Sometimes it is better to know less:

How known words influence referent selection and retention in 18 to 24-month-old children

Sarah C. Kucker[1], Bob McMurray[2,4], and Larissa K. Samuelson[3,4]

[1]Oklahoma State University (116 N Murray Hall, Stillwater, OK, USA), [2]The University of Iowa (E11 Seashore Hall, Iowa City, IA, USA), [3]The University of East Anglia (0.09B, Lawrence Stenhouse Building, Norwich NR4 7TJ, United Kingdom) and [4]The Delta Center (The University of Iowa, Iowa City, IA, USA)

In Press (Aug 25, 2019), Journal of Experimental Child Psychology

Authors Note

Address correspondences to Sarah Kucker, Oklahoma State University, Department of Psychology, 116 N Murray Hall, Stillwater, OK 75078; email: sarah.kucker@okstate.edu

Declarations of interest: none

Word count: 10,000

Abstract

Young children are surprisingly good word learners. Despite their relative lack of world knowledge and limited vocabularies, they consistently map novel words to novel referents and, at later ages, show retention of these new word-referent pairs. Prior work has implicated the use of mutual exclusivity constraints and novelty biases, which require that children use knowledge of well-known words to disambiguate uncertain naming situations. The current study presents evidence, however, that *weaker* vocabulary knowledge during the initial exposure to a new word may be better for retention of new mappings. Children between 18- and 24-months of age selected referents for novel words in the context of foil stimuli that varied in their lexical strength and novelty: well-known items (e.g. shoe), just-learned weakly known items (e.g. wif), and completely novel items. Referent selection performance was significantly reduced on trials with weakly known foil items. Surprisingly, however, children subsequently showed above chance retention for novel words mapped in the context of *weakly known* competitors, compared to those mapped with strongly known competitors or with completely novel competitors. We discuss implications for our understanding word learning constraints and how children use known words and novelty during word learning.

Keywords: word learning, referent selection, vocabulary knowledge, fast-mapping, novelty

Word Count: 194

Highlights

- Toddlers struggle to link new words to objects in the context of weakly known words
- New word-object links made in the context of weakly-known words can be retained
- Word-object mappings formed in the context of well-known words may not be retained
- Factors that support word-object mapping may not also support retention

Sometimes it is better to know less: How known words influence referent selection and retention in 18- to 24-month-old children

Young children's vocabulary building prowess is often captured in simple laboratory tasks: when presented with a novel word, a novel object, and multiple known objects, children as young as 18-months consistently reject the known objects and select the novel one. This is thought to suggest constraints or biases are needed to limit the referential ambiguity inherent in word learning. Many of these constraints suggest that children's rich prior knowledge of known words/objects and the novelty of unknowns supports children's selections. However, children's limited knowledge at this age means all words/objects are not equally well known – sometimes vocabulary knowledge coincides with familiar words presented, other times knowledge of familiar items is weak or absent. Here we examine word learning when individual familiar competitors are more or less well known.

What supports referent selection and retention?

A number of classic constraints such as mutual exclusivity (Baldwin & Markman, 1989; Markman & Wachtel, 1988) and novel-name nameless-category (N3C; Mervis & Bertrand, 1994) focus on prior knowledge to support referent selection and word learning. While much of the supporting evidence comes from referent selection tasks, recent work has demonstrated that real-time referent selection is at least partially distinct from long-term retention and learning (Horst & Samuelson, 2008; Kucker, McMurray, & Samuelson, 2016; Carey, 2010). Most prominently, Horst and Samuelson (2008) found two-year-old children were quite accurate at selecting the novel word in a referent selection task. However, when retention was tested five-minutes later, they were at chance. Thus, successful real-time referent selection may not guarantee retention (see also Bion, Borovsky, & Fernald, 2013; Kucker, McMurray, &

Samuelson, 2016). While some learning surely takes place when the referent of a novel word is selected, much more learning is necessary to transform initial mappings to ones that support retention (Carey, 2010; McMurray et al., 2012).

Further studies have clarified the relation between referent selection and retention. For example, retention rates change when altering the number of foils (Horst, Scott, & Pollard, 2010; Zosh, Brinster, & Halberda, 2013), biasing attention to targets (Axelsson, Churchley, & Horst, 2012), or providing prior familiarity with the novel items (Kucker & Samuelson, 2012). Likewise, successful retention results from repetition of the name-object pair before test (see Horst & Samuelson, 2008 for review).

There is also evidence that children's referent selection and retention abilities change with development. By 24-months children are good at referent selection, but retention is not reliable until at least 30-months-of-age (Bion et al., 2013). This fits with Carey and Bartlett's (1978) original demonstration of retention in preschoolers. Likewise, recent work has demonstrated the overall size of a child's vocabulary predicts performance in both referent selection (Kalashnikova, Mattock, & Monaghan, 2016) and retention (Bion et al., 2013). The picture presented by the accumulated literature then is one of malleable and changing abilities over development. Here we seek further understanding of these effects by examining the role of children's developing vocabulary knowledge in supporting selection of referents for novel words and retention of those mappings.

The Role of Vocabulary Knowledge

Vocabulary knowledge is usually defined in one of two ways: the overall size of a child's productive lexicon (Kalashnikova et al., 2016) or, as with Mutual Exclusivity, a child's knowledge of individual words (which helps exclude them as referents; Grassmann, Schulze, &

Tomasello, 2015; Markman & Wachtel, 1988). In both cases, strong vocabulary knowledge is required for mapping novel words to their referents. When the objects' names are well known, children avoid them as targets for a new word, instead selecting a novel object. This has been demonstrated in various ages (Bion et al., 2013; Mather & Plunkett, 2010; Mervis & Bertrand, 1994), using looking (Bion et al., 2013) and reaching measures (Gurteen, Horne, & Erjavec, 2011), in atypical populations (Alt & Plante, 2006), and to some extent, in bilinguals (Byers-Heinlein & Werker, 2013; Kalashnikova, Escudero, & Kidd, 2018).

However, a child's ability to use prior knowledge for word learning might be directly tied to how strong that knowledge is. Disambiguation tests typically pit well-known items against completely novel objects. This contrasts sharply with a child's natural learning environment in which the majority of words are likely to be weakly known; children hear up to 2,000 words an hour, yet produce only two new words a day (Hart & Risley, 1995), suggesting many words may be only partially known (see McMurray, 2007, for implications). Moreover, children's knowledge of individual words is still developing throughout the second year (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998) and into adulthood (McGregor, 2014), thus even "known" words vary considerably in strength. Some words are frequent in daily life (MacWhinney, 2000), and are part of dense semantic networks (Hills, Maouene, Riordan, & Smith, 2010; Hills, Maouene, Maouene, Sheya, & Smith, 2009a), whereas others are limited to a context and have only tangential associations with the rest of the lexicon (Hills, Maouene, Maouene, Sheya, & Smith, 2009b; Storkel, 2009). Given that children are also learning many words in parallel, real-time processing of words must occur in environments that are relatively ambiguous and with weak knowledge.

Few laboratory tests of word-referent mapping tap into this variability, and fewer have manipulated children's ability to draw on vocabulary knowledge during word learning. A recent study by Kalashnikova and colleagues (2016) pitted just-learned word-referent pairs against novel items, finding improvements in novel referent selection as overall vocabulary size increased. However, there was no comparison with well-known foils (e.g. *shoe, dog*), and no retention test. Another study by Merriman and Schuster (1991) made referent knowledge harder to recruit via word pronunciation distortions and non-typical exemplars of known items. Less typical pronunciations and referents reduced selection performance (see also Meints, Plunkett, & Harris, 1999). Conversely, novel referent selection is stronger in the context of words that children both comprehend and produce, compared to words they only comprehend (Grassmann et al., 2015). Together, these studies demonstrate that more lexical ambiguity is detrimental to referent selection of novel words.

It is not clear how the strength of known-word knowledge influences retention. The only studies of vocabulary knowledge addressing retention suggest that at least by 30-months-of-age, productive vocabulary size matters (Bion et al., 2013). Here, we present some of the first evidence that a critical factor in retention of new word-object mappings is not just overall vocabulary size, but strength of knowledge for the *individual* known words present during referent selection.

The Current Study

The current study examined children from 18- to 24-months-old, spanning a range of vocabularies. In each experiment, children were asked to select the referent of a novel word in the context of known foil items that children knew more or less well. In Experiments 1 and 2, novel items were presented with items that were well-known (e.g., "cup") or just-learned. We expected referent selection to be poorer when foil items were weakly known, and expected children's poor retention to remain unchanged. Experiment 3 used foil items without names to examine the role of novelty. Based on prior work (Horst et al., 2011), we expected children would pick the most novel item as a referent and did not have strong expectations for retention. We discuss the implications of word knowledge strength for theories of word learning and its importance for understanding children's language processing.

Experiment 1

Methods

A critical factor of this study is the need to differentiate well-known and weakly known word-referent pairs. There are methodological issues in assessing whether a child's current knowledge of particular words is strong or weak prior to test. Pre-tests may bias or prime that word. Tests after the experiment make it difficult to control stimulus exposure across children. Thus, we used parent report to determine highly familiar words, and a pre-training to create a set of words that were nominally weakly-known. We then checked this classification via a comprehension test.

