Toddlers Always Get the Last Word: Recency biases in early verbal behavior

Emily Sumner[⋆] (esumner@u.rochester.edu)
Erika DeAngelis[⋆] (erikadeangelis97@gmail.com)
Mara Hyatt[⋆] (hyattm@bc.edu)
Noah Goodman[◇] (ngoodman@stanford.edu)
Celeste Kidd[⋆,×] (ckidd@bcs.rochester.edu)

[⋆] Department of Brain & Cognitive Sciences, University of Rochester, Meliora Hall, Rochester, NY 14627-0268
[◇] Department of Psychology, Stanford University, Jordan Hall, Stanford, CA, 94305
[×] Center for Visual Science, University of Rochester, Meliora Hall, Rochester, NY 14627-0270

Abstract

A popular conception about language development is that comprehension precedes production. Although this is certainly true during the earliest stages of phonological development, once a child possesses the basic articulatory skills required for imitation, it need not necessarily be the case. A child could produce a word without possessing the fully formed lexical representation through imitation. In some cases, such as in response to questions containing fixed choices, this behavior could be mistaken for a deeper understanding of the words' semantic content. In this paper, we present evidence that 2- to 3-year-old children exhibit a robust recency bias when verbally responding to two-alternative choice questions (i.e., they select the second, most recently mentioned option), possibly due to the availability of the second word in phonological memory. We find further evidence of this effect outside of a laboratory setting in naturalistic conversational contexts in CHILDES (MacWhinney, 2000), a large corpus of transcribed child-adult interactions.

Keywords: Decision making; cognitive development; developmental experimentation; language acquisition; learning.

Comprehension Need Not Precede Production

Modern studies of language development commonly use measurements of children's early lexical productions as a proxy for their lexical knowledge. This practice seems reasonable given the common view that language comprehension precedes language production. At first glance, this assumption seems justified in light of language comprehension studies that demonstrate that infants can visually and manually locate the appropriate referents of spoken words before they can actually articulate those words themselves (e.g., Bergelson & Swingley, 2012). At such a young age, children have not overcome many of the rudimentary obstacles required for comprehensible speech production (e.g., they lack the fine-motor skills required for delicate speech articulation, in addition to teeth). Thus, the ordering of these early language milestones—comprehension before production—must necessarily be true for children's earliest lexical acquisitions.

However, once children have the physical and cognitive fundamentals required for speech production in place, speech production could theoretically precede comprehension. The current study examines the degree to which children's early verbal responses are prompted by their knowledge that a response is required (e.g., from high-level knowledge about the discourse context) more than their desire to communicate a semantically specific message. For example, consider the following exchange between a mother and her toddler from the Providence corpus in CHILDES (Demuth, Culbertson, & Alter, 2006; MacWhinney, 2000):

MOT: *Would you like some cereal or a bagel?*
CHI (2;1): *Bagel.*

In this exchange, the child responds in a manner consistent with understanding the word "bagel" (and maybe also "cereal", since her mother's question pressed her to weigh the value of the two options against each other before responding with the preferred choice). However, the child could also accomplish the same exchange while understanding far less. In order to answer her mother's question, the child only needs to: *(1)* recognize that a question was asked, which is easily accomplished on the basis of prosodic cues, *(2)* know that questions require a verbal response, and *(3)* know that "or" marks the onset of a response option. In other words, the child could simply detect that a response was being requested and answer with the most readily available response—the last word of the question.

Prosody is one of the earliest features of language to which children have access, starting during the third-trimester when they are still in the womb (e.g., Camras et al. 1992, 1998; Mehler et al. 1988). Further, previous literature has suggested that children combine their early detection of prosodic cues with a rudimentary understanding of the linguistic contexts in order to detect and interpret questions for different types (e.g., yes/no questions, wh-questions) (Cruttenden, 1981). Taken together with the fact that children's early vocabulary is extremely limited (McMurray, 2007), it is likely that children pass through a phase of language development in which they can detect that a speaker is asking a question that requires some verbal response without fully understanding the question itself. Though the phrase "I don't know" would be a useful one for toddlers, a quick search in CHILDES reveals that it is not used until much later in language development.

