sand with'
	pig /pɪg/	bargadoo /bagədu/	kas /kæs/	barbeque /babəkju/	scoop 'can scoop sand with'
	horse /hɔs/	pucky /pʌki/	noib /nɔɪb/	puppy /pʌpi/	cup 'can fill sand into'
	duck /dʌk/	norta /nɔtə/	gef /gɛf/	water /wɔtə/	press 'can hide sand in'

FIGURE 1 | Sand-play stimuli in the AFC task.

FIGURE 2 | Music-play stimuli in the AFC task.
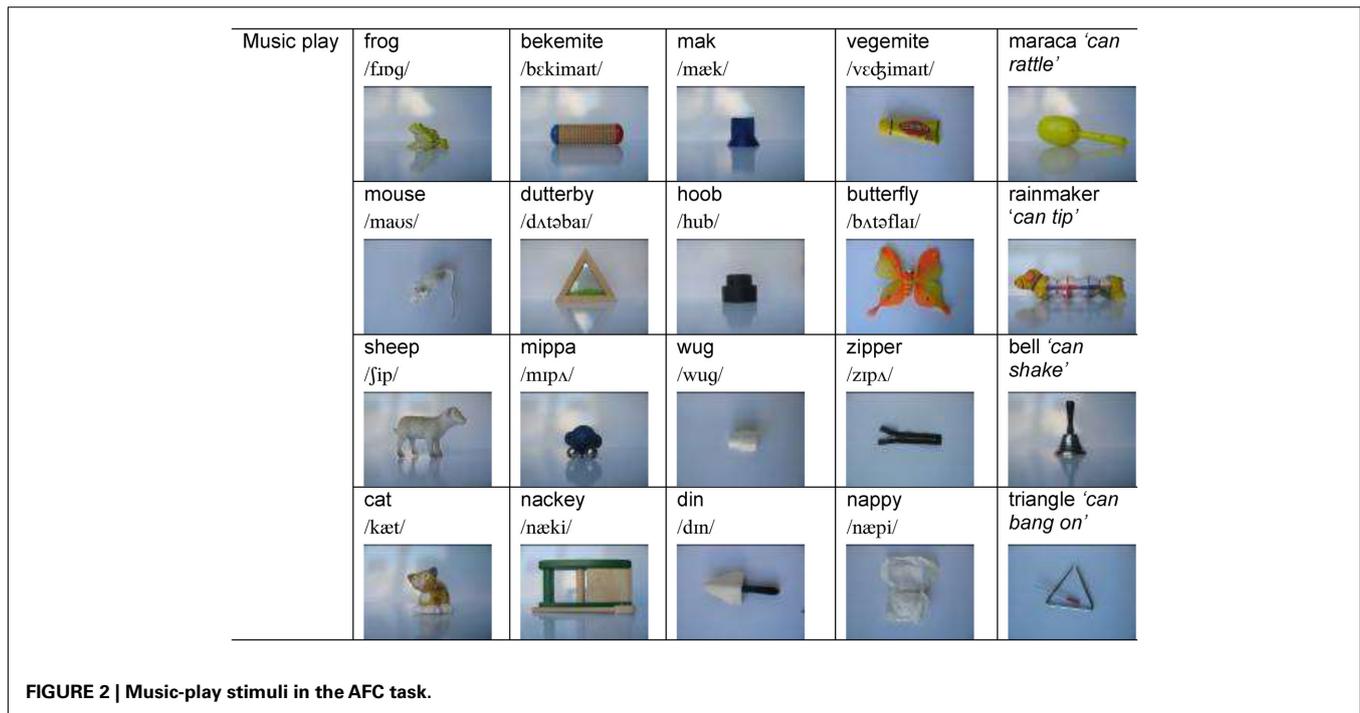
familiarity, the examiner showed and named pictures of each prior to the training procedure.

TRAINING PROCEDURE

Children attended two sessions. The first consisted of training and three production tests at immediate, 1-min, and 5-min retention intervals. The second, scheduled on average 4 days later (range = 1–7 days), consisted of a single retention test.

In the first session, children participated in two 20-min training episodes consisting of the sand- and music-play contexts. The children were introduced to the eight referent toys one at a time. Each referent toy was presented along with two toy foils (an unrelated highly familiar object, such as a frog, and an unrelated novel object) in the context of a discovery-and-play game.

The examiner followed a prepared script to ensure that the child was exposed to the foil and target objects and the words that named them. Each target word was produced six times. For example, for the target word *bekemite*, the script was: "I'm looking for a *bekemite*. This isn't it; it's a frog. Oh, this isn't it; it's a *mak*. Ah ha, here it is. This is a *bekemite*. This is a red and blue *bekemite*. You can rattle a *bekemite*. Let me show you. *Bekemite*! You say *bekemite*. It's your turn to play with it now." Note that this procedure was designed to prevent floor-level performance. The saliency of the link between word form and meaning was enhanced by ostensive naming and the word form itself was enhanced via multiple repetitions (see Horst and Samuelson, 2008, Experiment 2 for a similar procedure).