Participants. Forty-two 18- to 24-month-old children (17 females, *M*=22 months, 26 days; range 18;11-25;24) with a mean vocabulary of 242.9 words (range: 1-645, median: 184) participated. Data for three additional children were not included due to fussiness (2) and experimenter error (1). Children received a small prize for participating. Informed consent was obtained.

Stimuli. There were three sets of word-object pairs: highly familiar objects with well-known labels (strongly known), less familiar objects with newly-learned, weakly-known labels (weakly known), and completely novel objects with no label (novel). Strongly known stimuli were drawn from the LEX database and were ones which most children have prior experience (Dale & Fenson, 1993) and are known by 66% of 18-month-olds and 85% of 24-month-olds (Fenson et al., 1994). Prototypical, 3-D representations of each item easily categorized/labeled at this age were used (Mareschal & Tan, 2007; Nelson, Zhang, & McKinney, 2001). In contrast, the weakly known stimuli were initially novel items whose labels were learned during training prior to the main task. Thus, children's experience with strongly known stimuli was lengthier and more varied than their experience with the new weakly known stimuli, which was short and limited to a single exemplar in a single lab context– factors known to lead to significant differences in representations and learning (Perry, Samuelson, Malloy, & Schiffer, 2010; Wagar & Dixon, 2005). Novel items were distinct from the known items, were not seen by the child previously, and were confirmed by the parent to be unknown.

Prior to the main task, three known objects were selected from the pool of 16 strongly known items (Figure 1a). Parents confirmed their child was familiar with the name for each; most children were also reported to produce the label spontaneously. Items were replaced as needed. Thus, for the strongly known items, children had prior experience with at least one other exemplar in at least one context outside the laboratory. Likewise, 11 objects from a pool of 19 unfamiliar novel items were selected. Parents confirmed their status as unknown, with replacement as necessary. Three of these novel items became the weakly known items after training (Figure 1b); the remaining eight were used only on the referent selection trials and remained novel (Figure 1c).

Eleven novel words that conformed to the phonological rules of English but had no known referent were used as novel names. These words were drawn from a database of words from previous studies (Horst & Hout, 2014).

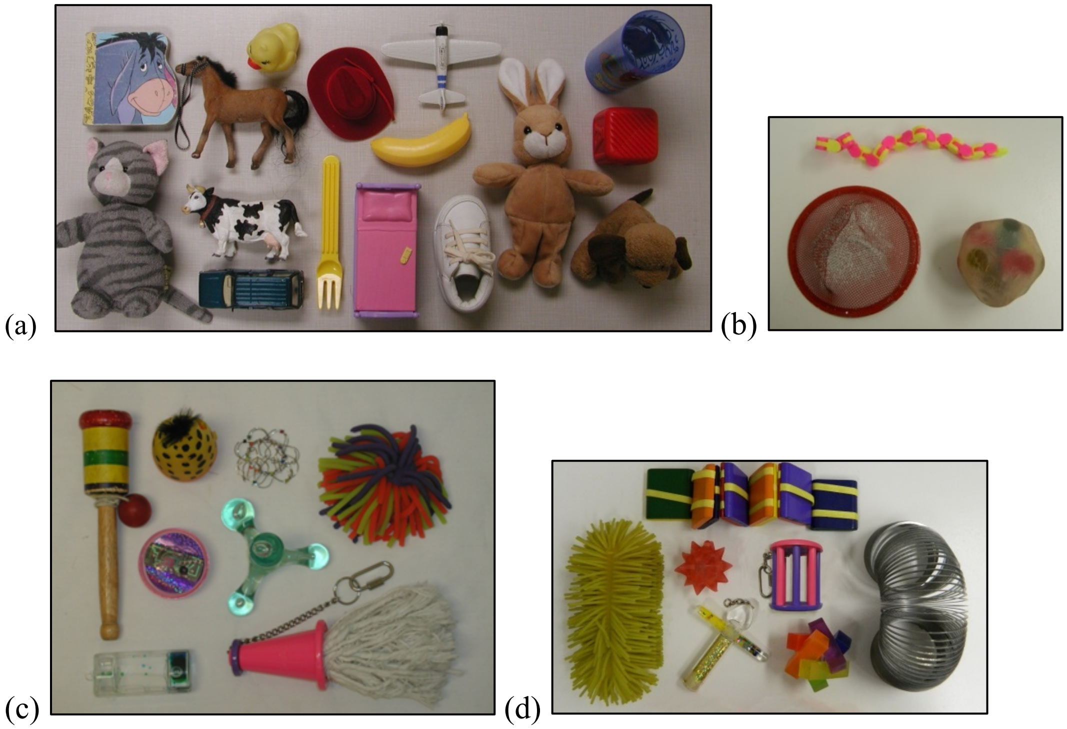

Figure 1. Stimuli. (a) Strongly known items; (b) items trained to be weakly known; (c) novel unnamed target items (d), and Experiment 3 items that were novel and not taught.

Procedure and design. The child sat across a table from the experimenter in a booster seat or on their parents' lap. Parents completed the MacArthur-Bates Communicative Development Inventory: Words and Sentences (MBCDI; Fenson et al., 1994) during the session and were instructed to avoid interacting with their child, offering minimal, neutral encouragement if needed. The child's total expressive vocabulary size was obtained from the MBCDI.

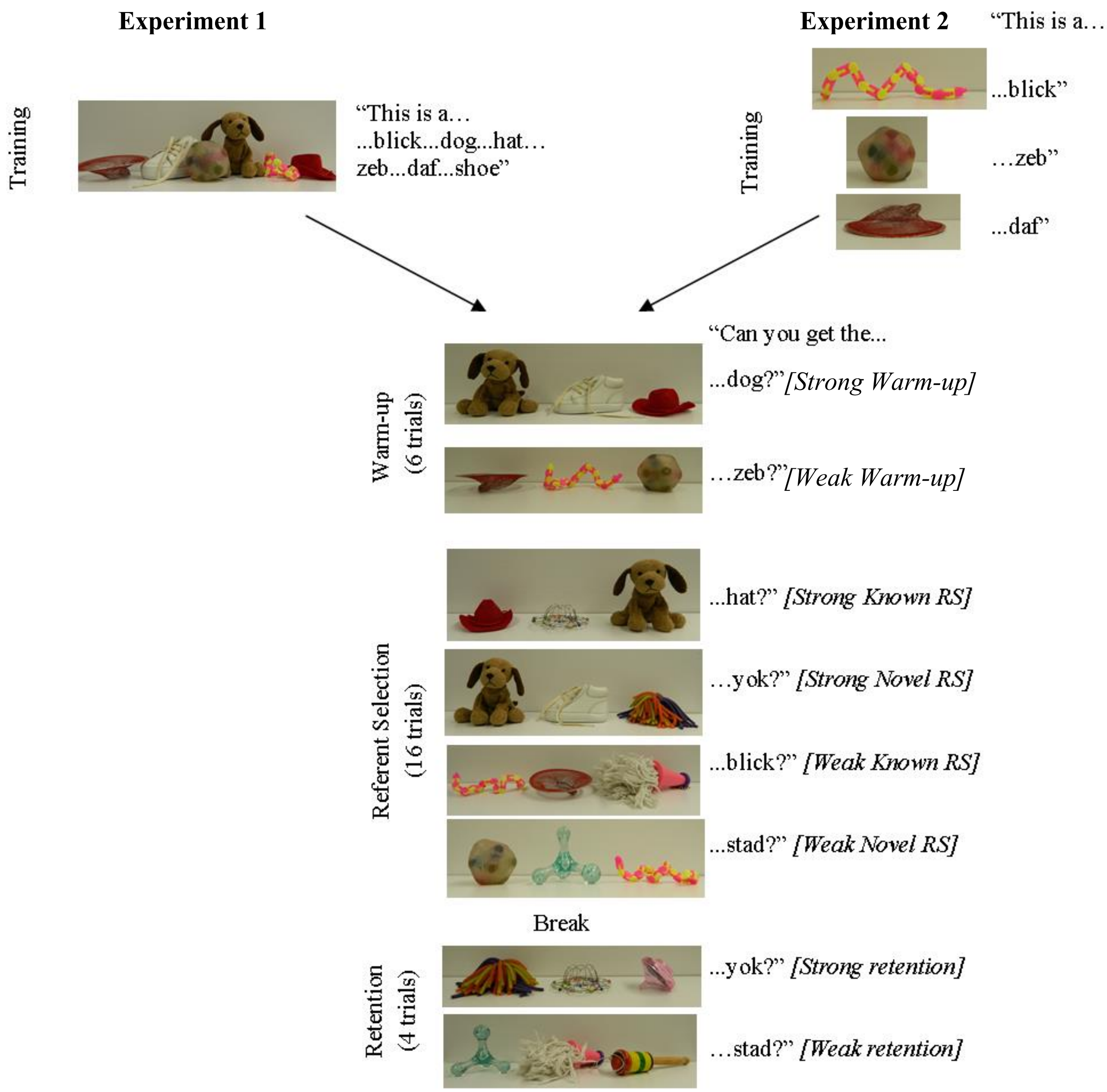

Figure 2. Procedure used in Experiments 1 and 2.