With this point in mind, consider the following exchange between a mother and her child from the Brown corpus in CHILDES (Brown, 1973):

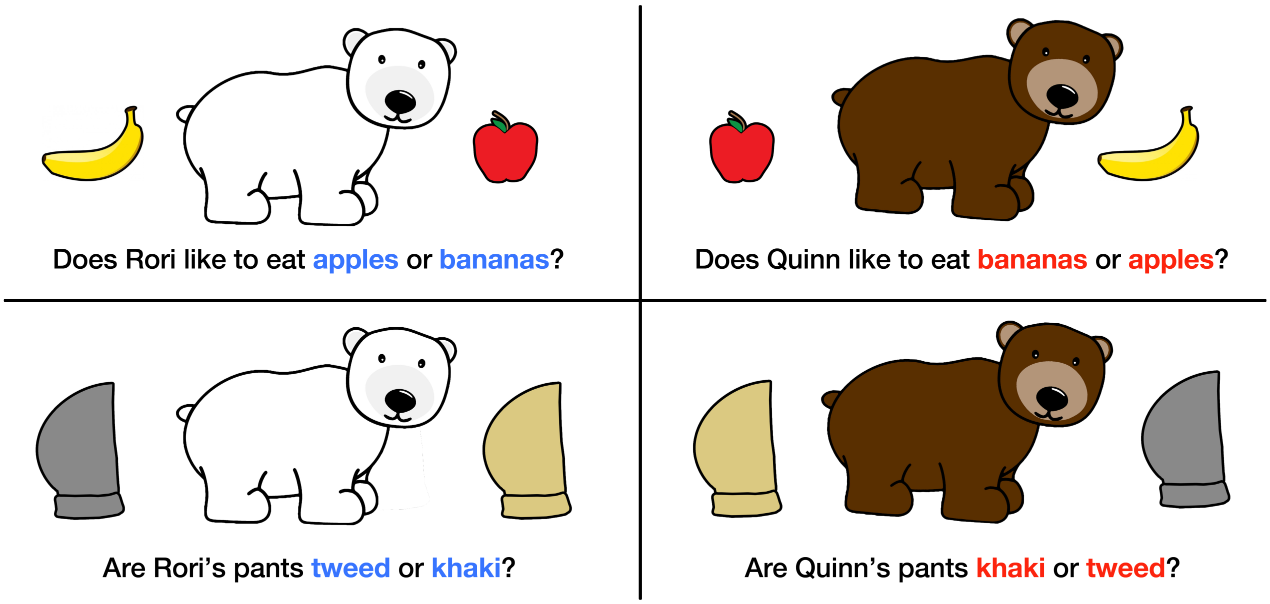

Figure 1: Static-cling stickers used in the experiment, along with verbally presented questions. Children were asked a series of questions about each bear—first in one order (e.g., choices in blue on the left) and later in the opposite order (e.g., choices in red on the right). Children were only shown the stickers after they responded to the question verbally or after two repetitions.

MOT: *Do I want coffee or tea?*
CHI (3;1): *Tea.*

The mother seemed to be asking this question aloud of herself in a rhetorical manner, but the child answered her question directly. Though it is not possible to ascertain the reason for the child's unexpected response, one possibility is that the child recognized the mother's utterance as a question and knew that questions are typically requests for a verbal response. The child may simply have responded with "tea" because he recognized it as a possible response. Further,"tea" would have been more easily accessible than other possible responses (e.g., "coffee") due to its temporal recency.

It is also important to consider that children often use words even before they have a complete, adult-like semantic understanding of those words. Research has shown that children may view words as prototypical, not necessarily definitional (e.g., Keil, 1992). Additionally, 2- and 3-year-olds sometimes acquire multi-word phrases (e.g.,"want to play") in advance of the meanings of the individual words that they contain (e.g., Bannard & Matthews, 2008).

