The Temporal Structure of Parent Talk to Toddlers About Objects

Lauren K. Slone,[1,2] Drew H. Abney,[1,3] Linda B. Smith,[1,4] and Chen Yu[1,5]

[1] Indiana University, Department of Psychological and Brain Sciences, Bloomington, 1101 E. 10th St., Bloomington, IN 47405-7007, USA

[2] Hope College, Department of Psychology, Holland, MI, 49423, USA

[3] University of Georgia, Department of Psychology, 125 Baldwin St., Athens, GA, 30602, UGA

[4] University of East Anglia, School of Psychology, Norwich, Norfolk, UK

[5] University of Texas at Austin, Department of Psychology, Austin, TX, 786712

Author note: Correspondence concerning this article should be addressed to Lauren K. Slone, Hope College, Department of Psychology, 35 East 12th Street, Holland, MI 49423. Email: slone@hope.edu

CRediT Author Statement: Lauren Slone: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing - original draft, Writing - review & editing; **Drew Abney**: Conceptualization, Methodology, Writing - review & editing; **Linda Smith**: Funding acquisition, Writing - review & editing; **Chen Yu**: Funding acquisition, Writing - review & editing.

Data Statement: Research data for this paper (utterance onsets, offsets, reference coding, and learning scores for study 2) are available on the OSF [https://osf.io/dzru7/?view_only=e6b44a6f86a14ec2a5b74cf75e687967].

Acknowledgments: This research was funded by the National Institutes of Health [grant numbers R01HD28675, R01HD074601, T32HD007475-22]; by the National Science Foundation [grant number BCS1523982]; and by Indiana University through the Emerging Area Research Grant – Learning: Brains, Machines, and Children. The authors appreciate the participation of the toddler and parent volunteers and the assistance of Computational Cognition and Learning Laboratory members in data collection and processing.

Abstract

Toddlers learn words in the context of speech from adult social partners. The present studies quantitatively describe the temporal context of parent speech to toddlers about objects in individual real-world interactions. We show that at the temporal scale of a single play episode, parent talk to toddlers about individual objects is predominantly, but not always, clustered. Clustered speech is characterized by repeated references to the same object close in time, interspersed with lulls in speech about the object. Clustered temporal speech patterns mirror temporal patterns observed at longer timescales, and persisted regardless of play context. Moreover, clustered speech about individual novel objects predicted toddlers' learning of those objects' novel names. Clustered talk may be optimal for toddlers' word learning because it exploits domain-general principles of human memory and attention, principles that may have evolved precisely because of the clustered structure of natural events important to humans, including human behavior.

Keywords: child development; language; learning; temporal structure

Highlights

1. Parent talk about objects is predominantly clustered, containing both repetition and spacing.
2. Clustered parent talk persists across different play contexts.
3. Clustered parent talk is associated with better word learning by the toddler.
4. Clustered talk may exploit domain-general learning and memory principles.

1. Introduction

Language is one of the most characteristic and influential aspects of human cognition, affecting human perception (Strange & Jenkins, 1978; Werker & Tees, 1984), attention (Carvalho, Vales, Fausey, & Smith, 2018), categorization (Lupyan, Rakison, & McClelland, 2007; Yoshida & Smith, 2005), encoding and remembering (Fausey & Boroditsky, 2010; Feist & Gentner, 2007), to name only a few. Unraveling the apparent ease and rapidity with which human toddlers learn language holds promise not only for advancing developmental science on early word learning, but also for understanding mechanisms of learning more generally, with potential implications for fields such as artificial intelligence (Smith & Slone, 2017) and education (Vlach, 2014).

Toddlers learn words in the context of speech from adult social partners. Much research has shown, unsurprisingly, that both the quantity and quality of adults' speech to their children – as measured by aggregated statistics like word frequency and lexical diversity – are predictive of a child's language ability as well as later school achievement (Cartmill et al., 2013; Hart & Risley, 1995; Hoff, 2013; Hurtado, Marchman, & Fernald, 2008; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991). To more fully understand the processes by which these outcomes come about, however, we must consider how language is actually experienced and learned *in time*. Speech is not experienced en masse, but rather it is taken in dynamically as it unfolds in time, and the processes by which children learn language are likely intricately related to the temporal properties of their language input.

Words unfolding over time are not random. People talk about what they see and what they are doing, which change with context (Montag, Jones, & Smith, 2018). Children may hear "socks" mentioned repeatedly when getting dressed in the morning, then not hear "socks" again until socks are taken off in the evening. Instead, they may hear talk about "swings" when at the park, talk about "flamingos" when at the zoo, and talk about "fossils" when at the museum, with none of these words likely mentioned again until that particular context is revisited. This clustered or "bursty," context-dependent property of language has been demonstrated at multiple time-scales, from conversations to whole texts (Abney, Warlaumont, Oller, Wallot, & Kello, 2017; Altmann, Cristadoro, & Esposti, 2012; Altmann, Pierrehumbert, & Motter, 2009). Burstiness has been quantified and modeled in large corpora of spoken and written language (Altmann et al., 2012, 2009; Church & Gale, 1995; Katz, 1996), in which words are shown to have a much higher probability of being encountered if they were just mentioned compared to their probabilities in the corpus of words as a whole. It is nearly inevitable that individual words would be bursty in corpora that span long time scales and therefore multiple contexts for talk. But the growth in children's vocabularies that can be observed over days, weeks, and months, is grounded in in-the-moment experiences of words that unfold on much shorter time scales. To the best that we can determine, the temporal properties of speech to young word learners has not been precisely quantified, despite considerable evidence that the repetitive structure of parent speech is relevant to early word learning (Brodsky, Waterfall, & Edelman, 2007; Hoff-Ginsberg, 1985, 1986, 1990).

Research examining the temporal structure of parent speech to children at shorter timescales (i.e., individual parent-child interactions) finds that parent speech is highly repetitive, with individual words and phrases often repeated across successive utterances (Brodsky et al., 2007; Broen, 1972; Frank, Tenenbaum, & Fernald, 2013; Messer, 1980; Rohde & Frank, 2014; Snow, 1972; Suanda, Smith, & Yu, 2016b). These parental self-repetitions correlate with children's language ability (Brodsky et al., 2007; Hoff-Ginsberg, 1985, 1986, 1990), and can even predict young children's learning of novel object labels when implemented in an experimental context (Schwab & Lew-Williams, 2016, 2017). However, despite the seeming importance of repeated talk on short timescales, research in this area remains largely qualitative because we lack clear quantitative descriptions of the timing properties of parent speech to young children in a single context and how this relates to the clustered temporal patterns we see at longer timescales. On short timescales, do parents mention an object in only one cluster of repeated talk and then move on, or do they intersperse multiple clusters of talk about an object over time?

The first aim of the present paper was to quantify temporal speech structure during a natural context for parent talk to their children: free-flowing play with toys. Because the timing of parent talk about individual toys might be influenced by the specific play context, we quantified parent speech in two contexts: play with a large set of real toys on the floor (Study 1), and play with three novel toys at a table (Study 2). The design of Study 2 also lends itself to our second aim: examining relations between the temporal structure of parent speech about individual novel objects and toddlers' learning of those objects' novel names.

Experimental studies of presentation timing have pitted the effects of massed (i.e., a single cluster) learning opportunities against spaced learning opportunities (Childers & Tomasello, 2002; Vlach, Ankowski, & Sandhofer, 2012; Vlach, Sandhofer, & Kornell, 2008; Vlach & Johnson, 2013). Counterintuitively, research demonstrates that spacing out repetitions of the same novel word in time can promote young children's learning and longer-term retention (Vlach et al., 2012, 2008). Nevertheless, this spacing effect is limited if the information spaced out in time has not yet been encoded strongly enough in memory so as not to be completely forgotten during the spacing interval (Appleton-Knapp, Bjork, & Wickens, 2005; Gagné, 1950;

Vlach & Johnson, 2013). This may be particularly important to keep in mind for young children, whose working memory, attention, knowledge base, and metamemory are still developing and may affect the ideal timing of presentations to support learning (Knabe & Vlach, 2020; Slone & Sandhofer, 2017). For instance, Vlach and Johnson (2013) found that 20-month-olds learned novel words via a spaced schedule, but 16-month-olds required a massed schedule with item presentations closer together to support learning.