Training. The procedure (Figure 2) began with a training period designed to produce distinct differences in the strength of lexical knowledge for the two sets of known words—robust recognition and knowledge for strongly known items and limited, transient knowledge of weakly known items. Because competition aids learning (Hills et al., 2010; Zosh et al., 2013), children simultaneously explored all six "known" objects–three previously known items that would become strongly known and three to-be-named items that would become weakly known. Both strongly known and weakly known word-referent pairs were taught ostensively: the experimenter pointed to and named each item as the child explored it (Axelsson et al., 2012). If the child did not engage with an item, the experimenter drew the child's attention to it and repeated the label in a conversational context. The goal was to name each object at least three times during the one-minute training period. However, care was taken to name items when the child was attending to eliminate confusion and increase learning (Axelsson et al., 2012). Because each child's attention and item exploration varied, so did the final number of naming events. Training ended with a review – all six items were lined up out of reach of the child, held up individually, and named. Items were then removed from the table.

Warm-up/comprehension. During the six warm-up/comprehension trials, the child was familiarized with the test procedure and their comprehension for the known items was measured. On each trial, three of the items from the training period were presented in a row on a white tray. The first three trials tested the three strongly known items (e.g. *shoe, hat, dog*); the last three trials tested weakly known items (e.g. *blick, daf, zeb*). While maintaining eye contact with the child, the experimenter placed the tray on the table out of reach. The child viewed the items for three seconds. The experimenter then requested an item by name (e.g., "Can you get the *shoe*?") and pushed the tray forward, maintaining eye contact. If the incorrect item was chosen, the child was re-prompted up to three times. After the third attempt, the experimenter corrected the child. When the child chose the wrong item it was counted as an error (see Coding, below). Once the child selected the correct item they were praised and the trial ended. Target locations and objects were randomized. Each object was the target once.

Referent Selection. All sixteen referent selection trials consisted of two known items from training and one never-before-seen novel item. On half the trials, children were asked for the known item by name (“Can you get the *dog*?”); these are collectively referred to as known RS. Four of the known RS trials included two randomly selected strongly known word-referent pairs from training (e.g., *dog*) and a never-before-seen novel item; these are the strong known RS trials. The other four known RS trials included two randomly selected weakly known items from the three newly-named training items (e.g. *blick*) and a novel item; these are the weak known RS trials.

On the other eight trials, children were asked to find a novel item by name (“Can you get the *cheem*?”); these are the novel RS trials. Four novel RS trials had two strongly known items as foils and a novel item; these are the strong novel RS trials. The other four novel RS trials had two weakly known foil items with a novel object; these are the weak novel RS trials. Across trials, no novel item repeated and location was randomized. Known foil items could repeat, but the number of repetitions was balanced and the target did not repeat on consecutive trials. Order/repetition did not impact performance. Trial types alternated in a repeating pattern of strong known RS, strong novel RS, weak known RS, and weak novel RS.

Break. A five-minute break followed referent selection during which children played in the playroom or colored quietly in the experiment room. No stimuli from the experiment were present.

Retention. Following the break, children received two warm-up trials (one strongly known and one weakly known) to re-engage them. Four retention trials immediately followed – two testing retention for words initially encountered on strong novel RS trials and two for words initially encountered on Weak Novel RS trials. On each trial, the target item was presented with one foil item that was previously a target on a novel RS trial and one foil item from a known RS trial. Foil items were always from the same trial-type as the target. Thus, all three items had previously been seen and two had previously been mapped to a novel word during referent selection. The trials testing retention for a novel name initially presented in the context of strongly known foils (strong retention) alternated with two weak retention trials. No items or prompts repeated and locations were randomized.

Coding. Naïve coders recorded children's final selections off-line. Data from 21 subjects (50%) were re-coded for reliability purposes. Inter-coder agreement was 98.4% across selections. Discrepancies were settled via discussion with a third, blind, coder. To examine the influence of re-prompts during warm-up, these trials were coded for the number of incorrect choices/errors made prior to a correct selection or termination of the trial due no responses, whichever came first. If a child was deemed to not be engaged in the task (e.g. refusing to make a choice), their subsequent choices on that trial were not included in the analysis (less than 3% of trials).

Analysis. Trial-by-trial performance (correct or incorrect) on each phase of the procedure (warm-up, known RS, novel RS, retention) was analyzed separately with a series of binomial mixed models comparing condition (strongly vs. weakly known competitor items, dummy coded with strongly known as the reference), vocabulary (total vocabulary size, centered)[1], and the interaction as fixed effects. Possible random effects included random intercepts of subject, a random slope of condition (strong vs. weak) on subject, and random intercepts for items (though item was not included for phases with too few to model). The random effects structure was the maximal model justified by the data (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017), using an AIC criterion (Seedorff, Oleson & McMurray, submitted). When computing significance of the fixed effects for linear models we used a Sattertwhaite approximation for the

[1] Analyses using the child's age (in days) instead of vocabulary resulted in nearly identical outcomes in every case. Because vocabulary is known to be a strong predictor of word learning, vocabulary size is used as a fixed factor instead of age.

degrees of freedom. Performance on each trial condition in each of the phases was compared to chance (33%).

Results and Discussion

See Figure 3 and Table 1 for means and standard deviations.

Training. The number of labeling events for each of the six word-referent pairs during training was coded to assess exposure to strongly versus weakly known items. The strongly known word-referent pairs were labeled an average of 3.69 times during training (range 1-9), less than the weakly known pairs, which were labeled an average of 7.53 times (range 2-14), $t(64.28)=-9.99$, $p<.0001$, $d=2.17$ (corrected for unequal variance). There were no systematic differences across children or items.

Warm-up/comprehension. The warm-up/comprehension trials served to both familiarize children with the task and check comprehension for known words. Because children were re-prompted and/or corrected on these trials, three different variables were analyzed with separate mixed model regressions: 1) the child's initial choice, ignoring subsequent selections after re-prompting, 2) the number of errors made on each trial, and 3) accuracy of the child's final choice following re-prompts/corrections.

Children's initial accuracy on each trial was high. Responses for all trial conditions was significantly better than chance: strongly known (78.86%), $t(40)=11.41$, $p<.0001$, $d=1.78$[2]; weakly known (53.17%), $t(41)=3.78$, $p<.0001$, $d=.58$. The best fitting binomial mixed effects model included a random intercept of subject. The model revealed a significant effect of condition, $\beta=-1.18$, $z=-3.92$, $p<.0001$, with better performance for the strong than weak words.

[2] One child failed to initially respond on the strongly known trials and thus, their data for initial choice is not included.

There was no effect of vocabulary, β=.33, z=1.41, *p*=.16, nor an interaction, β=-.46, z=-1.49, *p*=.14. Thus, children appeared to have at least some knowledge of the labels for the known words, even if their learning for the weakly known items was still developing.

We next analyzed children's average number of errors during warm-up. Here, all incorrect responses were counted. Thus, a child could make zero errors (selecting correctly on the first try) or multiple errors on a single trial. The number of errors in both conditions was significantly more than zero-children were far from ceiling; strong trials (.26 errors/trial), $t(41)=4.56$, $p<.0001$, $d=.70$, weak trials (.77 errors/trial), $t(41)=11.38$, $p<.0001$, $d=1.76$. A linear mixed effects model with number of errors as the dependent variable and subject as a random intercept revealed an effect of condition, β=.51, $t(252)=5.59$, $p=<.0001$, no effect of vocabulary, β=-.087, $t(252)=-1.36$, $p=.18$, and a marginal interaction, β=.15, $t(252)=1.64$, $p=.10$.

For final selections, children performed above chance on all trials; strongly known, (100%), weakly known (94.84%), $t(41)=23.92$, $p<.0001$, $d=3.69$. The best fitting binomial mixed model included a random intercept of subject and revealed no significant effects: condition, β=-131.53, z=-.62, *p*=.54; vocabulary, β=90.72, z=.28, *p*=.78; interaction, β=-90.29, z=-.28, *p*=.78.

Children performed well on warm-up trials, selecting the target above chance initially and by the end of the trial. However, as expected, initial performance was poorer for the newly learned, weakly known words, which were also more prone to error. This is consistent with prior work reviewed previously, that newly mapped words are less likely to be encoded robustly in the lexicon and more susceptible to errors upon retrieval (Gershkoff-Stowe, 2002; Munro, Baker, McGregor, Docking, & Arculi, 2012). Given the high rates of errors, it is feasible that children were still in the process of learning words during the warm-up trials (especially weak items).

Importantly though, children were well above chance for both strongly and weakly known words, even on their initial choices.

Known Referent Selection. The critical question in the known RS trials was whether children could correctly select both strongly known and just-learned weakly known targets. Performance was surprisingly quite poor – children selected the strongly known target 32.54% of the time, not significantly above chance, $t(41)=-.15$, $p=.89$, $d=-.02$. This is particularly surprising given that the strongly known words were those which the majority of 18-month-old children are reported to know, parents had confirmed the child was familiar with these items, and the same words/referents were selected accurately during the preceding warm-up. Children selected the weakly known target 11.00% of the time – significantly *below* chance, $t(41)=-5.91$, $p<.0001$, $d=-.91$, despite having also identified these items accurately in warm-up. This was analyzed with a similar mixed model as above, with accuracy on each trial as the outcome. The best fitting model included random intercepts for subject and item and a random slope of condition on subject (no covariance term). There were no main effects of trial condition, $\beta=-1.24$, $z=-1.54$, $p=.13$, vocabulary, $\beta=.61$, $z=1.49$, $p=.14$, or interaction, $\beta=-.44$, $z=-.92$, $p=.36$, suggesting children struggled on these trials.