Children were randomly assigned to one of four training conditions that varied in gestural support for mapping. Four word training conditions were counterbalanced across the sand and musical toys and included; (1) +phonological gesture/+semantic gesture; (2) −phonological gesture/+semantic gesture; (3) +phonological gesture/−semantic gesture; or (4) −phonological gesture/−semantic gesture (no gestures). Each time the target word was said during this script, the examiner concurrently gestured as appropriate for the training condition. The examiner gestured the phonology by using her index finger to tap the length of the word in syllables near her mouth as she spoke the word itself using normal prosody. She gestured the semantics of the referent by using her hands to portray the shape of the object referent as she spoke the word, again using normal prosody. As noted above, production performance did not vary with training condition.

To ensure procedural fidelity, the examiner used a written script and rehearsed the procedures with pilot participants until she reached a range above 98% accuracy. In addition, at the end of the data collection phase of the study, an independent research assistant reviewed the video footage and corresponding transcripts across all the sessions of each participant and calculated procedural fidelity. Procedural fidelity for the examiner on the training protocol was above 99%.

TESTING PROCEDURE

Following are the test procedures in the order they were presented.

1) Immediate repetition: the first production of the word was elicited via imitation during the training script ("you say *bekemite*"). The examiner provided no feedback on accuracy. This task provided a baseline against which to judge encoding and consolidation performance over time.

2) One-minute retention: following the imitation attempt, the child played with the novel referent toy for 1 min. The examiner then picked up the toy and said ("That was fun wasn't it. What was this one called?"). Again, no feedback was provided to the child about his/her production accuracy or lack of

recall as the case may be. This task tapped encoding, especially encoding of the word form and expressive link.

3) After all of the four toys in a given set were trained and tested for immediate and 1-min production, the examiner administered the recognition test for all words/referents. The order of testing was the same as the order of training. The children were introduced to Tommy the puppet and asked to look at an array of five objects consisting of the two foil objects used in word training (e.g., frog and *mak*) and two foil objects not used during training one being phonologically similar to the target (e.g., a tube of vegemite, in the case of *bekemite*) and the other being semantically similar (e.g., maraca, in cases where *bekemite* was the name of a musical instrument). The children were first asked to give the puppet the highly familiar object ("give Tommy the frog"), making this a 4-AFC probe. Next, the children were asked to give the puppet the target object ("give Tommy the *bekemite*"). If the child was incorrect, the examiner cued the child with a semantic gesture (the same used in the semantic gesture training condition) within the carrier phrase: "Give Tommy the (semantic gesture)" Given that gestures were iconic, this was done irrespective of whether the child received the semantic gesture during the training. If the child was again incorrect then the examiner moved to the next activity without giving the child any feedback on his or her performance. This task tapped encoding, especially encoding of the receptive link.

4) Five-minute retention: next, the examiner packed away all of the toys except for the target and then asked the child; "what's this called?" If an incorrect or no response was provided, the examiner cued the response by producing the first syllable of the novel word and a phonological gesture (the same used in the phonological gesture training condition). This occurred irrespective of whether the child had received the phonological gesture training condition for that particular word. Again, this task tapped encoding of the word form and expressive link.

5) Multiday retention: the recognition test and then the production test were repeated 1–7 days later without any subsequent training. These tasks tapped consolidation.

RELIABILITY OF TRANSCRIPTION

All data were video recorded. The examiner transcribed each child's responses on-line and verified her transcriptions with the video tape. Twenty percent of production responses were randomly selected and transcribed from the video tape by an independent research assistant. Inter-rater point by point phoneme agreement between the original examiner and the research assistant was above 98%.

RESPONSE ANALYSIS

As further evidence of the source of retention problems, we categorized children's uncued and cued productions. In addition to correct responses we considered errors of seven types: (1) approximations of the target word form (e.g., /fizimait/for/bɛkimaik/); (2) substitution of a familiar lexical neighbor (e.g., vegemite for *bekemite*); (3) substitution of a training neighbor (e.g., mippa for *bekemite*); (4) substitution of a semantic neighbor (e.g., music for *bekemite* if that word had been assigned to label a musical instrument); (5) no/don't know responses; and (6) non-compliant

behavior. In the case of cued responses only, there was also (7) exact repetition of the syllabic cue.

RELIABILITY OF RESPONSE CODING

Twenty-five percent of production responses were randomly selected and coded by an independent research assistant. Inter-rater point by point agreement between the original examiner and the research assistant was 92%.

RESULTS

There was no effect of play context on production outcomes. Therefore, in all analyses reported below, the data are collapsed across contexts.