Study 2 models toddlers' word learning outcomes, examining how different speech structures used by a parent to talk to their child about different objects relates to the child's learning of those objects' novel names. Specifically, we were interested in whether parent speech that intersperses multiple clusters of talk about an object over time in a single interaction may constitute a particularly effective training schedule. Such a training schedule provides close clustered repetitions of words in time, which may help learners resolve ambiguity of reference in the moment and help support initial encoding and short-term retention of word-object mappings (Kachergis, Yu, & Shiffrin, 2009; Suanda et al., 2016b; Vlach & Johnson, 2013; Weisleder & Fernald, 2014). Such a schedule also provides delays between clustered repetitions, which may support longer-term retention of those mappings (Atkinson & Shiffrin, 1968; Benjamin & Tullis, 2010; Brainerd & Reyna, 2002; Glenberg, 1979; Haebig et al., 2019; Landauer, 1969; Melton, 1970; Vlach et al., 2012; Wickelgren, 1970).

2. Study 1

2.1 Materials and Methods

2.1.1 Participants

Thirty-three parent-toddler dyads (*n* = 16 female toddlers) participated in this study when the child was between approximately 1 and 2 years of age (*M* = 19.0 months, *SD* = 3.2, *range*:

12.3-25.3). Families were recruited from a working and middle-class population of a Midwestern college town and given a small gift (e.g., a toddler book or t-shirt) for participating. Participants were treated in accordance with University IRB #0906000439. Informed parental consent was obtained for all dyads prior to participating in the experiment.

2.1.2 Setup and Stimuli

Parents and toddlers sat next to each other on the floor and were provided with 24 objects for play. Objects without a strong thematic structure were selected (e.g., car, snowman, block, flower, phone; see Figure 1A) so as not to impose a particular manner of play on the dyad. The parent's voice was recorded with a standard headset with a noise reduction microphone. A high-resolution camera (recording rate 30 frames per second) was mounted on the wall to the side of the floor/table, providing a side-on view of the interaction (see Figure 1). This camera provided visual information about the events that was used to annotate the referent of parents' speech.

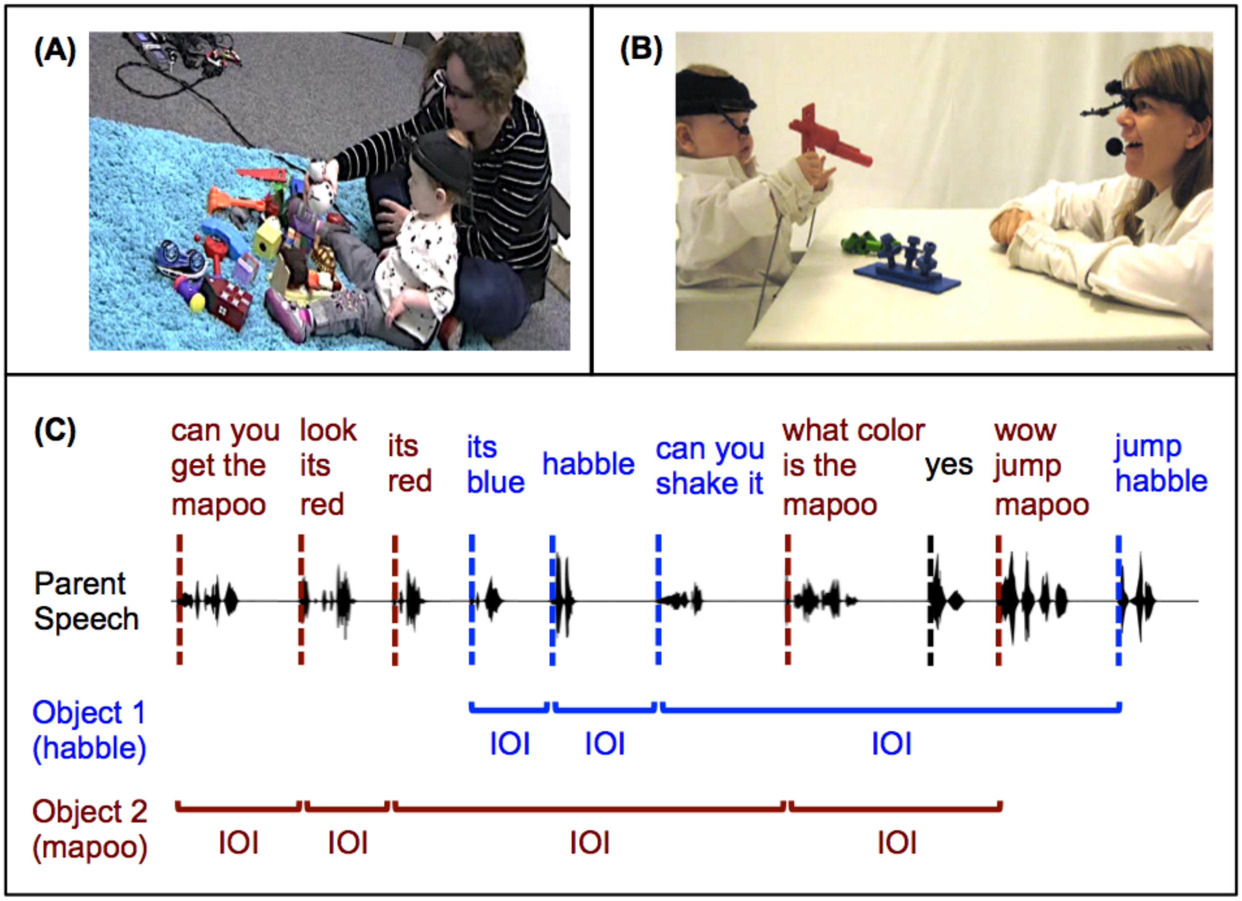

Figure 1. Experimental setup and illustration of parent speech coding. Parent and child playing with a set of toys in a free-flowing way (A) on the floor in Study 1, and (B) on a tabletop in Study 2. As shown here, 40 of the 63 dyads wore head-mounted eye trackers. All parents wore a microphone to record their speech. (C) Illustration of reference and inter-onset-interval (IOI) coding of sample parent speech. Words above utterances are transcriptions, color-coded to the object referenced (with non-reference utterances in black). Dashed vertical lines indicate utterance onsets. IOIs were computed separately for each object (only the two objects talked about in the sample speech are shown here).

2.1.3 Procedure and Coding

2.1.3.1 Procedure

All parents were told that the goal of the study was simply to observe how they and their toddler interacted with a set of toys and that they should try to play as naturally as possible. The play session began after an experimenter randomly distributed the 24 toys from a tub onto the floor in front of the dyad. The interaction lasted approximately 8 minutes or until the toddler no longer wanted to continue (*M* = 7.5 min, *SD* = 2.3). The experimenter monitored the session from a video feed in an adjacent room and re-entered the room briefly to readjust the recording equipment if it was bumped; in such cases, the resumption of play was marked as a new "trial" for coding purposes (see subsequent section).

2.1.3.2 Coding the temporal structure of parent speech

Parents' speech during each play trial was fully transcribed and divided into utterances, defined as segments of speech separated by periods of silence lasting at least 400 ms (Pereira, Smith, & Yu, 2014; Suanda, Smith, & Yu, 2016a; Yu & Smith, 2012). A number of researchers have argued and empirically demonstrated that all talk about an object has the potential to inform young children's object-name learning, not just those utterances containing the object's name (Clark, 2010; Frank et al., 2013; Messer, 1980; Schwab & Lew-Williams, 2017; Suanda et al., 2016b; Sullivan & Barner, 2016). For example, consider the two-utterance sequence, *"where's the zeebee" "there it is."* Even though the second utterance includes a pronoun rather than the object's name, it is part of a discourse context that can aid discovery of the object-name mapping. Thus, all utterances that contained reference to one of the objects were marked as referential utterances. These included utterances when parents named an object (e.g., "look a rattle"), employed a pronoun referring to an object (e.g., "can you shake it"), or used an alternate concrete noun referring to the object (e.g., "don't throw the toy"). For each referential utterance, a trained coder annotated the intended referent object by watching the video (see the supplemental material for more information). In rare cases where an utterance referenced more than one object, the first object referenced was coded as the target of the utterance. A second coder independently coded 25% of the recordings. Reliability of referential coding was determined by the Cohen's kappa (κ) statistic, and was high (κ = .77) based on conventional guidelines (Bakeman & Gottman, 1997).