To understand why, we examined the types of incorrect choices made. Instead of selecting the correct known target or the known foil, children selected the novel foil item the majority of the time on both trial types: strongly known (64.48%), $t(41)=5.53$, $p<.0001$, $d=.85$, weakly known (79.96%), $t(41)=10.07$, $p<.0001$, $d=1.55$. This suggests a robust novelty bias, consistent with recent work by Kucker, McMurray, and Samuelson (2018; see also Rocha Hidalgo, et al., 2018) showing strong biases toward novel foils can overtake word knowledge, especially in 18-month-old children.

To determine if this novelty bias was affected by condition or vocabulary, we conducted a binomial mixed effects model on novelty choices across known RS trials. Fixed effects included condition and vocabulary and a random intercept for subject. This revealed a significant effect of condition, β=1.16, z=3.61, *p*=.0003, with children selecting the novel item more often on the weakly known trials. There was no effect of vocabulary, β=-.51, z=-1.56, *p*=.12 and no interaction, β=.32, z=.98, *p*=.33. Children's low performance on these weakly known trials suggests either the novelty bias seen on strong known RS trials was stronger with weakly known competitors, or children's knowledge of the weakly known targets was so fragile they were unable to bring that knowledge to bear, instead selecting based on novelty (see Horst et al., 2011; Kucker et al., 2018). Both explanations support the notion that weak knowledge for individual words results in difficult real-time processing of novel words (see also Grassmann et al., 2015; Meints et al., 1999; Merriman & Schuster, 1991).

Novel Referent Selection. The critical question for novel RS is whether children can map a novel label to the novel referent in the context of both weakly and strongly known foil items. Children were quite good. Performance in both conditions was well above chance: strongly known (98.81%) $t(41)=55.00$, $p<.0001$, $d=8.49$, weakly known (90.48%) $t(41)=25.43$, $p<.0001$, $d=3.93$. Again, this was analyzed with a binomial mixed model. The best fitting model included random intercepts for subject and item and a random slope of condition on subject (no covariance term). There was no significant effect of condition, β=-9.62, z=-1.47, *p*=.14, vocabulary, β=1.31, z=.22, *p*=.83, or interaction, β=-1.08, z=-.18, *p* =.86.

Retention. The retention trials presented a critical test of memory for novel word-referent pairs. Of particular interest were differences between novel words that were initially mapped in the context of strongly versus weakly known foil items. On the strongly known retention trials, children selected the target 34.5% of the time, at chance, $t(41)=.22$, $p=.83$, $d=.03$. Children chose the foil items at equivalent rates – other "named" item (target from Novel RS) 35.7% and unnamed item (foil from Known RS) 29.8%, $F(2,82)=.208$, $p=.813$, $\eta^2=.005$. Counterintuitively, however, children demonstrated retention of the novel word-referent pairs mapped in weakly known contexts, selecting the target significantly above chance 46.4% of the time[3], $t(41)=2.03$, $p=.05$, $d=.31$. Selection of the other named foil item was 39.3%, and the unnamed foil 14.3%. The unnamed foil was selected at significantly lower rates than either named item, $F(2, 82)=5.735$, $p=.005$, $\eta^2=.123$. To determine the effect of condition and vocabulary, we conducted a mixed effects model. The best fitting model included a random intercept for both subject and item. There was no effect of condition, $\beta=.54$, $z=1.45$, $p=.15$; vocabulary, $\beta=-.06$, $z=-.21$, $p=.83$; or interaction, $\beta=-.37$, $z=-1.00$, $p=.32$.

These findings present inconclusive evidence for the effect of condition on retention. Within the weakly learned condition, children showed above chance retention (an ability typically not seen until 30-months; Bion et al., 2013; Horst & Samuelson, 2008), and were significantly more likely to select one of the previously named items than the unnamed foil, suggesting that even if retention was not perfect, they did recall which items had names. This pattern was not seen on retention for novel words from strong RS trials. However, performance in the two conditions did not differ significantly. Thus, the relationship between referent selection and retention is perhaps more complicated than previously proposed.

[3] Only items for which children selected correctly in RS are included. However, results are nearly identical when all trials are included: overall no effect of condition, $\beta=.35$, $z=1.25$, $p=.21$, vocabulary, $\beta=.04$, $z=.13$, $p=.89$, or interaction, $\beta=-.19$, $z=-.55$, $p=.59$. Performance against chance is also similar: strongly known (34.5%) $t(41)=.22$, $p=.83$, $d=.03$, weakly known (42.9%) $t(41)=1.72$, $p=.09$, $d=.27$.

Table 1. Average percent correct for children's choice in each of the trial types. Standard deviations are noted in parentheses. Known RS for Experiment 3 did not include a novel foil item and were thus, procedurally equivalent to Warm-up Initial.

Average percent correct across all experiments and phases

	Experiment 1		Experiment 2		Combined		Exp. 3
	Strong	Weak	Strong	Weak	Strong	Weak	
Warm-up Initial	0.79 (.26)	0.53 (.34)	0.75 (.44)	0.73 (.46)	0.77 (.43)	0.63 (.48)	n/a
Warm-up Final	1.00 (0)	0.95 (.17)	1.00 (0)	1.00 (0)	1.00 (0)	0.97 (.12)	.99 (.05)
Known RS	0.33 (.35)	0.13 (.22)	0.40 (.49)	0.21 (.41)	0.36 (.48)	0.17 (.36)	.74 (.22)
Novel RS	0.99 (.08)	0.91 (.30)	0.94 (.23)	0.84 (.37)	0.97 (.17)	0.87 (.33)	.70 (.70)
Retention (mapped)	0.35 (.36)	0.46 (.42)	0.34 (.47)	0.43 (.50)	0.34 (.48)	0.45 (.50)	.38 (.24)

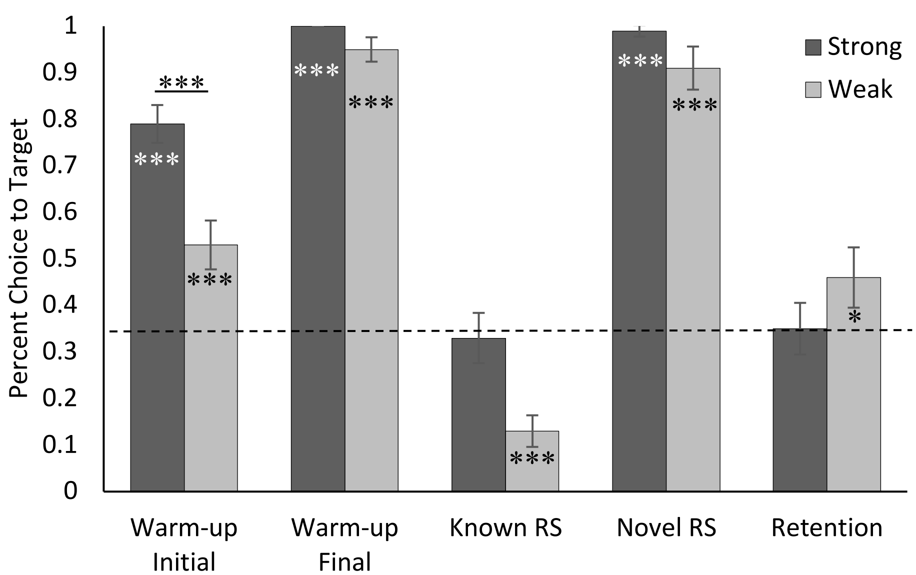

Figure 3. Performance in Experiment 1. Significance values above bars indicate comparison between strong and weak trials. Symbols within bars indicate difference from chance (33%)
Note: *<.05, ***<.001

Experiment 2

Experiment 1 demonstrated that individual word knowledge impacts recognition of those words, validating that these newly learned words were, in fact, more weakly represented. While foil word knowledge did not play a strong in-the-moment role on novel referent selection, there was some evidence it impacted retention. However, performance on the known RS trials was surprising; despite choosing the known objects correctly during warm-up, children selected the novel foil on known RS trials. This novelty bias is in line with recent findings from 18-month-old (Kucker et al., 2018) and bilingual children (Rocha Hidalgo et al., 2018). However, the current sample also included 24-month-olds. We found no interaction with vocabulary, and repeating the analyses with age instead of vocabulary revealed nearly identical results, suggesting this novelty bias extends to older children.

The findings are also surprising in the possibility of better retention of names mapped in the context of weakly known foils. However, the evidence for this was inconclusive—the conditions differed when compared to chance but not when compared against each other. Thus, the main goal of Experiment 2 was to replicate Experiment 1. Secondarily, we made a modification to training to enhance differences between familiar and newly learned words: removing the already known, strong word-object pairs from training and training each of the three weak word-object pairs sequentially without competitors present.

Methods

Participants. Forty, 18-to 24-month-old children (19 females, *M*=21 months, 17 days; range 17; 29-25;15) with a mean vocabulary of 189.4 words (range 0-673, median=141) participated. Data for three additional children were dropped due to fussiness (2) and experimenter error (1). Children received a small prize for participating and had not participated in Experiment 1. Informed consent was obtained from all participants.

Stimuli. Stimuli were identical to Experiment 1.

Procedure and Design. The procedure was identical to Experiment 1 except that during training children were only taught names for the new, weakly known items and these were trained one at a time instead of simultaneously. Each item was placed on the table, the child's attention was directed to it, and the label was given (e.g. "Look! This is a *blick*."). The child then explored the item for 30 seconds while it was labeled again. As before, items were named when children attended them.