As for overall recognition (whether cued or uncued), the children averaged 69% accuracy (SD = 18) at the 5-min retention interval and 73% at the multiday retention interval (SD = 22). Of all errors, selection of the foil object whose name was a lexical neighbor of the target was most common, this representing 54% of all erred responses at the 5-min retention interval and 52% at the multiday retention interval. A preliminary analysis of the recognition responses appeared in Munro et al. (2011) and more detailed analysis will appear in a separate paper. Here it is essential to note that the children demonstrated above chance levels of recognition at the 5-min and multiday intervals $ps < 0.0001$.

With these preliminary findings in mind, we proceeded to the analyses of primary interest which concerned the production data. We examined correct word productions, correct syllable productions, and correct phoneme productions. We examined uncued productions first, then directly compared cued versus uncued productions.
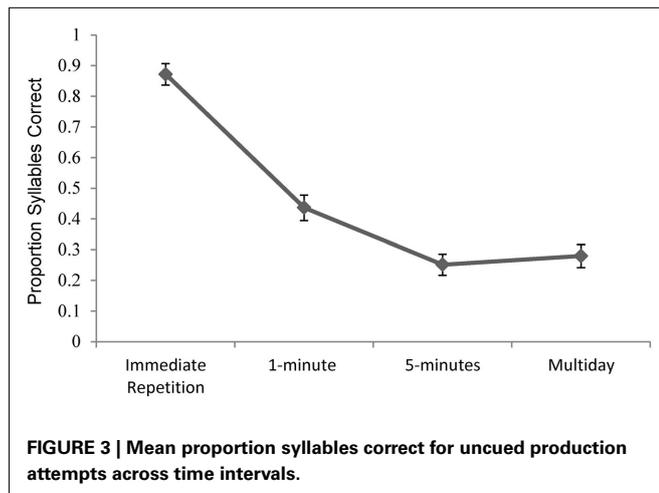
UNCUED PRODUCTIONS

Words

First, children's production responses were scored as either completely correct or incorrect. **Table 1** shows the mean correct responses (out of eight) at each time interval. The best performance is evident at immediate repetition where upon 71% (5.67/8) of words were produced correctly. Note however that, across individuals, performance ranged from no correct productions to eight correct productions. Therefore, *as a group*, the children demonstrated that they attended to, perceived, and could articulate the target word forms.

That said, group level accuracy declined sharply over time, with performance averaging only 19% (1.49/8) just 1 min later, $T = 1.5$, $z = 5.89, p < 0.0001$. There was additional decline from the 1- to 5-min interval, $T = 48.0, z = 3.79, p = 0.0001$. In terms of individual differences, there were some children who could correctly name all

Table 1 | Mean number of words produced correctly across time intervals.

	Immediate repetition	1-min retention	5-min retention	Multiday retention
Mean	5.67	1.49	0.43	0.33
SD	2.08	1.78	0.68	0.59
Range	0–8	0–8	0–2	0–2

FIGURE 3 | Mean proportion syllables correct for uncued production attempts across time intervals.

FIGURE 4 | Mean proportion phonemes correct for uncued production attempts across time intervals.

eight words at the 1-min interval; however, at the 5-min interval, the best performing children named only two words correctly. After these sharp declines, performance remained fairly stable from 5-min to multiday retention, $T = 72.5$, $z = 0.91$, $p = 0.37$.
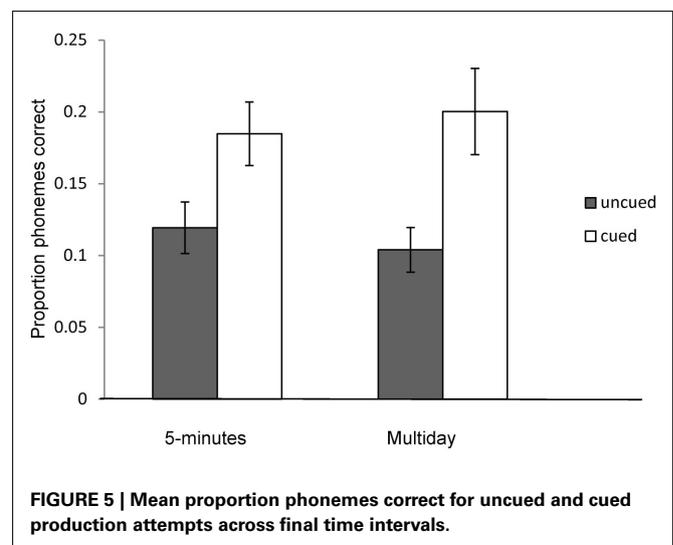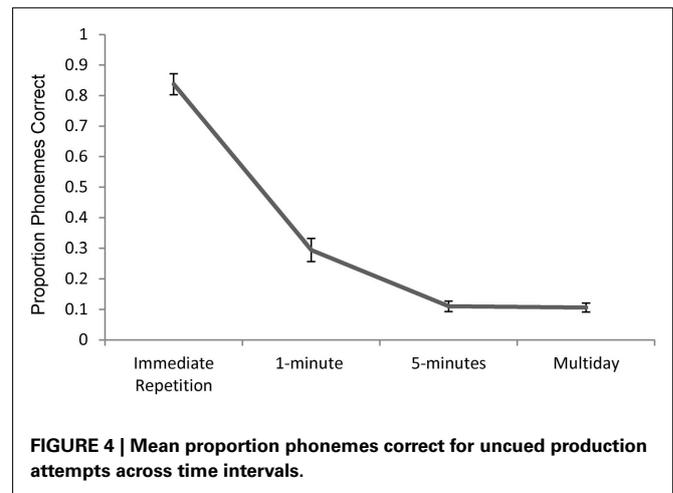
Syllables