The temporal structure of each parent's speech about each object was determined based on inter-onset-intervals (IOIs) of utterances about the same referent (see Figure 1C). IOIs of utterances were computed by subtracting the onset of a reference utterance from the onset of the subsequent reference utterance to that same object during that same trial. If an object was talked about during multiple trials, the vectors of IOIs for that object during each trial were concatenated. This resulted in up to 24 IOI distributions for each dyad, one for parent speech about each object.

2.2 Results

Parents produced on average 17.4 (*SD* = 3.1) total utterances per minute, 7.8 (*SD* = 2.1) of which were referential, with mean duration 1.5 s (*SD* = 1.2). Dyads did not play with the 24 objects equally frequently, but instead spent most of the time playing with only a few toys. Therefore, parent talk referred to a relatively few of the objects much more frequently than others. We analyzed the temporal structure of the on average 3.4 (*SD* = 2.5) object talk distributions per dyad that contained at least 5 IOIs (*M* = 7.9 IOIs, *SD* = 4.0); this resulted in a total of 102 IOI distributions analyzed (809 total IOIs).

The distribution of the durations of IOIs for speech about an individual object (Figure 2A) showed that short intervals occurred with high frequency, and there was also a long tail of longer IOIs, times when there was a long gap in talk about the same object. Given this skewed distribution of IOIs, as a first step in capturing the temporal structure of parents' referential utterances we classified each IOI as relatively "short" (repetition) or "long" (spacing) based on a 75th percentile split on the overall distribution of IOI durations (i.e., around the center of the distribution, see color coding in Figure 2A). By this operational definition, short and long IOIs were quite different – short IOIs were 4 s apart on average, whereas long IOIs were 71 s apart on average (Table 1).

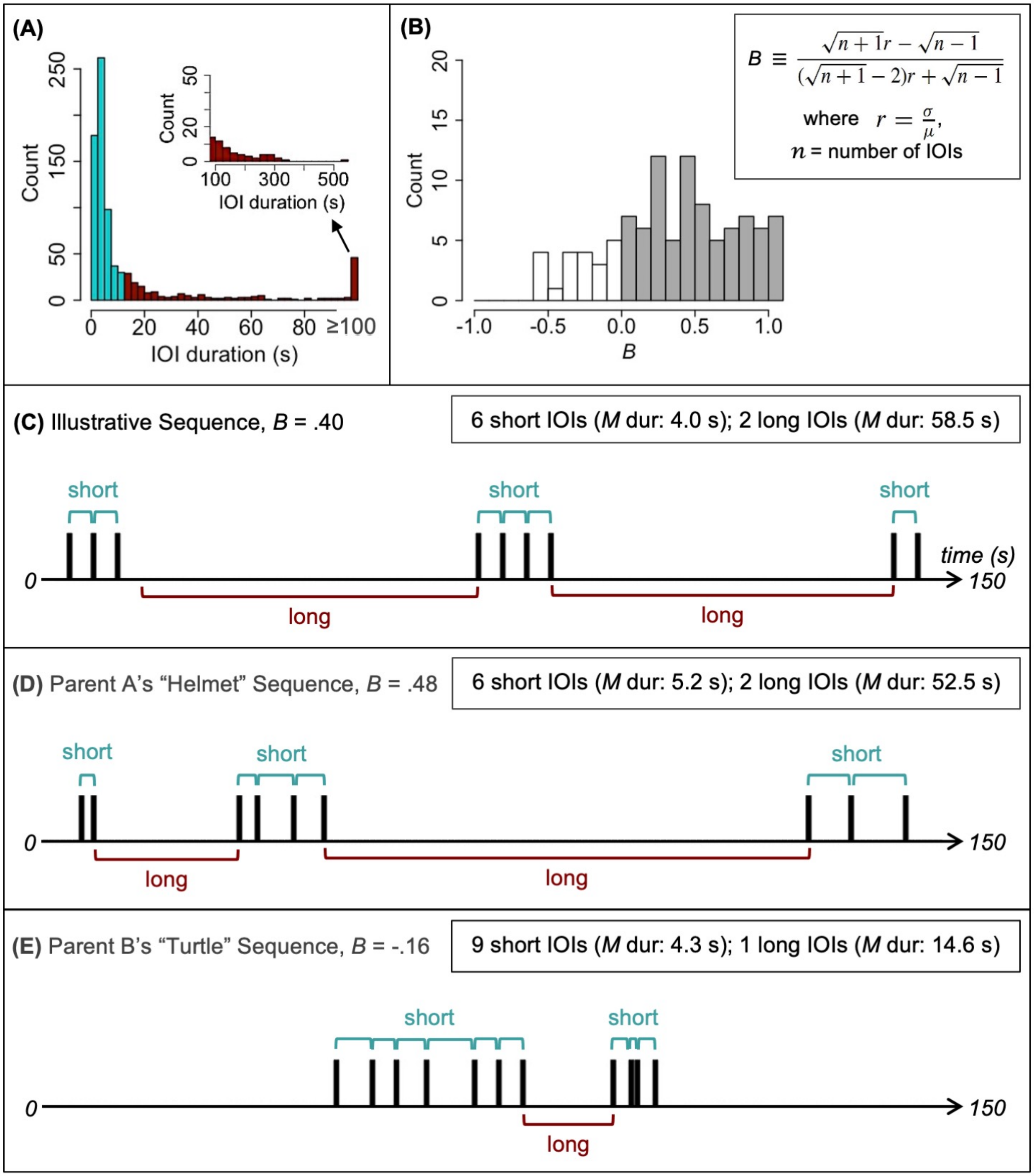

Figure 2. Illustrations of the metrics of temporal speech structure we employed. (A) Histogram of IOI durations. IOIs below the 75th percentile (cyan bars) of the IOI distribution were classified as "short;" IOIs above the 75th percentile (dark red bars) were classified as "long." (B)

Histogram of burstiness (*B)* values of IOI distributions. IOI distributions were classified as bursty (gray bars) or non-bursty (white bars) depending on whether their *B* value fell above or below 0. Inset shows the formula for calculating *B* (cf. Goh & Barabási, 2008; Kim & Jo, 2016), based on the number of IOIs and mean and standard deviation of the IOI distribution. (C) Fabricated sequence of parent utterances (black bars) to illustrate how short and long IOIs can be distributed in time and the associated metrics of temporal structure computed from the utterance sequence (see *B,* inset). (D-E) Real sequences of utterances (black bars; see the supplemental materials for the utterance transcripts) about an object from two different parents, and associated metrics of temporal structure (see *Bs*, insets), to illustrate bursty (D) and non-bursty (E) speech structure.

Table 1

Descriptive Statistics (Mean (SD)) for the Composition of Parent Referential Utterance Inter-Onset-Interval (IOI) Distributions, in Terms of Numbers (Num) and Durations (Dur) of Short and Long IOIs and Clusters, in Two Studies.

Study	**Num. total IOIs**	**Num. short**	**Num. long**	**Dur. short (s)**	**Dur. long (s)**	**Num. clusters**	**Cluster dur. (s)**	**Dur. between clusters (s)**
1	7.9 (4.0)	6.0 (3.4)	2.0 (1.6)	4.2 (2.6)	71.2 (80.5)	2.0 (1.0)	14.4 (9.2)	121.5 (113.1)
2	11.5 (4.5)	8.6 (4.2)	2.9 (1.3)	4.4 (2.9)	26.8 (12.0)	3.3 (1.1)	12.1 (6.4)	32.0 (13.0)

We next examined the composition of each IOI distribution in terms of short and long IOIs. As shown in Table 1, parent speech about each object contained, on average, 8 IOIs, composed of 6 short and 2 long IOIs. That is, for most streams of parent speech about an object, there were many more short IOIs than long IOIs. Figure 2C provides an illustration of how parents predominantly ordered speech to their children in time. Parents did not tend to produce long IOIs back-to-back (*M* = 0.6 times per object, *SD* = 0.9). Instead, they inserted clusters of close-in-time utterances (i.e., one or more short IOIs back-to-back) between lulls in talk about the object: On average, parents talked about an object in two clusters – each consisting of four to five close-in-time utterances (i.e., *M* = 3.6 short IOIs, *SD* = 2.3) and lasting around 14 seconds – separated by two minutes of no talk about the object (see Table 1). This pattern may constitute a particularly effective training schedule, as previous research suggests that spacing – long durations before repetition – is most beneficial for learning and memory when the information that is spaced out in time has already been encoded strongly enough in memory so as not to be completely forgotten during the spacing interval (Appleton-Knapp et al., 2005; Gagné, 1950; Haebig et al., 2019; Vlach & Johnson, 2013).