Coding. Naïve coders again recorded children's final selections off-line. Data from 25 subjects (62.5%) were re-coded for reliability purposes. The number of labelling events during training and the warm-up trials were coded as in Experiment 1. Inter-coder agreement was 98.4% and all discrepancies were settled via discussion with a third, blind, coder.

Analysis. Linear mixed models were again run with condition, vocabulary, and the interaction as fixed effects and the maximum random effect structure justified.

Results and Discussion

Means and standard deviations are reported in Table 1 and Figure 4.

Training. Objects were labeled 6.52 times during training (range 2-15), significantly less than the weakly known words in Experiment 1, $t(80)=2.36$, $p=.021$, $d=.52$. As before, there were individual differences in children's engagement with the stimuli and naming was contingent on children's attention to the objects.

Warm-up/comprehension. Separate models were run for initial responses, number of errors, and final response accuracy. For initial responses, children performed significantly above chance on both the strongly known trials (74.79% correct), $t(38)=9.06$, $p<.0001$, $d=1.45$, and the weakly known trials (72.97% correct), $t(36)=7.84$, $p<.0001$, $d=1.29$[4]. A mixed effects model examined the effects of condition or vocabulary. The best fitting model included a random intercept of subject and showed no effects of trial condition, β=-.26, z=-.81, $p=.42$, vocabulary, β=.40, z=1.59, $p=.11$, nor an interaction, β=-.005, z=-.014, $p=.99$.

Children's total number of errors was greater on weak known trials, .81 errors per trial (range 0-2.33) compared .39 errors per strong known trial (range 0-1.67). Both trial types were significantly above zero: strong known, $t(39)=5.13$, $p<.0001$, $d=.81$; weakly known, $t(39)=8.28$, $p<.0001$, $d=1.31$. The best fitting model included random intercepts for subject and item. The number of errors was significantly predicted by both condition, β=.42, $t(12.5)=2.77$, $p=.017$, and vocabulary, β=-.15, $t(239.75)=1.91$, $p=.058$. Children with lower vocabularies made more errors. There was no interaction, β=.11, $t(237.66)=-1.02$, $p=.31$.

Children's ultimate accuracy was 100% for all trials. Thus, like Experiment 1, children had some knowledge of both the strongly known and weakly known items, but made more initial errors on the weakly known trials. Further, the change in training still resulted in sufficient knowledge of the weakly known mappings.

Known RS. Much like Experiment 1, children performed poorly on both strongly and weakly known RS trials, selecting strongly known targets at chance (39.58% of the time), $t(39)=1.03$, $p=.31$, $d=-.098$, and weakly known targets *below* chance (20.83%), $t(39)=-3.44$, $p=.001$, $d=-.54$. The best fitting model included random intercepts for subject and item and a random slope of condition on subject (no covariance term). This model revealed no effect of

[4] Some children in each condition did not make initial choices on the warm-up trials.

condition, β=-1.24, z=-1.53, *p*=.125, vocabulary, β=.61, z=1.49, *p*=.13, or interaction, β=-.44, z=-.92, *p*=.36.

As in Experiment 1, instead of correctly choosing the target item, children chose the novel foil item the majority of the time; on both strongly known trials (56.67%), *t*(39)=3.94, *p*<.0001, *d*=.62, weakly known trials (65.4%), *t*(39)=7.16, *p*<.0001, *d*=1.13. Performance across all trials was analyzed in mixed effects model with choice of the novel foil as the outcome. The best fitting model had a random intercept of subject, revealing a significant effect of condition, β=1.16, z=3.61, *p*=.0003, with higher novelty biases on weakly known trials. Vocabulary did not predict this novelty bias, β=-.51, z=-1.56, *p*=.12, and did not interact with condition, β=.32, z=.98, *p*=.33.

Novel RS. Children selected the target well above chance levels on both trials with strongly known foils (94.38%), *t*(39)=32.19, *p*<.0001, *d*=5.09, and weakly known foils (84.38%), *t*(39)=15.41, *p*<.0001, *d*=2.44. The best fitting model included random intercepts for both subject and item, revealing a significant main effect of condition, β=-1.19, z=-2.70, *p*=.007, with better performance on the strongly known trials. There was no effect of vocabulary, β=-.18, z=-.53, *p*=.60, but there was a marginal interaction, β=.72, z=1.67, *p*=.09; children performed best on the strongly known trials when vocabulary was high.

Retention. Like Experiment 1, children performed at chance with novel words that were mapped in the context of strongly known foils, selecting the target only 33.75% of the time, *t*(39)=.072, *p*=.94, *d*=.013. They selected the other named foil item 38.75% of the time, and the unnamed foil 27.5% of the time – these three items were not selected at different rates, F(2,78)=.623, *p*=.439, η^2=.016. While retention of words mapped in a weakly known context was numerically greater

(43.2%), this was not significantly different from chance, $t(36)=1.53$, $p=.14$, $d=.25$[5]. On weak retention trials, children selected the other named foil item and the unnamed foil both 28.38% of the time. The best fitting model for retention included a random intercept of subject. As in Experiment 1, there was no effect of condition, $\beta=.54$, $z=1.53$, $p=.13$, vocabulary, $\beta=.08$, $z=.30$, $p=.76$, or interaction, $\beta=-.09$, $z=-.25$, $p=.80$[6].

---

[5] Only trials in which the child had initially selected correctly during RS were included. Similar results are found if all trials are included - strongly known trials at chance, (33.8%) $t(39)=.076$, $p=.94$, $d=.014$, weakly known above chance, (46.3%) $t(39)=2.24$, $p=.03$, $d=.355$.

[6] The results hold when all children and all trials are included, with condition as a marginal effect, $\beta=.58$, $z=1.75$ $p=.08$, no effect of vocabulary, $\beta=.04$, $z=.18$, $p=.86$, and no interaction, $\beta=.10$, $z=.30$, $p=.76$.

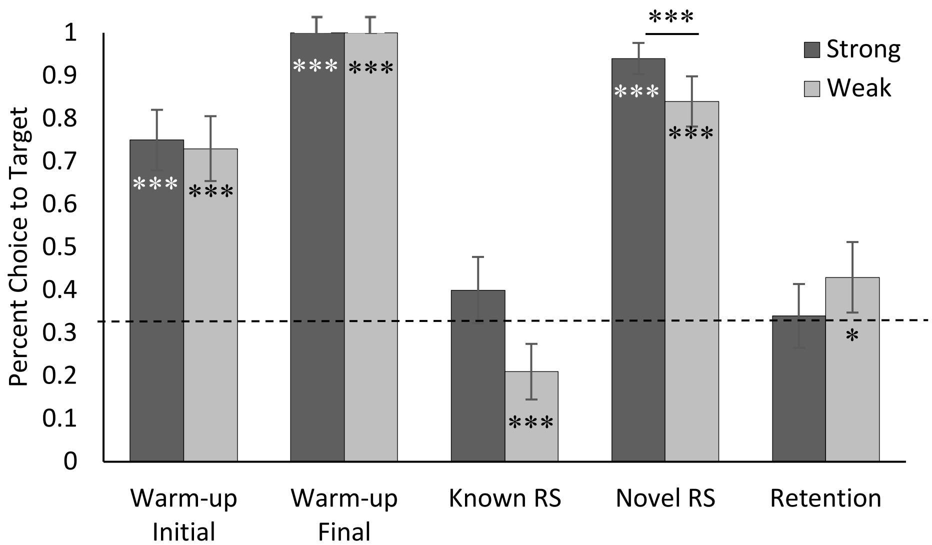

Figure 4. Performance in Experiment 2. Significance values above bars indicate comparison of strong and weak trials; within bars indicate difference from chance (33%). Note: *<.05, **<.01, ***<.001

Combined Analysis and Discussion

The results of Experiments 1 and 2 raise the possibility that children are more likely to demonstrate retention of novel word-object mappings made in the context of weakly known foils. However, results were mixed. In Experiment 1, retention was significantly above chance for weakly known foils and below chance for strongly known foils, though these did not differ significantly. In Experiment 2 neither contrast was significant, though performance was numerically almost identical to Experiment 1.

The sample size of both experiments was based on prior work examining referent selection and retention with well-known items. This prior work found large effect sizes, but had not examined individual word knowledge. The retention results of the current study show small effects (e.g. weakly known retention showed *d*=.31 in Experiment 1 and *d*=.25 in Experiment 2). This raises the possibility that each experiment (individually) was underpowered.

Nevertheless, both experiments showed similar patterns of performance at all stages and tested an identical sample of children. Thus a combined analysis of the referent selection and retention phases of both experiments was conducted. Binomial linear mixed models with maximum justified random effect structure were run as before. Experiment was added as a fixed effect and interaction. Only marginal/significant main effects/interactions are reported.