We calculated the proportion of syllables correct for each uncued production attempt across each time interval (**Figure 3**). These were calculated based on the child's response relative to the target's syllabic structure. For example, a child's response such as /bɛki/ for *bekemite* would equate to 0.66 syllables correct. A repeated measures ANOVA revealed a significant main effect for time interval $F(3,141) = 92.5239$, $p < 0.0001$, partial $\eta^2 = 0.658$. *Post hoc* pairwise comparisons with Bonferroni adjustments revealed that the children's syllable accuracy decreased significantly from immediate repetition to all retention intervals, as well as between the 1-min retention interval and all later retention intervals (ps < 0.002) but not between the 5-min and multiday retention intervals.

Phonemes

We calculated the proportion of phonemes correct for each uncued production attempt across each retention interval (**Figure 4**). These were calculated based on the child's response relative to the target. For example, a child's response such as /bɛki/ for *bekemite* would equate to 0.57 phonemes correct. There was a significant main effect for time interval $F(3,144) = 196.19$, $p < 0.0001$, partial $\eta^2 = 0.80$. *Post hoc* pairwise comparisons with Bonferroni adjustments revealed that the children's phoneme accuracy significantly decreased from immediate repetition to all retention intervals, as well as between the 1-min and all later retention intervals (ps < 0.0001) but not between the 5 min- and multiday retention intervals.

UNCUED VERSUS CUED: SYLLABLES

We compared uncued and cued productions directly. Recall that following incorrect or no responses, the examiner provided a cue that included the first syllable of the target. Therefore, to allow comparison, we did not count the child's production of the first syllable in either uncued or cued responses in this particular analysis. With this modified version of our dependent variable, we ran

FIGURE 5 | Mean proportion phonemes correct for uncued and cued production attempts across final time intervals.

a 2×2 repeated measures ANOVA for retention interval (5-min, multiday) and cue (uncued, cued). There were no main effects of retention interval $F(1,42) = 1.01$, $p = 0.32$, or cue, $F(1,42) = 1.29$, $p = 0.26$. The proportion of accurate syllable productions at the 5-min retention interval without a cue averaged 0.21 (SE = 0.03) and with a cue averaged 0.22 (SE = 0.04). At the multiday retention interval, accuracy without a cue averaged 0.20 (SE = 0.04) and with a cue averaged 0.28 (SE = 0.05).

UNCUED VERSUS CUED: PHONEMES

The uncued and cued proportion phonemes correct data were also subjected to a 2×2 repeated measures ANOVA for retention interval (5-min, multiday) and cue (uncued, cued) again after modifying the scoring of the uncued and cued data so that phonemes within the first syllable were not counted. A significant main effect was found for cue $F(1,42) = 17.35$, $p < 0.0002$, partial $\eta^2 = 0.29$, with cued productions being significantly more accurate than uncued productions. There was no effect for retention interval, $F(1,42) < 1$. **Figure 5** displays the mean proportion of phonemes correct for cued and uncued productions at the final two retention intervals.

Table 2 | Response types by time interval.