The current paper examines whether children exhibit systematic recency biases in their early verbal behavior, first through a behavioral study and then with a corpus analysis using CHILDES. In both studies, we present evidence that children exhibit recency biases when responding to questions verbally. We argue that this strategy could be very useful in enabling toddlers to engage in verbal exchanges even before they possess fully developed semantic representations for the words that they are using.

Experimental Data

Methods

Participants Twenty-four children (*mean*=24.29, *age range*=21.2 - 27.0 months) were tested for this study. The children were recruited from the database of the Rochester Baby Lab, which includes volunteer families from the greater Rochester, NY area. All participants had normal vision and hearing, according to parental report. They were also from home environments where they were exposed to at least 90% English. Families were compensated $10 and a child-sized t-shirt for their participation in the study. An additional 6 children were tested in the study, but were excluded from the final analysis because of failure to complete at least half of the total number of trials in the study (*mean*=24.84, *range*=21.1 - 26.0).

Procedure Children were run in a quiet testing room at the Rochester Baby Lab at the University of Rochester. Upon arrival at the lab, the researcher running the study began playing with them in the waiting room to give the child time to become comfortable with the researcher and the lab space. A second researcher described the study to the parent or guardian and obtained the appropriate consent paperwork before the experiment began. Children were tested without their parents present in order to prevent parental influence on children's responses. Parents remotely monitored their children (without the children's knowledge) throughout the study via a webcam in the testing room. The testing sessions were recorded so that coding could be completed by two coders after the testing sessions were over.

The testing room contained a small, child-sized table with a 17"x14" double-sided whiteboard easel. Each child par-

Table 1: Scale for rating verbal skills.

Rating	Criteria
1	Utterances are always a single word or phrase; no word-sequence combinations (sentences), even simple ones; may have some memorized phrases (e.g., "go bye-bye"); estimated mean length of utterance (MLU) = 1.0 - 1.5
2	Utterances contain some short (mostly 2-word) word-sequence combinations (e.g., "baby cry"); no longer or complex word-sequence combinations; estimated MLU = 1.5 - 2
3	Utterances contain a combination of short and some longer word-sequence combinations (e.g., "Malachai want cookie"), but with lots of agrammaticality and a moderate level of complexity; estimated MLU = 2 - 3
4	Utterances contain many full sentences that are fairly grammatical, with many novel word-sequence combinations; estimated MLU = 3+

ticipant was invited to sit on one side of the table, opposite the researcher on the other side. The researcher explained that they were going to play a sticker game together. The researcher introduced the child to a bear character (named either Rori, a polar bear, or Quinn, a grizzly bear) whose image was printed on a static-cling sticker and placed on the easel in front of the child. The researcher explained that the child was going to help Rori/Quinn make some choices and that for each answer, the child would get a "sticker" (i.e., static-cling) to place on the board (Figure 1).

The researcher then asked the child a set of questions about items for the bear character (e.g., "Is Rori's shirt red or yellow?"). The questions were asked verbally, in the absence of any visual cues as to the question's meaning. The static-cling stickers were kept inside a book that the researcher held behind the easel so that children could not peek and see what sticker options were available. When children gave an answer, they were given a static-cling sticker that depicted their choice to stick on the board. If the child did not make a choice after the question was asked the first time, the question was repeated up to two times. If the child still did not make a choice verbally, the researcher visually presented the two static-cling stickers that corresponded to the choices in the question and repeated the question up to two more times. The two static-cling stickers were presented in a randomized left-right configuration, each equidistant from the child. Children either responded to this final question verbally or by pointing, and were given the appropriate static-cling sticker to place on the board. This design allowed them to get a sticker to place on the board for each trial, regardless of their ability to make a verbal choice.

Once the child ran through the first set of choice questions, the child was asked to make the same choices for a different bear character (either polar bear Rori or grizzly bear Quinn, depending upon who they were introduced to in the first part of the experiment). Children were asked the same set of questions for the new bear character with the choices presented in the opposite order. For example, if they were asked if Rori's shirt was "red or yellow" during the first part of the experiment, they were asked if Quinn's shirt was "yellow or red" during the second part (Figure 1). The bear characters and order of the question sets was randomized across participants. If children appeared restless midway through the experiment, they were offered a brief free-play break.