Although clustered references to an object between lulls in talk about the object appeared to be the predominant *overall* pattern of parent speech in this corpus, it is possible that not all talk about objects was clustered. We used the burstiness metric (*B*) (Figure 2B) from Kim and Jo (2016) to measure in a single metric the temporal structure of utterances about each object by each parent. In other words, *B* values are calculated separately for each object. *B* measures temporal structure in terms of the relation between the mean and the standard deviation of the IOI distribution (Goh & Barabási, 2008). Positive *B* values indicate clustered or "bursty" event timing, characterized by an overdispersed distribution of IOIs in which the frequency of short and long IOIs is higher than in a random (Poisson process) signal comprised of an exponential distribution of IOIs. Negative *B* values indicate more uniform spacing of IOIs compared to that expected under a random signal, with *B* = -1 indicating perfectly even spacing, as in a metronome. As can be seen in Figure 2B, 79.4% of IOI distributions had positive *B* values (*M* = 0.36, *SD* = 0.43), significantly more than would be expected by chance ($\chi^2 = 35.29$, $p < .001$), indicating that parents' talk about individual objects was predominantly bursty. The predominantly bursty nature of parent speech about objects was observed at the level of individual dyads as well: on average, a parent spoke about most objects (M = 85.3%, SD = 19.1%) with bursty speech.

Nevertheless, parents' talk distributions spanned both negative and positive B values (range -0.58 to 1.06), such that 21% of object talk distributions were classified as non-bursty (i.e., negative B values). Bursty and non-bursty distributions contained similar numbers of short IOIs (bursty: M = 5.7, SD = 3.5; non-bursty: M = 7.0, SD = 3.1) and similar durations of short IOIs (bursty: M = 4.2 s, SD =2.6; non-bursty: M = 4.2 s, SD = 2.5); that is, both bursty and non-bursty parent talk about objects to their toddlers typically included repetition close in time. Where these temporal distributions primarily differed was in the number of long IOIs (bursty: M = 2.2, SD = 1.5; non-bursty: M = 1.0, SD = 1.5; $t(28) = 3.11$, $p = .004$) and durations of the long IOIs (bursty: M = 75.8 s, SD = 83.2; non-bursty: M = 29.3 s, SD = 21.6; $t(130) = 5.94$, $p < .001$) they contained, with bursty distributions exhibiting more spacing out in time (see Figure 2D) compared to non-bursty distributions (see Figure 2E). Thus, bursty speech more clearly exhibits the dual properties of repetition close in time *and* spacing out in time that may facilitate learning and memory.

2.3 Discussion

Statistical analyses of language over long timescales highlight the clustered nature of particular words, a structure that almost necessarily falls out of the context-dependent nature of speech combined with contexts that change over time. This may give the impression that zooming in on one episode unfolding in a single context would capture a single cluster of talk about a particular topic. Indeed, previous analyses of parents' speech to children in-the-moment emphasize the highly repetitive nature of speech on short time scales (Brodsky et al., 2007; Broen, 1972; Frank et al., 2013; Messer, 1980; Rohde & Frank, 2014; Snow, 1972; Suanda et al., 2016b). However, the present analyses make clear that even in a single 8-minute interaction in a single ordinary context, talk was predominantly distributed in a bursty manner not unlike the bursty timing seen on much longer time scales, with multiple clusters of talk about an object spaced out in time by lulls in talk about that object.

What might this mean for children's word learning? The present study suggests that pitting the effects of massed learning opportunities (i.e., a single cluster) versus spaced learning opportunities (single events spaced out in time), as is common in experimental studies (Childers & Tomasello, 2002; Vlach et al., 2012, 2008; Vlach & Johnson, 2013), may not align well with real world experiences as children's language learning environments contain a combination of both types of timing, even on short time scales. Given the theoretical importance and potential implications of this finding, we conducted Study 2 to examine whether the same patterns would be observed in a different play context. It is possible that parent speech to their toddlers was bursty in Study 1 because there were many toys that the dyads could play with, creating multiple different sub-contexts for play within the larger toy play context. Study 2 analyzed parent speech in a more limiting context – parent-toddler play with three novel toys at a time on a tabletop. The use of novel toys also allowed us to examine whether the names of the toys parents talked about with bursty speech were learned better than the names of toys talked about with non-bursty speech.

3. Study 2

3.1 Materials and Methods

3.1.1 Participants

Analyses were conducted on a corpus of audio-visual recordings of 30 parent-toddler dyads engaged in unscripted, free-flowing play with six novel objects. Toddler (*n* = 14 females) participants were between approximately 1 and 2 years of age (*M* = 21.6 months, *SD* = 2.9, *range*: 15.6-26.0). Analyses on a portion of the recordings in Study 2 have been reported previously (Bambach, Crandall, & Yu, 2013; Lee, Bambach, Crandall, Franchak, & Yu, 2014; Suanda, Foster, Smith, & Yu, 2013; Suanda et al., 2016b, 2016a; Yu & Smith, 2016, 2017; Yuan, Xu, Yu, & Smith, 2017), though none of the previous published reports has examined the bursty property of parent referential speech and its effects on infant word learning.

3.1.2 Setup and Stimuli

Parents and toddlers sat across from each other at a small table (Figure 1B). Dyads played with six unique novel objects, each of which was given a unique novel name. The specific object-name mappings differed across children. The novel names were disyllabic and adhered to the phonotactic constraints of English: “habble,” “mapoo,” “wawa,” “zeebee,” “tema,” and “dodi” (Pereira et al., 2014). The novel objects were custom made from clay, wood, or plastic to have unique shapes and textures, but be similar in size (about 250-300 cm^3). Objects were organized into two sets of three. Within each set, one object was painted blue, one red, and one green. Figure 1B shows one object sets on the tabletop during play. The parent’s voice was recorded and a high-resolution camera provided a side-on view of the interaction, as in Study 1.

3.1.3 Procedure and Coding

3.1.3.1 Procedure

The parent was told the names for each of the six novel objects prior to entering the experimental room and while the toddler played with an experimenter (see the supplemental material for more information). Parents were instructed to use these names when talking about the objects, but were not told that the purpose of the study was for them to teach their toddler these names. During the experiment, laminated cards listing the object-name pairings were taped to the parent's side of the table (out of the toddler's view) as reminders of objects' names. Once parents and toddlers were seated at the table, an experimenter put one set of three objects on the table and the play session began. After approximately 90 seconds, the experimenter removed the objects and replaced them with the next set of three objects. In this manner, the dyad cycled through both sets of three objects twice, resulting in four play trials. The whole interaction lasted about six minutes, with a brief break between trials for switching object sets.

3.1.3.2 Coding the temporal structure of parent speech

Parents' speech during each play trial was fully transcribed, divided into utterances, and coded for reference to one of the objects as in Study 1. Reliability of referential coding was high ($\kappa = .81$), as in Study 1. The temporal structure of each parent's speech about each object was determined based on IOIs of utterances about the same referent, as in Study 1. This resulted in up to 6 IOI distributions for each dyad, one for parent speech about each object.

3.1.3.3 Object-name learning test

Immediately after the play session, an experimenter tested the toddler in an object-name learning task. Toddlers had passed a warm-up test with familiar objects to screen for task comprehension prior to the novel object-name testing trials. During warm-up trials the experimenter placed a flower, a horse, and an apple on a tray and presented the tray to the child while asking the child for one of the three items (e.g., "where is the *apple*, get the *apple*"). After the child made a selection, the objects were taken away, shuffled, and presented to the child again while the experimenter asked for one of the other objects (e.g., "where is the *horse*, get the *horse*"). The warm-up ended when the child had chosen the correct object on two trials (out of up to three trials).