Response		Time interval					
		Imm	**1 min**	**5 min uncued**	**5 min cued**	**Multiday uncued**	**Multiday cued**
Correct (complete representation)	M	0.70	0.18	0.05	0.04	0.04	0.07
	(SD)	(0.27)	(0.21)	(0.08)	(0.09)	(0.07)	(0.11)
Partial representation	M	0.18	0.15	0.06	0.23	0.05	0.19
	(SD)	(0.16)	(0.15)	(0.12)	(0.21)	(0.10)	(0.21)
Approximation	M	0.16	0.12	0.04	0.20	0.05	0.17
	(SD)	(0.14)	(0.13)	(0.07)	(0.19)	(0.10)	(0.19)
Lexical neighbor	M	0.02	0.02	0.02	0.02	0.01	0.02
	(SD)	(0.06)	(0.05)	(0.08)	(0.06)	(0.02)	(0.05)
No representation	M	0.11	0.68	0.89	0.73	0.90	0.74
	(SD)	(0.24)	(0.29)	(0.16)	(0.24)	(0.12)	(0.26)
Training neighbor	M	0.00	0.07	0.07	0.03	0.18	0.06
	(SD)	(0.00)	(0.12)	(0.13)	(0.07)	(0.23)	(0.13)
Semantic neighbor	M	0.01	0.10	0.13	0.02	0.11	0.01
	(SD)	(0.03)	(0.14)	(0.15)	(0.05)	(0.15)	(0.04)
No/DK	M	0.10	0.50	0.68	0.33	0.59	0.35
	(SD)	(0.23)	(0.29)	(0.24)	(0.26)	(0.32)	(0.37)
Non-compliant	M	0.00	0.00	0.01	0.09	0.01	0.05
	(SD)	(0.00)	(0.00)	(0.03)	(0.12)	(0.03)	(0.09)
Repetition of cue	M				0.26		0.25
	(SD)				(0.27)		(0.35)

Imm, immediate; DK, don't know.

UNCUED AND CUED RESPONSE TYPE ANALYSIS

Response types appear in **Table 2** with subtypes grouped into categories that evince either complete, partial, or no representation of the target word form. Upon immediate repetition, errors were few and fell overwhelmingly into two subtypes: approximations and no/don't know responses. During the retention intervals, as errors increased in number, more subtypes of error were represented; however, in terms of raw numbers no/don't know responses were the single most frequent subtype at all retention intervals whether cued or uncued. At times, cueing did enable shifts from errors that evinced no representation to errors that evinced partial representation at both the 5-min retention interval (increase in partial representation from uncued to cued: $T = 22.5$, $z = 4.6$, $p < 0.0001$; decrease in no representation from uncued to cued: $T = 93$, $z = 3.64$, $p = 0.0003$) and at the multiday retention interval (increase in partial representation from uncued to cued: $T = 54$, $z = 3.8$, $p = 0.0001$; decrease in no representation from uncued to cued: $T = 70$, $z = 4.01$, $p < 0.0001$). Once accuracy leveled off, error types remained largely stable; that is, the response profiles for the 5-min and multiday retention intervals appeared remarkably similar. The one exception was the increased proportion of training neighbor substitutions from the uncued 5-min retention interval to the uncued multiday retention interval, $T = 59$, $z = 3.12$, $p = 0.002$.

INDIVIDUAL DIFFERENCES

To better describe relationships between phonological short term memory, extant vocabulary, and word learning in the moment, we ran a series of regression analyses. Percentage of phonemes accurately produced at each retention interval served as the outcome variable and phonological short term memory estimated by percentage of phonemes accurately produced on the TENR, extant vocabulary estimated by PPVT-IV raw scores, and chronological age in months were the independent variables. The two independent variables of primary interest, the PPVT-IV and the TENR, did not correlate with each other, $r = 0.002$, $p = 0.99$. The regression models yielded significant fit at the 1-min retention interval, cued 5-min retention interval (but not the uncued 5-min intervals) and uncued and cued multiday interval. In all cases it was the PPVT-IV scores and only those scores that accounted for variance in accuracy of productions, $r^2 = 0.21$. At 1 min, accuracy of word form production correlated with PPVT-IV scores, $r_{partial} = 0.40$, $p = 0.007$ but not TENR, $r_{partial} = 0.13$, $p = 0.38$, or age, $r_{partial} = -0.02$, $p = 0.92$. Again, at 5 min, accuracy of cued word form production correlated with PPVT-IV scores, $r_{partial} = 0.47$, $p = 0.001$ but not TENR, $r_{partial} = -0.01$, $p = 0.93$, or age, $r_{partial} = -0.09$, $p = 0.57$. Finally, at the multiday retention interval, accuracy of unprompted word form production again correlated with PPVT-IV scores, $r_{partial} = 0.39$, $p = 0.008$ but not TENR, $r_{partial} = 0.04$, $p = 0.78$ or age, $r_{partial} = 1.22$, $p = 0.13$. The same pattern held for cued word form production at the multiday interval, PPVT-IV scores, $r_{partial} = 0.57$, $p = 0.00004$, TENR, $r_{partial} = 0.17$, $p = 0.26$, and age, $r_{partial} = 1.12$, $p = 0.43$. We conclude that individual differences in the production retention data were related to size of the extant vocabulary but not phonological short term memory abilities.

DISCUSSION

The goal of the current study was to determine why memory for words encountered for the first time is so fragile. Toddlers were taught eight new words and their referents. They were very good at selecting the correct referents from an array as the examiner requested as measured at the 5-min interval and they retained that ability as measured at the multiday interval. In this sense, they were the amazing fast mappers much discussed in the literature. Presumably, they had encoded acoustic-phonetic representations that supported recognition and those representations stabilized during consolidation. However, they did not retain the ability to *produce* the new words over a 1-min interval. That is, their fast mapping experience did not yield a memory trace that could support production. In this sense as well, they performed according to other reports in the literature (Carey, 1978; Horst and Samuelson, 2008). The data suggest that it was not the post-encoding process of consolidation, but the process of encoding itself that resulted in children's relatively poor retention of word forms.