Stimuli The stimuli consisted of two sets of 20 questions, each containing two choices. The second set of questions was identical to the first except that choices were presented in the opposite order. Question choices varied in terms of their commonality and frequency in child-directed speech. Some words were highly familiar to most 2-year-olds (e.g., red, cat) and some were likely to be unfamiliar (e.g., ganache, khaki). Questions also varied in terms of where the choices appeared within them. Some questions contained choices at the end (e.g., "Does Rori like to eat apples or bananas?") and some contained choices earlier (e.g., "Should Rori bring a backpack or a lunchbox to school").

Coding Two researchers coded each child's responses from the video recording of the testing session. Additionally, the child's verbal skills were ranked on a scale of 1 to 4 independently by two researchers who observed the participant in the waiting room in advance of the study, according to a predetermined set of rating criteria (detailed in Table 1). The two coders assigned the same verbal-skill rating to children 70.83% of the time and differed by one 29.17% of the time. When coders differed by one, the two verbal-skill rating values for that child were averaged. Child participants had a median verbal-skill rating of 2 (*mean*=2.1, *range*=1-4).

Analysis Our primary analysis examined the proportion of second-choice responses children made when verbally responding to the questions. This analysis allowed us to determine if children have a bias to respond with the most easily accessible choice option—a *recency bias*. We used a Wilcoxon signed-rank test in order to compare this proportion to the proportion we would expect by chance—*mu*=0.5. We also compared the proportion of second-choice responses that children made verbally to those that they made non-verbally. If, in fact, children exhibit a recency bias because recent linguistic material is more readily available in their phonological loop, we would expect to see an effect only for verbally answered questions.

We also used a generalized linear mixed model with random intercepts for items and subjects in order to evaluate the influence of whether the response was verbal—

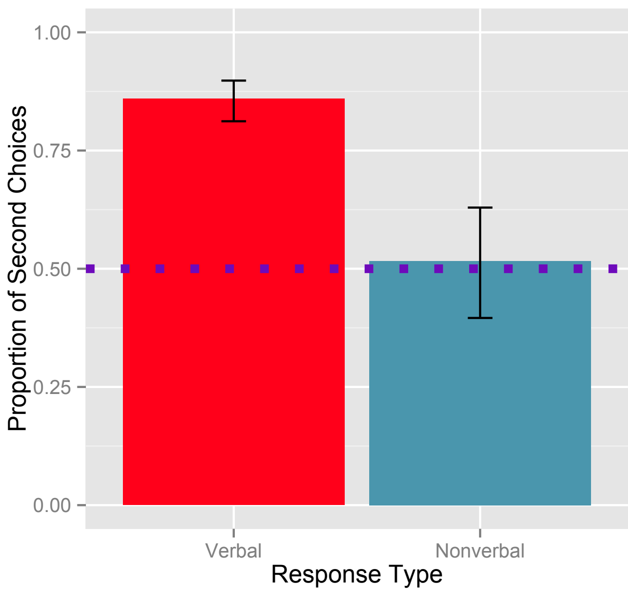

Figure 2: Proportion of second choices made for verbal responses (red) and nonverbal responses (blue). Dotted line indicates chance. Error bars represent 95% confidence intervals.

along with other factors (age, verbal skills, choice familiarity/frequency, and whether or not the choices occurred at the end of the question)—on children's likelihood of responding with the second choice. Age (in months) was scaled before it was entered into the regression analysis. Frequency of the choice words—which was estimated using the Google Books NGram corpus for English (Michel et al., 2010)[1]—was logged and then scaled before it was entered.

Results

Recency bias in verbal responses Participants responded verbally 78.7% of the time. Across all responses (verbal and nonverbal), we found that children were more likely to pick the second choice during this task. On average, participants chose the second option 69.93% of the time. A Wilcoxon signed-rank test showed that this was significantly above chance (V=803.5, p<0.0001). A comparison of verbal to nonverbal responses suggests that this bias is limited to the verbal domain (Figure 2).