Toddlers then completed 12 novel object-name comprehension trials. The order of the 12 testing trials was randomly determined, with two blocks of six trials in which each object name was tested once and thus twice overall. The experimenter sat across the table from the toddler. The parent sat behind the toddler and was explicitly asked not to interact with the toddler. On each trial, the experimenter put three objects – the target object plus two foils – onto a tray out of view of the toddler. Foils were pseudo-randomly selected objects from the set of six objects, with the constraint that foils could not match the target object in color. The experimenter then brought the tray into view and prompted the child to choose an object by saying "where is the *novel name*, get the *novel name.*" The experimenter provided neutral feedback (e.g., ‘‘thank you") after each selection. Each trial lasted approximately 30 seconds. Naïve coders who knew when the prompt was given but did not know the target object, coded the video for the first object the toddler touched or pointed to after the prompt on each trial. An object name was scored as "learned" only if the target object was the first object the toddler touched or pointed to after the prompt on both of the testing trials for that object name.

3.1.4 Statistical Analyses

To examine whether the object names a parent talked about with bursty speech were learned better than the names the parent talked about with non-bursty speech, we computed for each dyad two learning outcomes: the proportion of objects spoken about in a bursty way that were learned, and the proportion spoken about in a non-bursty way that were learned. To examine whether speech structure would predict object name learning controlling for the amount of parent speech about those objects and toddler age, we conducted two linear mixed effects models using the lmer function of the R package lme4 (Doran, Bates, Bliese, & Dowling, 2007);

the R code used for all models is provided in the supplementary materials. The null model included proportion names learned as the dependent variable, the mean number of IOIs (roughly equivalent to the mean number of utterances) for the objects talked about by the parent with that speech structure and toddler age as fixed effects, and by-dyad random intercepts (Baayen, Davidson, & Bates, 2008). The alternative model added speech structure (bursty versus non-bursty) as a fixed effect. Chi-squared (χ^2) tests were used to compare the null and alternative models to determine whether the addition of the speech structure variable significantly increased model fit.

We also conducted exploratory analyses to examine the possible relation between the *B value* of parent speech and toddler word learning. That is, because *B* values were calculated for each object, we can ask whether the burstiness value for an individual object predicts learning of that specific object's name. We considered these analyses to be exploratory because, although there is strong theoretical motivation based on the memory literature for treating *B* categorically and for hypothesizing that the categories of bursty and non-bursty utterance distributions, which exhibit categorical differences in the temporal distributions of their utterances, should have meaningfully different effects on word learning, there is not such clear motivation for hypothesizing a linear effect of *B values* on word learning. Such a hypothesis does make intuitive sense based on our hypothesis that the categories of bursty speech, which exhibits positive *B* values, and non-bursty speech, which exhibits negative *B* values, should differentially predict word learning. Nevertheless, in the published work on burstiness, the theorized maximal value of $B = 1$ (Goh & Barabási, 2008) has not been documented and it is not known what the ceiling *B* value for natural behavior is. Additionally, it is not yet clear whether or not it is appropriate to treat *B* as an interval scale and we are not aware of any previous findings to suggest that *B* should be linearly related to psychological outcomes.

To examine whether or not the *B value* of parent speech about an object predicted the binary learning outcome for that object's name, we conducted two generalized linear mixed models (Jaeger, 2008) using the glmer function of the R package lme4. The null model included the binary learning outcome for each object (learned, not learned) as the dependent variable, the number of IOIs (roughly equivalent to the number of utterances about the object) and toddler age as fixed effects, and by-dyad random intercepts (Baayen et al., 2008). The alternative model added the *B* value as a fixed effect. We used the most complex (maximal) random effect structure permitted by the design, removing only terms required to allow a non-singular fit (i.e., by-object random effects and by-subject random slopes were removed due to singular fit) (Barr, Levy, Scheepers, & Tily, 2013). Chi-squared ($\chi 2$) tests were used to compare the null and alternative models to determine whether the addition of the B variable significantly increased model fit. Note that, because specific object-name mappings differed across children, "object" could be defined based on either the physical items or the novel names; because the outcome is learning of the novel names, we defined "object" as the novel label used.

3.2 Results

3.2.1 Temporal structure of parent speech

Parents produced on average 20.3 (*SD* = 3.1) utterances per minute, 13.5 (*SD* = 2.5) of which were referential, with mean duration 1.3 s (*SD* = 1.0). As in Study 1, we analyzed the temporal structure of the on average 5.7 (*SD* = 0.5) object talk distributions per dyad that contained at least 5 IOIs (*M* = 11.5 IOIs, *SD* = 4.5); this resulted in a total of 170 IOI distributions analyzed (1957 total IOIs). The distribution of the durations of IOIs for speech about individual objects was skewed, with high frequencies of short intervals and a long tail of longer IOIs (Figure 3A). Based on a 75th percentile split on the overall distribution of IOI durations, short IOIs (repetition) were 4 s apart on average, whereas long IOIs were 27 s apart on average (see Figure 3A, Table 1).

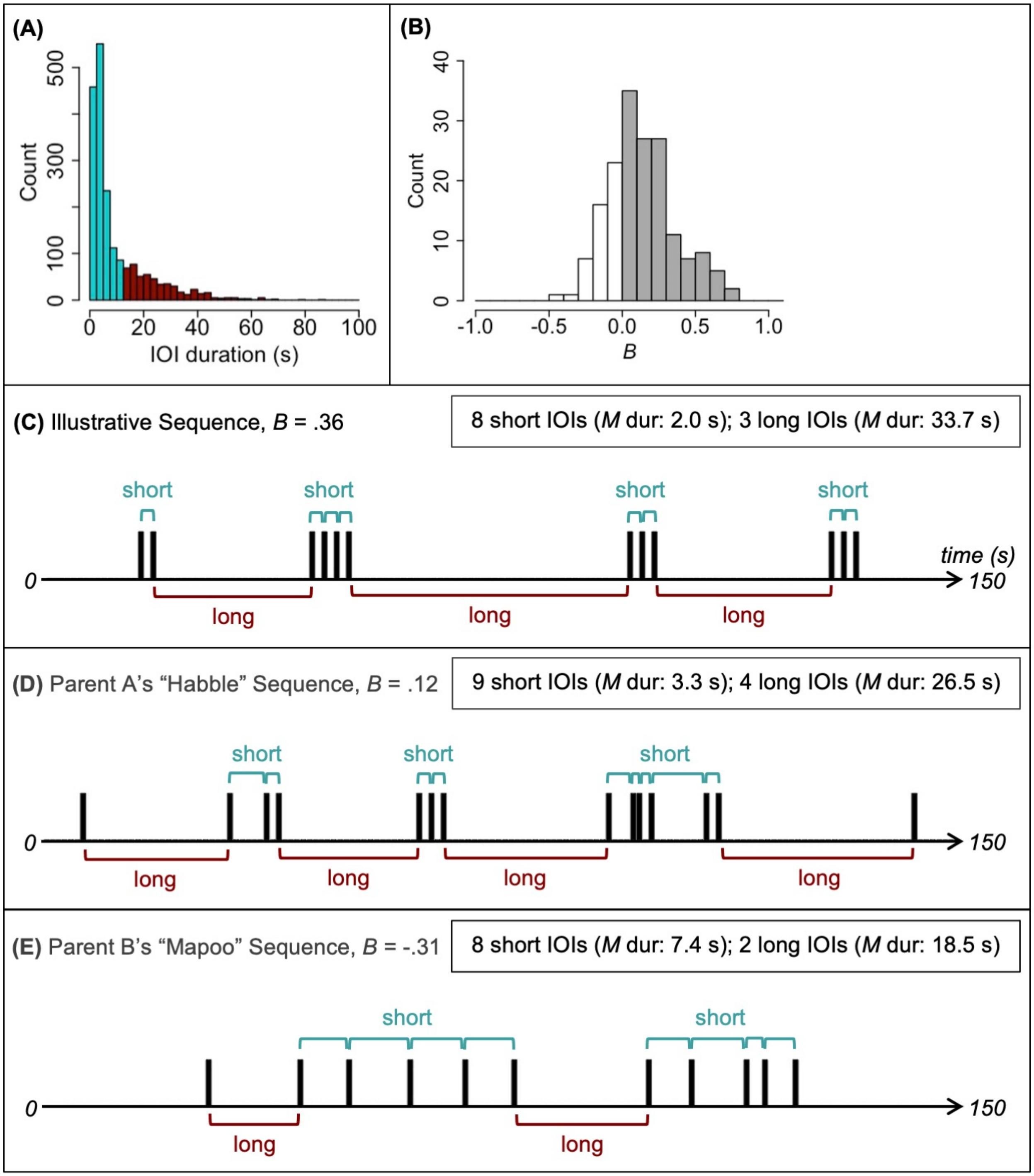

Figure 3. Illustrations of the metrics of temporal speech structure we employed. (A) Histogram of IOI durations. IOIs below the 75th percentile (cyan bars) of the IOI distribution were classified as "short;" IOIs above the 75th percentile (dark red bars) were classified as "long." (B)

Histogram of burstiness (*B)* values of IOI distributions. IOI distributions were classified as bursty (gray bars) or non-bursty (white bars) depending on whether their *B* value fell above or below 0. (C) Fabricated sequence of parent utterances (black bars) to illustrate how short and long IOIs can be distributed in time and the associated metrics of temporal structure computed from the utterance sequence (see *B,* inset). (D-E) Real sequences of utterances (black bars) about an object from two different parents, and associated metrics of temporal structure (see *Bs*, insets), to illustrate bursty (D) and non-bursty (E) speech structure.