Known RS

Combined performance on the strongly known RS trials was at chance (35.98%) $t(81)=.651$, $p=.52$, $d=.072$, and below chance for the weakly known (16.77%), $t(81)=-6.55$, $p<.001$, $d=-.72$. The best fitting model included random intercepts for subject and item and a random slope of condition on subject (no covariance term). There was a marginally significant effect of condition, β=-.876, z=-1.67, $p=.09$, and experiment, β=1.01, z=1.68, $p=.092$, and a significant effect of vocabulary, β=.92, z=2.92, $p=.004$. Children performed better on strongly known RS trials, when they had a higher vocabulary, and in Experiment 2. A model of novel foil item selection (random effects were identical to the known RS model) revealed a novelty bias that was negatively correlated with vocabulary, β=-.86, z=-2.95, $p=.003$, and marginally higher in Experiment 1, β=-.99, z=-1.78, $p=.08$, with no effect of condition, β=.55, z=1.26, $p=.21$.

Novel RS

Children were well above chance on all trials; strongly known (96.65%), $t(82)=56.12$, $p<.0001$, $d=6.20$, weakly known (87.5%), $t(81)=27.09$, $p<.0001$, $d=2.99$. The best fitting model included random intercepts for subject and item and a random slope of trial condition on subject (no covariance term). There was a marginal effect of condition, β=-4.51, z=-1.91, $p=.056$, with children performing best on the strong novel RS trials. There was no effect of vocabulary, β=.43, z=.33, $p=.74$, or experiment, β=-2.54, z=-1.20, $p=.23$.

Retention

Retention of words mapped in a strongly known context was at chance levels, (34.1%) $t(82)=.21$, $p=.84$, $d=.021$. However, retention of words learned in weak contexts was significantly above chance, (44.9%) $t(78)=2.55$, $p=.013$, $d=.29$.[7] The best fitting model for retention included random intercepts for subject and item. There was a significant effect of condition, $\beta=.50$, $z=1.93$, $p=.05$, with children retaining more words learned in weakly known contexts.

Discussion

These combined analyses buttress the findings of the individual experiments: the strength of children's knowledge of individual foil items when making new word-object mappings matters both for their mapping and later retention. Children did better at selecting strongly known items during referent selection, although they demonstrated a bias to select novel items on *all* referent selection trials. Children demonstrated the highest levels of retention for words that were initially mapped in the context of weakly-known foils.

These results are surprising, first, because they show a counter-intuitive effect of foil knowledge. From the standpoint of mutual exclusivity, stronger name-object links for familiar items being rejected as possible referents should help, not hinder, encoding of new mappings (Merriman & Bowman, 1989) and subsequent retention. The null effect of children's overall vocabulary size on retention was also surprising; even younger children with smaller vocabularies demonstrated retention of mappings formed in the context of weakly known foils. This contrasts with prior work suggesting that young children do not often retain mappings

[7] The results hold when all trials (not just those which children mapped) are included; condition, $\beta=.48$, $z=1.97$, $p=.049$, vocabulary, $\beta=.08$, $z=.43$, $p=.66$, experiment, $\beta=.04$, $z=.11$ $p=.91$; strongly known against chance, (34.1%) $t(82)=.21$, $p=.83$, $d=.022$, weakly known against chance, (44.5%) $t(82)=2.81$, $p=.006$, $d=.31$.

formed in referent selection unless they are given additional support (Kucker & Samuelson, 2012; Mervis & Bertrand, 1994).

These results also highlight the dissociation of referent selection and retention (Kucker, et al., 2016); good novel RS performance does not necessarily translate to good retention. The present work adds a new wrinkle, however, in that slightly less-good referent selection performance *does* lead to slightly higher retention in the case of weakly known foils. That is, factors that support selection are not necessarily the same factors that support retention.

Interestingly, the most complete explanation for the current findings is that relative novelty, not knowledge, underpinned selections. The novel target object was always more novel than the familiarized foils. Thus, selection driven by novelty would always result in high target choices during novel RS, as we found. The fact that children also selected the novel item when it was a foil in *known* RS also fits this idea. These data thus coincide with prior work showing a role for novelty in referent selection (Horst et al., 2011; Kucker et al., 2018; Mervis & Bertrand, 1994). However, while more novelty is known to support selection, *retention* from novelty-based referent selection is less well documented. The current results reveal above-chance retention only for novel items correctly selected during the weakly known RS trials in which the foil items were relatively novel (having just been learned). Thus, higher novelty of the *competing foil items* (i.e. contextual novelty) during initial word-referent mapping may cascade to boost retention. A critical question, then, is if it is *only* the novelty of the foil items that boosts retention, or if those items must also have a name. We test this in Experiment 3.

Experiment 3

Children in Experiments 1 and 2 retained new words mapped with weakly known competitors. This suggests that increased novelty of competing foils during referent selection supports learning. Though a small body of work has looked at the role of novel foils in referent selection (Horst et al., 2011), none have measured retention. Moreover, it is unknown if novelty alone boosts retention or if a weakly learned name is required. The task used here is similar to that in Experiments 1 and 2, but the "known" foil items were seen only minimally prior to the RS trials and never named. Thus, these novel RS trials ask children to map a name to a novel object in the context of foils that are even less well known, and thus more novel, than the weakly-known foils in the prior experiments.

Methods

Participants. Forty-three, 18- to 24-month-old children (19 females, *M*=22 months, 15 days; range 17:22-25:20) with a mean vocabulary of 242.6 words (range 1-657, median=243) participated. Data from two additional children were dropped due to fussiness (1) and experimenter error (1). Children received a small prize for participating and did not participate in Experiment 1 or 2. Informed consent was obtained from all participants.

Stimuli. The same pool of stimuli and novel names were used (Figure 1).

Procedure and design. As before, children were seated across a table from the experimenter in a booster seat or on their parents' lap. Parents completed the MBCDI during the session and were instructed to avoid interacting with their child. The procedure was based on that used by Horst and colleagues (2011), but with an additional retention test (Figure 5).

Familiarization. During the short familiarization period, children were presented with eight novel items in two sets of four. Children explored each set for approximately one minute. If an item was not explored, the experimenter pointed and said "look". No names were given. After the child had been familiarized with each item and one minute had passed, they were removed and the child was given the second set to explore. These eight items were used as foils during the novel referent selection trials.

Warm-up. A subset of six known items was randomly selected from the pool of known items. This subset was used throughout the rest of the procedure. Warm-up proceeded identically to Experiments 1 and 2.

Referent selection. On eight of the 16 RS trials, children were presented with three known items randomly drawn from the subset of six from warm-up and asked to find a known item by name. No novel items were present. Because there were only six possible known items, items could repeat over the course of referent selection but the target was never the same for two consecutive trials, and the same set of three were never presented back-to-back.

On the other eight alternating novel RS trials, children were presented with three unnamed novel items. Two of these items were objects from the familiarization phase. The third object was completely novel; known as ‘supernovel’. Thus, items on the novel RS trials only differed on the amount of familiarity/relative novelty and none had labels. On these novel RS trials children were asked to find a novel item by name. The unnamed, familiarized items were each repeated twice as foils – once in the first block of eight referent selection trials and once in the second block. There was no effect of trial order. Supernovel items never repeated and location was randomized. No correction or praise was given, but children were prompted up to three times if necessary.

Break. Referent selection was followed by a five-minute break.

Retention. A single warm-up trial and four retention trials followed the break. On each, two novel items that had previously been the targets on novel RS trials were present along with a third item that had served as a familiarized novel foil item. Items and targets were not repeated.

Coding. Naïve coders blind to the hypothesis coded children's final selections off-line. Data from 22 subjects (51.1%) were recoded for reliability purposes. Inter-coder agreement was 96.72%. All discrepancies were settled via discussion with a third, blind, coder.

Analysis. As in Experiments 1 and 2, trial-by-trial performance (correct or incorrect) on each phase of the procedure was analyzed with a binomial mixed model with vocabulary as a fixed effect. The maximum random effect structure justified in each phase included a random intercept of subject. Overall performance was also compared to chance (33%). See Figure 6 and Table 1 for means and standard deviations.

Results and Discussion

The warm-up trials were coded for the number of errors children made as well as their final choices. Children made an average of .52 errors per trial, significantly more than zero, $t(42)=6.98$, $p<.0001$, $d=1.06$. Vocabulary negatively predicted the number of errors, $\beta=-.24$, $z=-3.84$, $p<.0001$. However, children ultimately chose the target item 98.84% of the time regardless of vocabulary size, $\beta=1.83$, $z=1.16$, $p=.25$, above chance, $t(42)=80.63$, $p<.0001$, $d=12.30$.

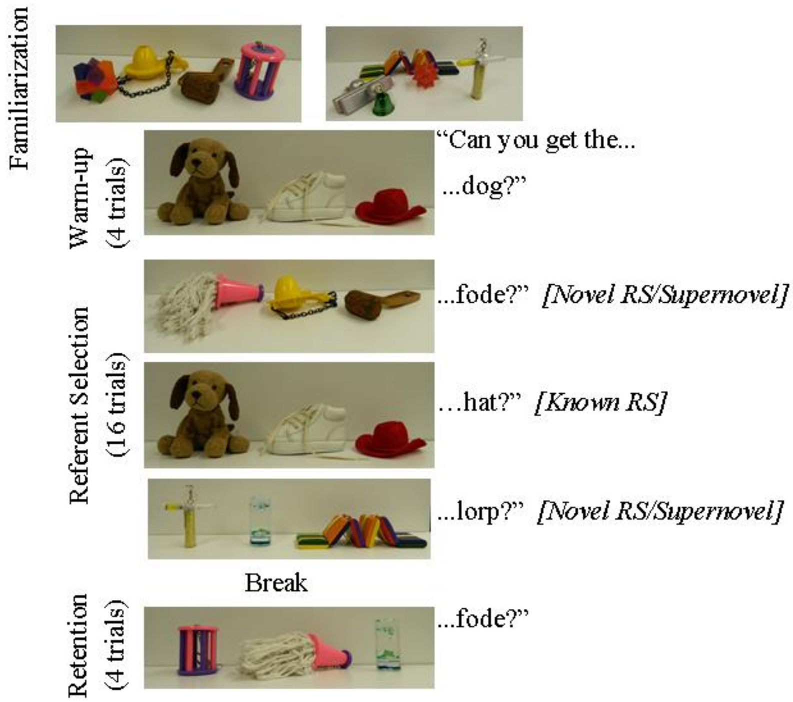

Figure 5. Procedure of Experiment 3.