For verbal responses only, participants chose the second choice 88.9% of the time—a value that a Wilcoxon signed-rank test confirmed to be significantly above chance (V=231, p<0.0001). For nonverbal responses, participants chose the second choice only 51.6% of the time, which was not significantly different from chance (V=102, p = 0.79). The proportion of second choices was significantly higher for verbal than non-verbal responses (W = 384.5, p<0.0001).[2] Further, 22 of the 24 subjects showed significantly more second-choice responses in an analysis of individual subjects (Figure 3).

Table 2: Generalized linear mixed model results.

Factor	Coef.	SE	z	p
Intercept	-0.30	0.44	-0.69	0.49
VerbalResponse	2.18	0.26	8.22	<2e-16 ***
scale(AgeMo)	-0.31	0.15	-2.07	<0.04 *
scale(log(Freq))	-0.10	0.14	-0.74	0.46
Trial	0.00	0.01	0.12	0.90
VerbalSkill	-0.04	0.15	-0.26	0.80
EndChoiceLoc	0.31	0.29	1.07	0.28

Effects of verbal response-type and age A generalized linear mixed model with random intercepts for items and subjects revealed that verbal responses and age were significant predictors of second-choice responses (Table 2). Consistent with the results reported above, verbal responses generated more second-choice responses (β=2.18, z=8.22, p<0.0001). Older children also made fewer second-choice responses than younger children, as revealed by the significantly negative coefficient for age in the regression results (β=-0.31, z=-2.074, p<0.04). No other factors—verbal skills, choice familiarity/frequency, and choice position in the question—reached significance.

More repetitions for choice-medial questions Though there was no effect of the location of the choices on the likelihood of second-choice responses in the mixed-model analysis, the question structure did impact how many repetitions were necessary before children answered the questions. Fewer repetitions were required when choices were at the end of the sentence (*mean*=0.499) than when the choices were embedded earlier in the question (*mean*=0.979). This result suggests that children struggled more with comprehension and response formulation when the choices were earlier in the question rather than at the end.

CHILDES Data

We performed an analysis using CHILDES (MacWhinney, 2000), a large corpus of transcribed caretaker-child interactions, in order to test whether the second-choice bias we observed existed in more naturalistic conversational contexts involving children and adults. We also sought to determine until what age the observed bias persisted, following up on the significant coefficient for age in the mixed-model analysis. The mixed-model results suggested that children make fewer second-choice responses as they age, which is consistent with the idea that the behavior is a heuristic which is convenient in

[1] Though children's familiarity with words is more traditionally estimated with child-specific resources such as the MacArthur-Bates Communicative Development Inventories (Fenson et al., 2007), we opted instead to use the much larger Google Books NGram corpus because the low-frequency words in our stimuli could not be accurately estimated with smaller corpora.

[2] In this within-subject analysis, three participants responded verbally 100% of the time so they were excluded from this particular analysis.

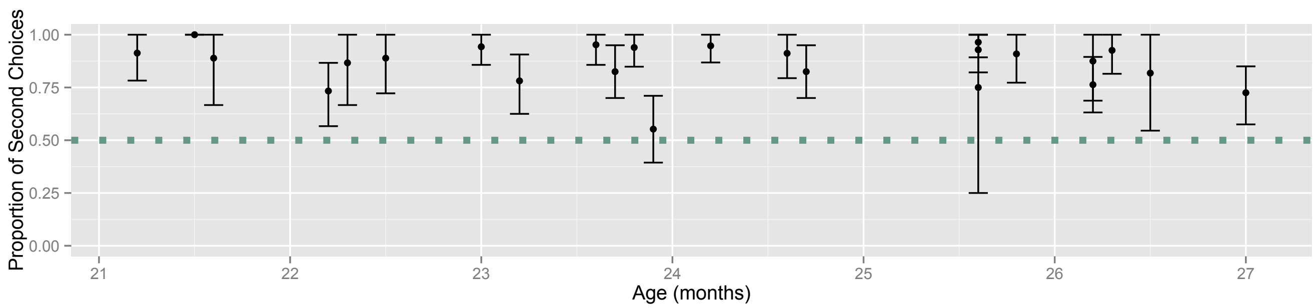

Figure 3: Each dot represents an individual's proportion of second choices during trials in which they responded verbally. The error bars depict 95% confidence intervals. The dotted line represents chance. Twenty-two of the twenty-four participants exhibited significantly more second-choice responses than would be predicted by chance.