As shown in Table 1, parent speech about each object contained, on average, 11 IOIs, composed of 8 short and 3 long IOIs. Figure 3C provides an illustration of how parents predominantly ordered speech to their children in time. Parents did not tend to produce long IOIs back-to-back (*M* = 0.5 times per object, *SD* = 0.8). Instead, they inserted clusters of close-in-time utterances (i.e., one or more short IOIs back-to-back) between lulls in talk about the object: On average, parents talked about an object in three clusters – each consisting of three to four close-in-time utterances (i.e., *M* = 2.7 short IOIs, *SD* = 1.1) and lasting around 12 seconds – separated by 32 seconds of no talk about the object (see Table 1).

We next used the burstiness metric to measure in a single metric the temporal structure of utterances about each object by each parent. In other words, *B* values are calculated separately for each object. As can be seen in Figure 3B, parents' talk about individual objects was predominantly bursty, with 71.8% of IOI distributions possessing positive *B* values (*M* = 0.13, *SD* = 0.23), significantly more than would be expected by chance ($\chi^2 = 32.21$, p < .001). The predominantly bursty nature of parent speech about objects was observed at the level of individual dyads as well: on average, a parent spoke about most objects (*M* = 71.7%, *SD* = 19.4%) with bursty speech.

Nevertheless, parents' talk distributions spanned both negative and positive *B* values (range -0.44 to 0.74), such that 29% of object talk distributions were classified as non-bursty. Bursty and non-bursty distributions contained similar numbers of short IOIs (bursty: *M* = 8.7, *SD* = 4.4; non-bursty: *M* = 8.0, *SD* = 3.6), indicating they both typically included repetition close in time. Where these temporal distributions primarily differed was in the durations of the short and long IOIs they contained, with bursty distributions exhibiting closer repetitions (i.e., shorter 'short' IOIs; bursty: *M* = 4.0 s, *SD* = 2.6; non-bursty: *M* = 5.5 s, *SD* = 3.4; $t(595) = 7.82$, $p < .001$) as well as more spacing out in time (i.e., longer 'long' IOIs; bursty: *M* = 29.5 s, *SD* = 13.3; non-bursty: *M* = 21.3 s, *SD* = 5.7; $t(480) = 9.55$, $p < .001$), as illustrated in Figure 3D-E. Thus, as in Study 1, bursty speech more clearly exhibits the dual properties of repetition close in time and spacing out in time that may facilitate learning and memory.

3.2.2 Word learning

Linear mixed effects models demonstrated that the type of speech structure (bursty versus non-bursty) accounted for significant variance in toddlers' object-name learning scores (*B* = .136, *SE* = .045, $t = 3.02$, $p = .005$) when added to a null model including toddler age and the average number of parent utterances. Moreover, the addition of the speech structure variable to the null model significantly increased model fit ($\chi^2 = 7.73$, $p = .005$). Toddlers were more likely to learn the names of objects talked about with bursty temporal structure compared to those talked about with non-bursty structure, regardless of how much the parent talked about the objects or how old the toddler was.

Additionally, because *B* values were calculated for each object, we can ask whether the burstiness value for an individual object predicts learning of that specific object's name. Generalized linear mixed models demonstrated that *B* values accounted for a marginally significant amount of variance in the binary learning outcomes (B = 1.75, SE = 0.90, z = 1.94, p = .052) when added to a null model including toddler age and the number of parent utterances. The addition of the *B* value variable to the null model significantly increased model fit (χ^2 = 3.84, p < .05). Toddlers were (marginally) more likely to learn the names of objects talked about with higher (more bursty) *B* values compared to lower (less bursty) *B* values, regardless of how much the parent talked about the object or how old the toddler was.

3.3 Discussion

Study 2 shows that burstiness characterizes parent naming on a shorter time scale than Study 1 and in the context of fewer potential referents. Study 2 also shows a link between the temporal structure of human behavior and toddler learning. Specifically, the category of bursty parent speech, compared to the same parents' non-bursty speech, resulted in the best object-name learning by their children. Additionally, the *B* value for parent speech about an object was a marginally significant predictor of learning the object's novel name. This finding suggests that the *degree* of burstiness, not just the category of busty speech, may be an important predictor of word learning. To our knowledge, this is the first demonstration that burstiness values may be linearly related to a psychological outcome. Because the optimal manner in which clusters of repetitions are spaced out in time may depend on the developmental state of the learner and individual differences in memory, attention, and prior knowledge (Knabe & Vlach, 2020; Samuelson, 2021), an important avenue for future work will be to replicate and extend the present findings to other populations and contexts, as well as examine possible interactions between burstiness, age, and task difficulty. For instance, it is possible that for younger populations or for more challenging material, *B* may exhibit curvilinear relations with learning, for instance if too long of spacings, which may be associated with the largest *B* values, become detrimental to learning (Vlach & Johnson, 2013).

4. General Discussion

Burstiness is a pervasive property of the complex systems that generate many natural events including human behavior (Eckmann, Moses, & Sergi, 2004; Goh & Barabási, 2008; Vázquez et al., 2006), and thus provides the evolutionary and developmental context for human learning. The present studies demonstrate that even on the timescale of a single play episode, regardless of the number of potential referents, parent talk to toddlers is predominantly bursty, containing not only repeated references to a single object close together in time, but also spacing out of clusters of repeated talk about that object.

4. 1 Why is Parent Speech Bursty, and Why Does This Promote Children's Word Learning?

Zipf (1949) argued that power-law distributions (e.g., in words' rank frequencies) are a fundamental property of language due to the competing needs of speakers and hearers and the desire to communicate efficiently with least effort. Recent research and theory on language evolution suggests that language structure and use have been shaped by repeated processes of transmission by adults and acquisition by children (Chater & Christiansen, 2010). Bursty speech may emerge from similar processes. Language is fundamentally about communication, depending on acquisition and use by humans, and therefore contingent upon general properties of human memory, attention, and learning. Language has likely been adapted to the brain, with features of language use that enhanced its learnability by young humans being retained and magnified over time (Chater & Christiansen, 2010; Christiansen & Chater, 2008). Bursty parent speech about objects may be selected because it facilitates toddlers' word learning by engaging domain general attentional and memory processes. Those attentional and memory processes, in turn, may have the properties they do because human behaviors in general – and many natural phenomena in the world – have a bursty temporal structure. More specifically for toddler word learning, bursty talk combines repeated references to the same object, which helps word learners resolve ambiguity of reference in the moment and promotes encoding and short-term retention of word-object mappings (Kachergis et al., 2009; Suanda et al., 2016b; Vlach & Johnson, 2013; Weisleder & Fernald, 2014), with spacing of these repetitions, which promotes longer-term retention of those mappings (Atkinson & Shiffrin, 1968; Benjamin & Tullis, 2010; Brainerd & Reyna, 2002; Glenberg, 1979; Haebig et al., 2019; Landauer, 1969; Melton, 1970; Vlach et al., 2012; Wickelgren, 1970).