FAST MAPPING AND RETENTION IN THE CURRENT DATA SET

The children demonstrated fast mapping by performing at levels significantly above chance on a 4-AFC recognition probe at the 5-min retention interval. The probe itself was designed to require discrimination between the target and the foils based on shared episodic information (the unrelated foil that was also present during training), shared semantics (the foil that was a semantic neighbor), or shared phonology (the foil whose name was a lexical neighbor). Therefore, given their good performance, we can be assured that the children had mapped some of the receptive links as well as some critically distinctive acoustic-phonetic information about the word form and referent.

Importantly, the children were equally good when the recognition test was repeated after a multiday interval. That is, they *did* retain this particular information. This should not be taken as contradicting the thesis that retention following fast mapping is fragile. It is commonly accepted that the receptive link is easier to establish and maintain than the expressive link (Dollaghan, 1985; Gray, 2003, 2004; Gupta, 2005; Booth et al., 2008). Moreover, the training environment used in the current study was highly supportive of a good outcome as it involved ostensive labeling of the referent (Horst and Samuelson, 2008), multiple repetitions of the target word form (McGregor et al., 2007), multiple opportunities for the child to attempt the word prior to the recognition test, familiar categories (i.e., sand and music toys; Mayor and Plunkett, 2010), semantic contrasts (e.g., "this is not it! This is a frog!"; Gottfried and Tonks, 1996), established lexical neighborhoods (e.g., *bekemite* being a neighbor of vegemite; Storkel, 2001), and the chance to imitate the word form (Masur, 1995) and manipulate the referent (Scofield et al., 2009), each of which has been proven to facilitate learning.

On the other hand, consider that the majority of errors that did occur on the 4-AFC recognition probe were selections of the lexical neighbor foil. It could be that, in these cases, the child retained only underspecified acoustic-phonetic representations of the target word, or indeed had such limited representations that the examiner's use of the target word simply activated the more richly specified representation of a lexical neighbor. This suggestion is

more clearly borne out in the production data that we present below. To begin to account for the oblivescence that follows fast mapping, we first turn to the memory model that motivated our approach to the problem.

GROUNDING FAST MAPPING AND WORD LEARNING IN MEMORY

We began with a two-part model of memory in which the ordered processes of encoding and consolidation support word learning and retention. While these processes serve ultimately to build the long term lexicon, the extant lexicon serves to scaffold these processes in real time (e.g., Metsala and Walley, 1998). In fact, regression models revealed relationships between extant lexical knowledge and encoding and consolidation success. These relationships suggest the validity of our approach: if we consider our task to measure word learning, a lack of correlation between performance on our task and what children have already learned about words would have been problematic. Given this support for our general approach, we next critically evaluate evidence for two possible explanations of limited retention after fast mapping.

EXPLANATIONS FOR LIMITED RETENTION AFTER FAST MAPPING
Encoding of word forms and expressive links

Upon immediate repetition, the children averaged more than five perfectly produced words out of eight. Only 5 min later, they averaged less than one. This pattern of change held whether uncued productions were analyzed at the level of whole words, syllables, or phonemes.

The poor production accuracy comes as no surprise given numerous reports of floor-level production performance following fast mapping exposures (Dollaghan, 1985; Gray, 2003, 2004; Gupta, 2005; Booth et al., 2008; Horst and Samuelson, 2008). The new contribution here is that the comparison between immediate repetition and 1- and 5-min recall probes clearly identifies the encoding process as a bottleneck. Had the limitations been in perceptual processes that limit encoding or in articulatory processes that limit the child's ability to express what had been encoded, then we should have seen poor performance upon immediate repetition because immediate repetition also depends upon perception and articulation. Although a few individual children were poor at immediate repetition, the group means were strong: 71% (5.67/8) of the target words (86% of the phonemes) were produced correctly.

Which aspect of encoding was problematic? We think it unlikely that limitations in phonological short term memory were at play as immediate repetition was strong and as an independent measure of phonological short term memory, the TENR, bore no relationship to encoding performance at the 1- or 5-min retention intervals.

The error patterns in the uncued and cued productions hold some clues as to a more likely locus of difficulty. We reasoned that, if limitations occur in linking word meanings to forms, then the child's errors may involve training neighbor substitutions. Also, the child's production should improve following cues that convey information about the word form, in this case the production of the first syllable accompanied by a beat gesture highlighting word length in syllables, because that information specifies the expressive link for the child. There were cases of training substitutions; however, they represented only

7% of all responses at the 1- and 5-min retention intervals. Cueing reduced the proportion to 3%. These low frequencies suggest that encoding the expressive link was not particularly problematic.