the earliest stages of language acquisition that becomes unnecessary as the child ages and gains more language knowledge. Presumably, children must reach an age and point in language development at which they no longer depend on simple heuristics for responding to questions. Though recency biases have been attested in adults, so have primacy biases and other serial position effects (e.g., Deese & Kaufman, 1957; Greene, 1986; Murdock, 1962; Neath & Knoedler, 1994). We expected that, if such questions occurred in naturalistic contexts in CHILDES, the strength of the bias should decrease with age.

Methods

We limited our analysis to CHILDES transcript files that involved only two participants—one adult and one child—and that included the age of the child. This way, we could insure that questions we extracted from adult speech were most likely directed to a child of a known age (as opposed to a sibling or other adult). From this predefined subset, we extracted 534 two-alternative choice questions. The questions were posed to children ranging in age from 0.74 to 4.54 years of age (*mean* = 2.51). We then coded children's responses to the question for whether they chose the first option, the second option, or neither option (e.g., by failing to respond, or responding with an irrelevant verbal response).

Results

Out of all of the two-choice questions adults asked children in CHILDES, children responded with the first or second option 58.17% of the time. The proportion of first and second-choice responses they gave binned by age appears in Figure 4. As evident from the plot, binned responses from 2-year-olds are significantly more biased towards the second-choice option, as compared to chance. This bias is no longer significant in the 3- and 4-year-olds, who are equally likely to select the first and second choice. Figure 5 shows the raw data (displayed as circles at the top and bottom of the plot to represent the 1's and 0's encoding whether each response was for the second-choice option), as well as a LOESS curve fitted to the data. The smoothed fit to the data suggests that young chil-

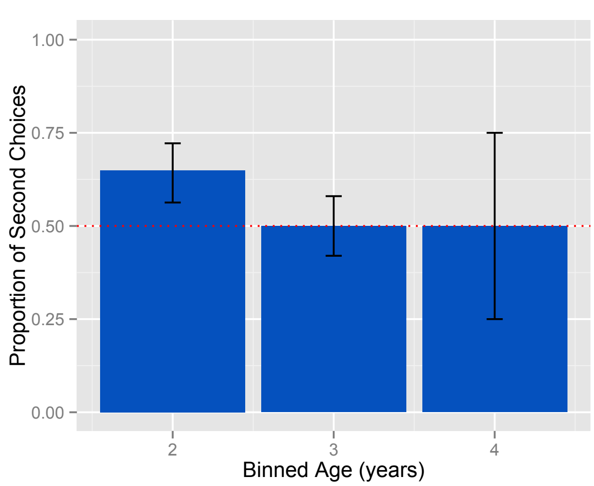

Figure 4: Proportion of second-choice responses in CHILDES binned by age of the children in years.

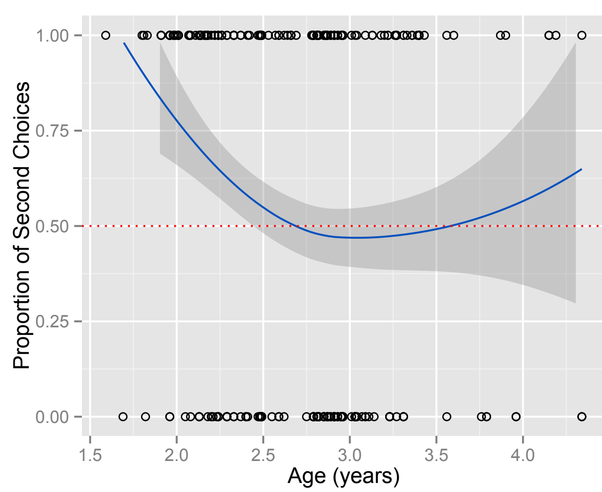

Figure 5: A smoothed LOESS curve fitted to the raw data from CHILDES (empty dots).

dren exhibit a strong second-choice bias, but that choices are at chance starting at between 2.5 and 3 years of age. The data suggests that recency biases persist until around the third year of life.