It is important to note that the particular metric analyzed in the present studies – parent speech to their toddler – is one index of a whole system of behaviors that go together in fluid parent-child interaction (e.g., Karmazyn-Raz & Smith, 2022). We show that bursty parent speech is part of that complex system. There are likely many factors that conspire to make parent speech bursty in such a complex, multimodal system (e.g., locations of objects in space; motor constraints; memory; attention; the coherence of conversations – if you jump around evenly to everything, that is not a fluid conversation). Moreover, we know that parents are sensitive to the behavior of their infants, and recent research demonstrates that parents adapt the timing of their vocal behavior to that of their infants (Abney et al., 2017; Ritwika et al., 2020). Thus, children may play an important role in driving bursty parent speech, both on the timescale of conversations and over the course of evolutionary time.

4.2 Directions for Future Research

Considerable research makes clear that the quality of parent talk is a significant factor in the size and rate of growth of children's vocabulary, which in turn is a significant factor in long-term outcomes in school achievement (Carvalho et al., 2018; Lupyan et al., 2007; Strange & Jenkins, 1978; Werker & Tees, 1984; Yoshida & Smith, 2005). The finding that bursty parent talk supports object name learning and the finding – in both studies – that not all parent talk about objects is bursty, raise critical questions about just when and why talk is and is not bursty (Childers & Tomasello, 2002).

Most studies of the bursty structure of human language have focused on demonstrating that language is overall bursty and not on conversational structure or conversational contexts that support bursty talk, nor how much the burstiness of talk varies across individual components of a conversation, the context, or individuals (cf., Abney, Dale, Louwerse, & Kello, 2018; Altmann et al., 2012, 2009). These are critical questions for understanding the properties and variability of parent talk that supports learning, as well as for understanding the kinds of conversations and real-time behaviors that create burstiness and that support learning more generally. A structure like the present one – with 75% of topics bursty and 25% not – might emerge naturally in narratives in which one toy is the protagonist (or core) of play and parent speech, and other toys play a supporting role by being related to that protagonist. Might parents create this structure themselves or, rather, might this structure be inherently tied to communicating responsively in a social context (e.g., if parents continue to talk about objects that elicit a response from the child, and otherwise move on to talk about a different object)? Future research that systematically measures verbal and nonverbal behaviors of both children and parents will be essential for understanding which factors in fluid interaction conspire to produce bursty behavior.

Moreover, future studies should explore how the present findings generalize to everyday contexts beyond toy play. One potentially high-impact context to study is conversation that is principally didactic in its goals, such as when a parent or teacher intends to impart a piece of knowledge or skill to a learner or group of learners. To the extent that such teaching-focused situations may be less responsive and less conversational, driven instead by adults' beliefs about how learning happens, instruction may show a less bursty structure and thus be less effective in meeting its own goals. For instance, research on adults' metacognitive judgments of their own learning demonstrates that adults often show a bias for massed learning schedules (Knabe & Vlach, 2020).

Future research should also test experimentally the attentional and memory processes that may underlie the benefits of bursty speech for language learning. Elucidating these processes holds promise not only for better understanding early word learning, but for understanding learning, memory, and social interaction more generally. The present research is the first research to show that burstiness is associated with a consequence – better word learning by children – setting the stage for further work to consider the potential consequences of bursty events in various fields, particularly learning fields such as artificial intelligence and education, with potential clinical applications (Haebig et al., 2019; Leonard et al., 2019).

4.3 Conclusion

The distribution of time intervals between successive parent utterances about an individual object during play with their toddler – a common context for toddler word learning – typically, but not always, shows a bursty structure. Bursty talk, but not other kinds of talk, is associated with toddlers' learning the individual object names from parent talk. Conflicting experiments on human memory and word learning have shown benefits of both massed exposure to to-be-learned material and spaced exposure to that material. Both of these effects may emerge from evolutionary coordination of the timing of natural events, including human language, and human mechanisms of learning, memory, and social interaction. The present findings unify and link the remarkable proficiency of young children in word learning to the bursty structure of the natural world and human behavior and a memory that has evolved to learn in this dynamically complex environment.

References

Abney, D. H., Dale, R., Louwerse, M. M., & Kello, C. T. (2018). The Bursts and Lulls of Multimodal Interaction: Temporal Distributions of Behavior Reveal Differences Between Verbal and Non-Verbal Communication. *Cognitive Science*, *42*(4), 1297–1316. https://doi.org/10.1111/COGS.12612

Abney, D. H., Warlaumont, A. S., Oller, D. K., Wallot, S., & Kello, C. T. (2017). Multiple Coordination Patterns in Infant and Adult Vocalizations. *Infancy*, *22*(4), 514–539. https://doi.org/10.1111/infa.12165

Altmann, E. G., Cristadoro, G., & Esposti, M. D. (2012). On the origin of long-range correlations in texts. *Proceedings of the National Academy of Sciences*, *109*(29), 11582–11587. https://doi.org/10.1073/pnas.1117723109

Altmann, E. G., Pierrehumbert, J. B., & Motter, A. E. (2009). Beyond word frequency: Bursts, lulls, and scaling in the temporal distributions of words. *PLoS ONE*, *4*(11), e7678. https://doi.org/10.1371/journal.pone.0007678

Appleton-Knapp, S. L., Bjork, R. A., & Wickens, T. D. (2005). Examining the Spacing Effect in Advertising: Encoding Variability, Retrieval Processes, and Their Interaction. *Journal of Consumer Research*, *32*(2), 266–276. https://doi.org/10.1086/432236

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. *Psychology of Learning and Motivation*, *2*, 89–195. https://doi.org/10.1016/S0079-7421(08)60422-3

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bakeman, R., & Gottman, J. M. (1997). *Observing Interaction: An Introduction to Sequential Analysis*. Cambridge, UK: Cambridge University Press.

Bambach, S., Crandall, D. J., & Yu, C. (2013). Understanding embodied visual attention in child-parent interaction. In *2013 IEEE 3rd Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL-EPIROB)* (pp. 1–6). IEEE. https://doi.org/10.1109/DevLrn.2013.6652555

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/J.JML.2012.11.001

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, *61*(3), 228–247. https://doi.org/10.1016/J.COGPSYCH.2010.05.004

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-Trace Theory and False Memory. *Current Directions in Psychological Science*, *11*(5), 164–169. https://doi.org/10.1111/1467-8721.00192

Brodsky, P., Waterfall, H. R., & Edelman, S. (2007). Characterizing motherese: On the computational structure of child-directed language. In *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 833–838). Austin, TX: Cognitive Science Society.

Broen, P. A. (1972). The Verbal Environment of the Language-Learning Child. *American Speech and Hearing Asssociation Monographs*, *17*.

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

Carvalho, P. F., Vales, C., Fausey, C. M., & Smith, L. B. (2018). Novel names extend for how long preschool children sample visual information. *Journal of Experimental Child Psychology*, *168*, 1–18. https://doi.org/10.1016/J.JECP.2017.12.002

Chater, N., & Christiansen, M. H. (2010). Language Acquisition Meets Language Evolution. *Cognitive Science*, *34*(7), 1131–1157. https://doi.org/10.1111/j.1551-6709.2009.01049.x

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, *38*(6), 967–978. https://doi.org/10.1037/0012-1649.38.6.967

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(05), 489–509. https://doi.org/10.1017/S0140525X08004998

Church, K. W., & Gale, W. A. (1995). Poisson mixtures. *Natural Language Engineering*, *1*(2), 163–190. https://doi.org/10.1017/S1351324900000139

Clark, E. V. (2010). Adult offer, word-class, and child uptake in early lexical acquisition: *First Language*, *30*(3–4), 250–269. https://doi.org/10.1177/0142723710370537

Doran, H., Bates, D., Bliese, P., & Dowling, M. (2007). Estimating the Multilevel Rasch Model: With the lme4 package. *Journal of Statistical Software*, *20*(2), 1–18. https://doi.org/10.1111/j.1467-9868.2007.00600.x

Eckmann, J.-P., Moses, E., & Sergi, D. (2004). Entropy of dialogues creates coherent structures in e-mail traffic. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(40), 14333–14337. https://doi.org/10.1073/pnas.0405728101

Fausey, C. M., & Boroditsky, L. (2010). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin & Review 2010 18:1*, *18*(1), 150–157.

https://doi.org/10.3758/S13423-010-0021-5

Feist, M. I., & Gentner, D. (2007). Spatial language influences memory for spatial scenes. *Memory and Cognition*, *35*(2), 283–296. https://doi.org/10.3758/BF03193449

Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and Discourse Contributions to the Determination of Reference in Cross-Situational Word Learning. *Language Learning and Development*, *9*(1), 1–24. https://doi.org/10.1080/15475441.2012.707101

Gagné, R. M. (1950). The effect of sequence of presentation of similar items on the learning of paired associates. *Journal of Experimental Psychology*, *40*(1), 61–73.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*(2), 95–112.