In contrast, encoding of word forms themselves, appeared to be a challenge. The three most frequent errors at the 1- and 5-min retention intervals were no/don't know, semantic substitutions, and approximations. We originally viewed no/don't know and semantic substitutions as indicative of absent form representations. However, when cueing followed these response types, approximations increased. Moreover, we must consider that recognition performance was strong at the 5-min interval and some knowledge of the word form is necessary even in a recognition task. Therefore, it is likely that some no/don't know and semantic responses indicated partial form representations (and a conservative responder). Overall, we conclude that, over the time course of encoding, articulatory-phonetic representations decayed to a point that they were too weak or underspecified to support production.

Consolidation of word forms and expressive links
The production accuracy, whether uncued or cued, demonstrated at the 5-min retention interval, though low, was maintained at the multiday retention interval. This was evident whether accuracy was measured at the word, syllable, or phoneme level. Moreover, response types and response to cueing were very similar at the 5-min and multiday intervals. We conclude that consolidation was not the primary bottleneck to retention of word forms. This accords with other studies of children's word learning wherein performance days after training was as good or better than performance immediately post-training (Rice et al., 1994; Storkel, 2001; Booth, 2009; McGregor et al., 2009; Norbury et al., 2010).

The only exception to the stability of response profiles was an increased proportion of training neighbor substitutions after the multiday interval. We hypothesize that our ordering of the recognition probe prior to the production probe during the multiday posttest may account for this increase. Because the children heard each word form once during the recognition probe, the forms themselves were primed but perhaps the children did not have robust memories of the expressive links that enabled them to produce these forms in response to the appropriate referents. This observation highlights a limitation of the study. Namely, the expressive consolidation measure was not purely a reflection of consolidation but also of the one additional exposure that the children had during recognition testing.

WHY ENCODING IS DIFFICULT
To effectively support comprehension and production, acoustic-phonetic, articulatory-phonetic, and lexical aspects of word

forms must be represented in long term memory (Rvachew and Brosseau-Lapré, 2010). Therefore, part of the encoding bottleneck might reflect the large size of the problem space. Another way to think about this is that representations of words in the long term lexicon must involve both fine-grained acoustic-phonetic information as well as coarser-grained, context-independent generalizations about phonological structure (Pierrehumbert, 2003; Buchwald and Miozzo, 2011; Munson et al., 2011). The latter, in particular, would take time to develop as generalizations are necessarily abstracted over multiple instances.

Despite our continual use of the term "bottleneck," we recognize that children's limitations in the encoding of new word forms might be adaptive. This recognition is prompted by the descriptions of infants' word learning in Hollich et al. (2002) and the connectionist model in McClelland et al. (1995). Hollich and colleagues describe children as conservative word learners who require much evidence before they will add a new word to their lexicons. McClelland and colleagues find that rapid sequential acquisition of new data can lead to catastrophic interference in a learning network. Both groups point out that learning that involves small, gradual changes will allow abstraction of general patterns and will prevent undue influence from individual exemplars. To take a concrete example, children must be able to recognize known words when they hear them produced by different speakers and in different contexts. Therefore it would be problematic if novelty (e.g., an unfamiliar voice) always triggered the creation of a new lexical entry. Children's conservatism is certainly not consciously strategic but limitations on the amount of information that can be encoded at any point in time might indirectly force this conservatism.

CONCLUSION
Although children can and do fast map information about word forms, referents, and the links between them, this information is susceptible to oblivescence. We found that the ability to produce newly encountered word forms was particularly fragile. The outcomes of encoding were limited but memory consolidation was relatively robust. Thus, encoding limitations ensure that word learning is not fast.

ACKNOWLEDGMENTS

REFERENCES
Booth, A. E. (2009). Causal supports for early word learning. *Child Dev.* 80, 1243–1250.

Booth, A. E., McGregor, K. K., and Rohlfing, K. J. (2008). Socio-pragmatics and attention: contributions to gesturally guided word learning in toddlers. *Lang. Learn. Dev.* 4, 179–202.

Buchwald, A., and Miozzo, M. (2011). Finding levels of abstraction in speech production: evidence from sound production impairment. *Psychol. Sci.* 22, 1113–1119.

Carey, S. (1978). "The child as word learner," in *Linguistic Theory and Psychological Reality*, eds M. Halle, J. Bresnan, and G. A. Miller

(Cambridge, MA: The MIT Press), 264–293.

Dollaghan, C. (1985). Child meets word: "fast mapping" in preschool children. *J. Speech Hear. Res.* 28, 449–454.

Dumay, N., and Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of

spoken words. *Psychol. Sci.* 18, 35–39.