On the known RS trials there was a significant effect of vocabulary, β=.57, *z*=2.85, *p*=.004. Overall, children performed well above chance, selecting the target 74.12% of the time, *t*(42)=12.41, *p*<.001, *d*=1.89. This was higher than performance on strong known RS trials in Experiments 1 and 2. There was no novel foil item here, suggesting a high novelty bias, rather than not knowing the referent, is the cause of children's failure on the known RS of the prior experiments (see also Kucker et al., 2018).

As in Experiments 1 and 2, performance on the novel RS trials was also high. Children selected the target above chance 69.71% of the time, *t*(42)=13.08, *p*<.0001, *d*=1.99. There was no effect of vocabulary, β=-.15, *z*=-1.23, *p*=.22.

If retention for words mapped in the context of weakly versus strongly known foils in Experiments 1 and 2 was due to the weakly known foils being more novel, we should expect to see above chance retention here. That was not the case. Children selected the target at chance levels, only 38.00% of the time, $t(40)=1.07$, $p=.29$, $d=.17$[8], similar to strong retention trials in Experiments 1 and 2. There was no effect of vocabulary, $\beta=.23$, $z=1.16$, $p=.24$. Thus, while relative novelty of possible referents may drive in-the-moment behavior, novelty alone may not support retention. Interestingly, this finding suggests a possible ‘goldilocks effect’ (Kidd, Piantadosi, & Aslin, 2012): some knowledge of the names of the foils is needed to support mapping and retention but too much knowledge of foils works against retention.

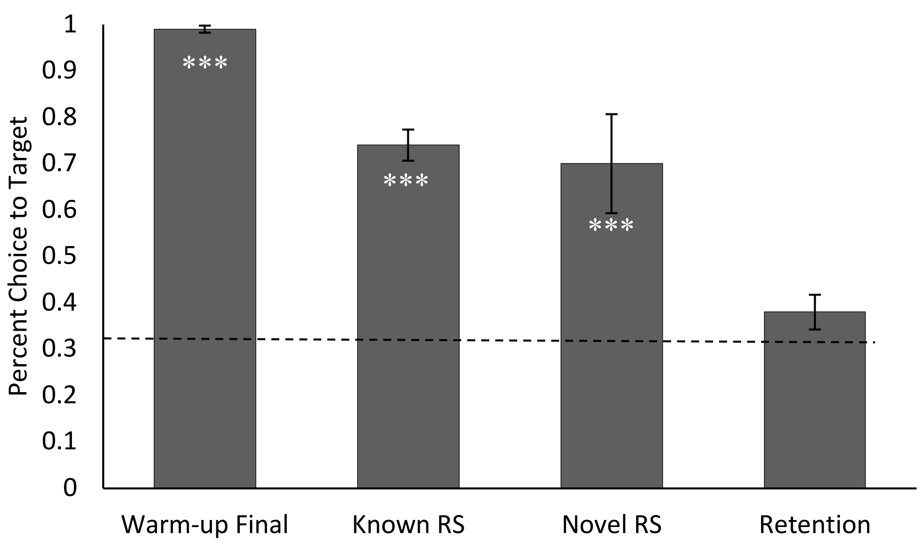

Figure 6. Results of Experiment 3. Note, ***p<.001 represents above chance performance (33%; dotted line).

General Discussion and Conclusions

The current study adds a surprising twist to the literature on early word learning: while *both* strong and weak foils support the novel referent selection, strong prior knowledge might not lead to retention. Experiment 1 revealed that children’s ability to select a novel object as the

[8] When all trials are included in the analysis, not just those which children correctly chose in RS, results are similar (37.79%), $t(42)=1.20$, $p=.24$, $d=.18$, but with a significant effect of vocabulary, $\beta=.34$, $z=2.09$, $p=.037$.

referent for a novel word was reliably high for both strongly and weakly known foils. Critically, this suggests that classic referent selection constraints only capture part of the story – children map novel words even when knowledge is not robust. Moreover, though retention results were mixed, only novel words mapped in the context of weakly known foils were retained at above chance levels—a finding replicated in Experiment 2. Experiment 3 found that novelty alone was not able to support retention. Overall, the findings stand in partial contrast to prior work and classic approaches to word learning such as mutual exclusivity (Markman & Wachtel, 1988), and N3C (Mervis & Bertrand, 1994), that suggest robust knowledge of previously-learned word-object mappings is used to build new lexical knowledge. While the current results do not rule out the use of mutual exclusivity or similar constraints, they do suggest that learning *and* retention might be best facilitated by *weaker* knowledge of individual foils – a possibility prior work has not tested.

We did find some evidence for an influence of vocabulary size on warm-up and known RS in the combined analysis and Experiment 3. However, unlike prior work (Bion et al., 2013), vocabulary size did not predict retention. These prior studies, however, tested only well-known foil items. This suggest that while in-the-moment processes underlying children's word learning are driven by both the mapping context and vocabulary, the cascading effects of what is learned in-the-moment may be more closely related to the context itself drawing on *specific knowledge* of the words and objects that are relevant to that naming event.

The current results may be usefully viewed in light of a small, but growing body of work in language suggesting that challenging initial contexts may result in better long-term outcomes (Storkel & Adlof, 2009; Vlach & Sandhofer, 2014). In the current study, disambiguating a novel object from an array of weakly-known items could reasonably be assumed to be harder than selecting from an array of well-known items. Thus, children had to do more “work” to be successful in finding the novel referent in the context of weaker knowledge. It has been proposed that real time processing of difficult material may be deeper, which can cascade to benefit learning on a longer timescale (Vlach & Sandhofer, 2014; and see McMurray, et al., 2012). Thus, the strength of a child’s individual word knowledge leads to subtle changes during referent selection, which then lead to larger differences in learning.

In a similar vein, both the weakly known items (Experiment 1 and 2) and the familiarized novel objects (Experiment 3) were items with presumably weaker semantic networks. It is feasible that the strong semantic network of well-known items interfered with encoding of the novel labels, whereas less dense networks allowed new words to be more easily integrated into the network space (Zhao, Packard, McMurray, & Gupta, in press; but see McClelland, 2013). Similarly, it is feasible the networks of the well-known items were strongly tied to contexts outside of the lab, making them relatively more novel in the current task (Hills et al., 2009b). The lack of retention in Experiment 3, however, suggests that some knowledge of competitor’s label is important. In particular, the lack of a lexical network might have impacted the lexical connections built during referent selection. If word learning is a competitive process in which strengthening a connection between one word and its referent means simultaneously pruning the spurious connections between that word and other foil referents (McMurray et al., 2012), then the absence of a name for the foils might mean less pruning was done. Applied to all the data, then, the proposal would be that the best learning results from a context that balances the challenge of selecting the right referent with enough knowledge on the child’s part to maximize changes to the lexical network.

In a broader context, the current work is in line with other recent studies that highlight how the dynamic, time-extended nature of the word learning process is both robust and surprising. In particular, the results confirm that learning can occur in very ambiguous cases (e.g. Smith & Yu, 2008), and highly variable (Perry et al., 2010), uncertain contexts (Zosh et al., 2013) can produce positive learning outcomes. Our findings may also be usefully viewed from the perspective of computational models that suggest a lack of competition during initial encoding may help initial referent selection, but such direct, easy training, is detrimental to long-term learning (McMurray, Zhao, Kucker, & Samuelson, 2013). Such models, along with prior work (Fitneva & Christiansen, 2011; Vlach & Sandhofer, 2014) demonstrate how contexts that make disambiguation difficult might lead to lower rates of initial success (see also Meints et al., 1999; Merriman & Schuster, 1991), but can still lead to strong learning (Kucker et al., 2015). In the current study, children demonstrated only minimal differences in novel referent selection performance across conditions. However, there appear to be subtle differences in the mapping specifics that cascade to differences in retention.

Acknowledgements

This research was supported by R01HD045713 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development to LKS. The content is solely the responsibility of the authors. We wish to thank the members of the Language and Categorization Lab, John Spencer, Prahlad Gupta, and Karla McGregor for helpful discussions. This research was part of the first author's doctoral dissertation submitted to the University of Iowa.

References

Alt, M., & Plante, E. (2006). Factors that influence lexical and semantic fast mapping of young children with specific language impairment. *Journal of Speech, Language and Hearing Research*, *49*(5), 941-954. doi: 10.1044/1092-4388(2006/068)

Axelsson, E.L., Churchley, K., & Horst, J.S. (2012). The right thing at the right time: Why ostensive naming facilitates word learning. *Frontiers in Psychology*, *3,* 1-8. https://doi.org/10.3389/fpsyg.2012.00088

Baldwin, D.A., & Markman, E.M. (1989). Establishing word object relations: A first step. *Child Development*, *60*, 381–398. doi: 10.2307/1130984

Bion, R.A.H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, *126*(1), 39–53. https://doi.org/10.1016/j.cognition.2012.08.008

Byers-Heinlein, K., & Werker, J.F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition*, *128*(3), 407–416. https://doi.org/10.1016/j.cognition.2013.05.010

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, *15*, 17–29.