Conclusion and Discussion

In both our experimental data and CHILDES analysis, we found that young children exhibit a strong recency bias when responding to questions that present two choices. The CHILDES analysis confirms that this effect exists outside of a laboratory setting, and that it persists until around the third year of life.

We often view language production as an indicator of a child's comprehension and language knowledge. Our findings have demonstrated that this is not always the case. Our results show how children may begin engaging in conversation even without a complete semantic understanding of what they are saying (e.g., Skinner, 1957). These findings also have obvious and direct implications for the field of developmental psychology, especially work with children younger than 3 years of age. Researchers of language development should be especially careful in how they phrase questions, and ensure that questions that pose two-alternatives are appropriately counterbalanced. More importantly, however, production measures in language acquisition research may strongly reflect factors relating to memory and processing rather than true knowledge of language.

Though we have demonstrated here that this bias is robust across young children, we have not yet tested the causal mechanism underlying it. While we mentioned the possibility that children simply lack sufficient language knowledge in order to make a choice based on a genuine value comparison, this bias could be considerably more general. It could be that this bias appears in the face of a cognitive overload generally (either because of limited understanding of the words, or because of some other factor that limits processing speed and capacity, such as exhaustion). Limited cognitive processing resources (e.g., memory) could mean that adults could show this same bias under high cognitive load (when sleep deprived or choosing between long, low-frequency options). The second option is likely to be far more accessible in these circumstances due to its availability in the phonological loop (e.g., Baddeley, Gathercole, & Papagno, 1998).

Acknowledgments

We thank R. Malachai Goodman for the inspiration and unsolicited pilot data for these experiments; Steve Piantadosi and Shirlene Wade for assistance with analyses; Holly Palmeri, Claire Hart, Anetaijia Porter-Monroe, Ashley Rizzieri, Emily Roemer, and Kelsey Spear for recruiting and scheduling participants; Carla Macias, Eric Partridge, and Madeline Pelz for coding; the University of Rochester for funding; and members of the Kidd Lab for helpful feedback and discussions.

References

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning the effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, *19*(3), 241–248.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Camras, L. A., Oster, H., Campos, J., Campos, R., Ujiie, T., Miyake, K., et al. (1998). Production of emotional facial expressions in european american, japanese, and chinese infants. *Developmental Psychology*, *34*(4), 616.

Camras, L. A., Oster, H., Campos, J. J., Miyake, K., & Bradshaw, D. (1992). Japanese and american infants' responses to arm restraint. *Developmental Psychology*, *28*(4), 578.

Cruttenden, A. (1981). Falls and rises: meanings and universals. *Journal of Linguistics*, *17*(01), 77–91.

Deese, J., & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of experimental psychology*, *54*(3), 180.

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of english. *Language and Speech*, *49*(2), 137–173.

Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J. (2007). *Macarthur-bates communicative development inventories*.

Greene, R. L. (1986). Sources of recency effects in free recall. *Psychological Bulletin*, *99*(2), 221.

Keil, F. (1992). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, *317*(5838), 631–631.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Team, T. G. B., et al. (2010). Quantitative analysis of culture using millions of digitized books. *Science*.

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482.

Neath, I., & Knoedler, A. J. (1994). Distinctiveness and serial position effects in recognition and sentence processing. *Journal of Memory and Language*, *33*(6), 776–795.

Skinner, B. F. (1957). *Verbal behavior*. Acton, MA: Copley Publishing Group.