Dunn, L. M., and Dunn, D. M. (2007). *Peabody Picture Vocabulary Test*, 4th Edn. Bloomington, MN: NCS Pearson, Inc.

Fenn, K. M., Nusbaum, H. C., and Margoliash, D. (2003). Consolidation during sleep of perceptual

learning of spoken language. *Nature* 425, 614–616.

Gais, S., Molle, M., Helms, K., and Born, J. (2002). Learning-dependent increases in sleep spindle density. *J. Neurosci.* 22, 6830–6834.

Goldman, R., and Fristoe, M. (2000). *Goldman Fristoe test of Articulation: 2.* Circle Pines, MN: American Guidance Service.

Gottfried, G. M., and Tonks, J. M. (1996). Specifying the relation between novel and known: input affects the acquisition of novel color terms. *Child Dev.* 67, 850–866.

Gray, S. (2003). Word-learning by preschoolers with specific language impairment: what predicts success? *J. Speech Lang. Hear. Res.* 46, 56–67.

Gray, S. (2004). Word learning by preschoolers with specific language impairment: predictors and poor learners. *J. Speech Lang. Hear. Res.* 47, 1117–1132.

Gupta, P. (2005). What's in a word? A functional analysis of word learning. *Perspect. Lang. Learn. Educ.* 12, 4–8.

Hollich, G., Jusczyk, P., and Luce, P. (2002). "Lexical neighborhood effects in 17-month-old word learning," in *Proceedings of the 26th Annual Boston University Conference on Language Development* (Boston, MA: Cascadilla Press).

Horst, J. S., and Samuelson, L. K. (2008). Fast mapping but poor retention in 24-month-old infants. *Infancy* 13, 128–157.

MARCS Auditory Laboratories. (2004). *Australian English Developmental Vocabulary Inventory – OZI.* Sydney: MARCS Auditory Laboratories, University of Western Sydney.

Martin, N., and Gupta, P. (2004). Exploring the relationships between word processing and verbal short-term memory: evidence from associations and dissociations. *Cogn. Neuropsychol.* 21, 213–228.

Masur, E. F. (1995). Infants early verbal imitation and their later lexical development. *Merrill Palmer Q.* 41, 286–306.

Mayor, J., and Plunkett, K. (2010). A neuro-computational account of taxonomic responding and fast mapping in early word learning. *Psychol. Rev.* 117, 1–31.

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–457.

McGregor, K. K., Rohlfing, K., Bean, A., and Marschner, E. (2009). Gesture as a support for word learning: the case of under. *J. Child Lang.* 36, 807–828.

McGregor, K. K., Sheng, L., and Ball, T. (2007). Complexities of word learning over time. *Lang. Speech Hear. Serv. Sch.* 38, 1–12.

Metsala, J., and Walley, A. (1998). "Spoken vocabulary growth and the segmental restructuring of lexical representations: precursors to phonemic awareness and early reading ability," in *Word Recognition in Beginning Literacy*, eds J. L. Metsala and L. C. Ehri (Mahwah, NJ: Erlbaum), 89–120.

Munro, N., Baker, E., McGregor, K. K., Arciuli, J., and Docking, K. (2011). "Iconic gestures support toddlers' retention of word-referent pairings," in *12th International Congress for the Study of Child Language*, Montreal.

Munson, B., Edwards, J., and Beckman, M. E. (2011). "Phonological representations in language acquisition: climbing the ladder of abstraction," in *Handbook of Laboratory Phonology*, eds A. C. Cohn, C. Fougeron, and M. K. Huffman (Oxford: Oxford University Press), 288–309.

Norbury, C., Griffiths, H., and Nation, K. (2010). Sound before meaning: word learning in autistic disorders. *Neuropsychologia* 48, 4012.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Lang. Speech* 46, 115–154.

Rice, M., Oetting, J., Marquis, J., Bode, J., and Pae, S. (1994). Frequency of input effects on SLI children's word comprehension. *J. Speech Hear. Res.* 59, 106–122.

Rvachew, S., and Brosseau-Lapré, F. (2010). "Speech perception intervention," in *Treatment of Speech Sound Disorders in Children*, eds L. Williams, S. McLeod, and R. McCauley (Baltimore, MD: Paul Brookes Publishing Co.), 295–314.

Scofield, J., Hernandez-Reif, M., and Keith, A. B. (2009). Preschool children's multimodal word learning. *J. Cogn. Dev.* 10, 306–333.

Stokes, S. F., and Klee, T. (2009). The diagnostic accuracy of a new test of early nonword repetition for differentiating late talking and typically developing children. *J. Speech Lang. Hear. Res.* 52, 872–882.

Storkel, H. L. (2001). Learning new words: phonotactic probability in language development. *J. Speech Lang. Hear. Res.* 44, 1321–1337.

Walker, M. P. (2005). A refined model of sleep and the time course of memory formation. *Behav. Brain Sci.* 28, 51–104.