Carey, S. (2010). Beyond fast mapping. *Language Learning and Development, 6*, 184-205. doi: 10.1080/15475441.2010.484379

Dale, P. S., & Fenson, L. (1993). *LEX: A lexical development norms database*. Seattle, WA: University of Washington, Department of Psychology

Fenson, L., Philip, P.S., Reznick, J., Bates, E., Thal, D.J., & Pethick, S.J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5), i–185. doi: 10.2307/1166093

Fernald, A., Pinto, J.P., Swingley, D., Weinberg, A., & McRoberts, G.W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228–231. https://doi.org/10.1111/1467-9280.00044

Fitneva, S.A., & Christiansen, M.H. (2011). Looking in the wrong direction correlates with more accurate word learning. *Cognitive Science*, *35*(2), 367–380. https://doi.org/10.1111/j.1551-6709.2010.01156.x

Gershkoff-Stowe, L. (2002). Object naming, vocabulary growth, and the development of word retrieval abilities. *Journal of Memory and Language*, *46*(4), 665–687. https://doi.org/10.1006/jmla.2001.2830

Grassmann, S., Schulze, C., & Tomasello, M. (2015). Children's level of word knowledge predicts their exclusion of familiar objects as referents of novel words. *Frontiers in Psychology*, *6*, 1–8. https://doi.org/10.3389/fpsyg.2015.01200

Gurteen, P.M., Horne, P.J., & Erjavec, M. (2011). Rapid word learning in 13- and 17-month-olds in a naturalistic two-word procedure: Looking versus reaching measures. *Journal of Experimental Child Psychology*, *109*(2), 201–217. https://doi.org/10.1016/j.jecp.2010.12.001

Hart, B., & Risley, T.R. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore: Paul H Brookes Publishing.

Hills, T.T., Maouene, J., Riordan, B., & Smith, L.B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*(3), 259–273. https://doi.org/10.1016/j.jml.2010.06.002

Hills, T.T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009a). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological Science*, *20*(6), 729–739. https://doi.org/10.1111/j.1467-9280.2009.02365.x

Hills, T.T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009b). Categorical structure among shared features in networks of early-learned nouns. *Cognition*, *112*(3), 381–396. https://doi.org/10.1016/j.cognition.2009.06.002

Horst, J.S., Scott, E.J., & Pollard, J.A. (2010). The role of competition in word learning via referent selection. *Developmental Science*, *13*(5), 706–713. https://doi.org/10.1111/j.1467-7687.2009.00926.x

Horst, J.S., & Hout, M.C. (2014). The Novel Object and Unusual Name (NOUN) database: A collection of novel images for use in experimental research. *Behavioral Research Methods, 48*(4), 1393-1409. doi: 10.3758/s13428-015-0647-3

Horst, J.S., & Samuelson, L.K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–157. https://doi.org/10.1080/15250000701795598

Horst, J.S., Samuelson, L.K., Kucker, S.C., & McMurray, B. (2011). What's new? Children prefer novelty in referent selection. *Cognition*, *118*(2), 234–244. https://doi.org/10.1016/j.cognition.2010.10.015

Kalashnikova, M., Escudero, P., & Kidd, E. (2018). The development of fast-mapping and novel word retention strategies in monolingual and bilingual infants. *Developmental Science, 21*, 1-11. doi: 10.1111/desc.12674

Kalashnikova, M., Mattock, K., & Monaghan, P. (2016). Mutual exclusivity develops as a consequence of abstract rather than particular vocabulary knowledge. *First Language*, *36*(5), 451–464. https://doi.org/10.1177/0142723716648850

Kidd, C., Piantadosi, S.T., & Aslin, R.N. (2012). The Goldilocks Effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, *7*(5), e36399. https://doi.org/10.1371/journal.pone.0036399

Kucker, S.C., McMurray, B., & Samuelson, L.K. (2015). Slowing down fast mapping: Redefining the dynamics of word learning. *Child Development Perspectives*, *9*(2), 74–78. https://doi.org/10.1111/cdep.12110

Kucker, S.C, McMurray, B., & Samuelson, L. K. (2018). Too much of a good thing: How novelty biases and vocabulary influence familiar and novel referent selection and retention in 18-month-old children and associative models. *Cognitive Science, 42*(Supp 2). 463-493. doi: 10.1111/cogs.12610

Kucker, S.C., & Samuelson, L.K. (2012). The first slow step: Differential effects of object and word-form familiarization on retention of fast-mapped words. *Infancy*, *17*(3), 295–323. https://doi.org/10.1111/j.1532-7078.2011.00081.x

MacWhinney, B. (2000). *Child Language Data Exchange System Project*. Mahwah, NJ: Lawrence Erlbaum.

Mareschal, D., & Tan, S.H. (2007). Flexible and context-dependent categorization by eighteen-month-olds. *Child Development*, *78*(1), 19–37. https://doi.org/10.1111/j.1467-8624.2007.00983.x

Markman, E.M., & Wachtel, G.F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

Mather, E., & Plunkett, K. (2010). Mutual exclusivity and phonological novelty constrain word learning at 16 months. *Journal of Child Language*, *38*(05), 933–950. https://doi.org/10.1017/S0305000910000401

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McGregor, K.K. (2014). What a difference a day makes: Change in memory for newly learned word forms over Twenty-four hours. *Journal of Speech, Language, and Hearing Research, 57*(5), 1842-1850. doi: 10.1044/2014_JSLHR-L-13-0273

McMurray, B, Zhao, L., Kucker, S.C., & Samuelson, L.K. (2013). Pushing the envelope of associative learning: Internal representations and dynamic competition transform association into development. In L. Gogate & G. Hollich (Eds.). In *Theoretical and computational models of word learning: Trends in psychology and artificial intelligence* (pp. 49–80). Hershey, PA: IGI Global.

McMurray, B. (2007) Defusing the childhood vocabulary explosion. *Science, 317*(5838), 631. doi: 10.1126/science.1144073

McMurray, B., Horst, J.S., & Samuelson, L.K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. https://doi.org/10.1037/a0029872

Meints, K., Plunkett, K., & Harris, P.L. (1999). When does and ostrich become a bird? The role of typicality in early word comprehension. *Developmental Psychology*, *35*(4), 1072-1078. doi: 10.1037//0012-1649.35.4.1072

Merriman, W.E., & Schuster, J.M. (1991). Young children's disambiguation of object name reference. *Child Development*, *62*(6), 1288–1301. https://www.jstor.org/stable/1130807

Merriman, W.E., & Bowman, L.L. (1989). The Mutual Exclusivity Bias in children's word learning. *Monographs of the Society for Research in Child Development*, *54*(3/4), i. https://doi.org/10.2307/1166130

Mervis, C.B., & Bertrand, J. (1994). Acquisition of the novel name–nameless category (N3C) principle. *Child Development*, *65*(6), 1646–1662. https://www.jstor.org/stable/1131285

Munro, N., Baker, E., McGregor, K.K., Docking, K., & Arculi, J. (2012). Why word learning is not fast. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00041

Nelson, D.L., Zhang, N., & McKinney, V.M. (2001). The ties that bind what is known to the recognition of what is new. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1147-1159. doi: 10.1037/0278-7393.27.5.1147

Perry, L.K., Samuelson, L.K., Malloy, L.M., & Schiffer, R.N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, *21*(12), 1894–1902. https://doi.org/10.1177/0956797610389189

Rocha Hidalgo, J., Barr, R., Blanchfield, O., Rusnak, S., Feller, M., & Lozano, G. (October, 2018). Using the referent selection task to test mutual exclusivity in monolingual and bilingual 2-year-olds. Poster presented at the International Society of Developmental Psychobiology, San Diego, CA.

Seedorff, M., Oleson, J., & McMurray, B. (2019). Maybe maximal: Good enough mixed models optimize power while controlling Type I error. Preprint available at https://psyarxiv.com/xmhfr/

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Storkel, H.L. (2009). Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *Journal of Child Language*, *36*(02), 291. https://doi.org/10.1017/S030500090800891X

Storkel, H.L., & Adlof, S.M. (2009). The effect of semantic set size on word learning by preschool children. *Journal of Speech, Language, and Hearing Research*, *52*(2), 306–320. doi: 10.1044/1092-4388(2009/07-0175)

Vlach, H.A., & Sandhofer, C.M. (2014). Retrieval dynamics and retention in cross-situational statistical word learning. *Cognitive Science*, *38*(4), 757–774. https://doi.org/10.1111/cogs.12092

Wagar, B.M., & Dixon, M.J. (2005). Past experience influences object representation in working memory. *Brain and Cognition*, *57*(3), 248–256. https://doi.org/10.1016/j.bandc.2004.08.054

Zosh, J.M., Brinster, M., & Halberda, J. (2013). Optimal contrast: Competition between two referents improves word learning. *Applied Developmental Science*, *17*(1), 20–28. https://doi.org/10.1080/10888691.2013.748420