Goh, K. I., & Barabási, A. L. (2008). Burstiness and memory in complex systems. *Europhysics Letters*, *81*(4), 84002–84005. https://doi.org/10.1209/0295-5075/81/48002

Haebig, E., Leonard, L. B., Deevy, P., Karpicke, J., Christ, S. L., Usler, E., … Weber, C. (2019). Retrieval-Based Word Learning in Young Typically Developing Children and Children With Development Language Disorder II: A Comparison of Retrieval Schedules. *Journal of Speech, Language, and Hearing Research : JSLHR*, *62*(4), 944. https://doi.org/10.1044/2018_JSLHR-L-18-0071

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

Hoff-Ginsberg, E. (1985). Some contributions of mothers' speech to their children's syntactic growth. *Journal of Child Language*, *12*(2), 367–385. https://doi.org/10.1017/S0305000900006486

Hoff-Ginsberg, E. (1986). Function and Structure in Maternal Speech. Their Relation to the

Child's Development of Syntax. *Developmental Psychology*, *22*(2), 155–163. https://doi.org/10.1037/0012-1649.22.2.155

Hoff-Ginsberg, E. (1990). Maternal speech and the child's development of syntax: A further look. *Journal of Child Language*, *17*(1), 85–99. https://doi.org/10.1017/S0305000900013118

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Developmental Psychology*, *49*(1), 4–14. https://doi.org/10.1037/a0027238

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, *11*(6), F31-9. https://doi.org/10.1111/j.1467-7687.2008.00768.x

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early Vocabulary Growth: Relation to Language Input and Gender. *Developmental Psychology*, *27*(2), 236–248. https://doi.org/10.1037/0012-1649.27.2.236

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/J.JML.2007.11.007

Kachergis, G., Yu, C., & Shiffrin, R. M. (2009). Temporal Contiguity in Cross-Situational Statistical Learning. In *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*. Austin, TX.

Karmazyn-Raz, H. & Smith, L.B. (2022) Discourse with few words: Coherence statistics, parent-infant actions on objects, and object names. *Language Acquisition.*

Katz, S. M. (1996). Distribution of content words and phrases in text and language modelling.

Natural Language Engineering, *2*(pt 1), 15–59. https://doi.org/10.1017/S1351324996001246

Kim, E. K., & Jo, H. H. (2016). Measuring burstiness for finite event sequences. *Physical Review E*, *94*(3), 1–7. https://doi.org/10.1103/PhysRevE.94.032311

Knabe, M. L., & Vlach, H. A. (2020). When are Difficulties Desirable for Children? First Steps Toward a Developmental and Individual Differences Account of the Spacing Effect. *Journal of Applied Research in Memory and Cognition*, *9*(4), 447–454. https://doi.org/10.1016/J.JARMAC.2020.07.007

Landauer, T. K. (1969). Reinforcement as consolidation. *Psychological Review*, *76*(1), 82–96. https://doi.org/10.1037/h0026746

Lee, S., Bambach, S., Crandall, D. J., Franchak, J. M., & Yu, C. (2014). This Hand Is My Hand: A Probabilistic Approach to Hand Disambiguation in Egocentric Video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 543–550).

Leonard, L. B., Karpicke, J., Deevy, P., Weber, C., Christ, S., Haebig, E., … Krok, W. (2019). Retrieval-Based Word Learning in Young Typically Developing Children and Children With Developmental Language Disorder I: The Benefits of Repeated Retrieval. *Journal of Speech, Language, and Hearing Research : JSLHR*, *62*(4), 932. https://doi.org/10.1044/2018_JSLHR-L-18-0070

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1083. https://doi.org/10.1111/J.1467-9280.2007.02028.X

Melton, A. W. (1970). The Situation with Respect to the Spacing of Repetitions and Memory.

Journal of Verbal Learning and Verbal Behavior, *9*, 596–606.

Messer, D. J. (1980). The episodic structure of maternal speech to young children. *Journal of Child Language*, *7*(1), 29–40. https://doi.org/10.1017/S0305000900007017

Montag, J. L., Jones, M. N., & Smith, L. B. (2018). Quantity and Diversity: Simulating Early Word Learning Environments. *Cognitive Science*, *42*(2), 375–412. https://doi.org/10.1111/COGS.12592

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, *21*(1), 178–185. https://doi.org/10.3758/s13423-013-0466-4

Ritwika, V. P. S., Pretzer, G. M., Mendoza, S., Shedd, C., Kello, C. T., Gopinathan, A., & Warlaumont, A. S. (2020). Exploratory dynamics of vocal foraging during infant-caregiver communication. *Scientific Reports 2020*, *10*(1), 1–14. https://doi.org/10.1038/s41598-020-66778-0

Rohde, H., & Frank, M. C. (2014). Markers of Topical Discourse in Child-Directed Speech. *Cognitive Science*, *38*(8), 1634–1661. https://doi.org/10.1111/COGS.12121

Samuelson, L. K. (2021). Toward a Precision Science of Word Learning: Understanding Individual Vocabulary Pathways. *Child Development Perspectives*, *15*(2), 117–124. https://doi.org/10.1111/CDEP.12408

Schwab, J. F., & Lew-Williams, C. (2016). Repetition Across Successive Sentences Facilitates Young Children's Word Learning. *Developmental Psychology*, *52*(6), 879–886. https://doi.org/10.1037/dev0000125

Schwab, J. F., & Lew-Williams, C. (2017). Discourse continuity promotes children's learning of new objects labels. In *Proceedings of the 39th Annual Conference of the Cognitive Science*

Society (pp. 3101–3106). Austin, TX: Cognitive Science Society.

Slone, L. K., & Sandhofer, C. M. (2017). Consider the category: The effect of spacing depends on individual learning histories. *Journal of Experimental Child Psychology*, *159*, 34–49. https://doi.org/10.1016/J.JECP.2017.01.010

Smith, L. B., & Slone, L. K. (2017). A developmental approach to machine learning? *Frontiers in Psychology*, *8*(DEC). https://doi.org/10.3389/fpsyg.2017.02124

Snow, C. E. (1972). Mothers' Speech to Children Learning Language. *Child Development*, *43*(2), 565. https://doi.org/10.2307/1127555

Strange, W., & Jenkins, J. J. (1978). Role of Linguistic Experience in the Perception of Speech. In R. D. Walk & J. H. L. Pick (Eds.), *Perception and Experience* (pp. 125–169). Springer, Boston, MA. https://doi.org/10.1007/978-1-4684-2619-9_5

Suanda, S. H., Foster, S. B., Smith, L. B., & Yu, C. (2013). Attentional constraints and statistics in toddlers' word learning. In *2013 IEEE 3rd Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL-EPIROB)* (pp. 1–6). IEEE. https://doi.org/10.1109/DevLrn.2013.6652542

Suanda, S. H., Smith, L. B., & Yu, C. (2016a). More than Words : The Many Ways Extended Discourse Facilitates Word Learning Coding : Parent Speech. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1835–1840). Austin, TX: Cognitive Science Society.

Suanda, S. H., Smith, L. B., & Yu, C. (2016b). The Multisensory Nature of Verbal Discourse in Parent–Toddler Interactions. *Developmental Neuropsychology*, *41*(5–8), 324–341. https://doi.org/10.1080/87565641.2016.1256403

Sullivan, J., & Barner, D. (2016). Discourse bootstrapping: preschoolers use linguistic discourse to learn new words. *Developmental Science*, *19*(1), 63–75. https://doi.org/10.1111/DESC.12289

Vázquez, A., Oliveira, J. G., Dezsö, Z., Goh, K. Il, Kondor, I., & Barabási, A. L. (2006). Modeling bursts and heavy tails in human dynamics. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *73*(3), 036127. https://doi.org/10.1103/PhysRevE.73.036127

Vlach, H. A. (2014). The Spacing Effect in Children's Generalization of Knowledge: Allowing Children Time to Forget Promotes Their Ability to Learn. *Child Development Perspectives*, *8*(3), 163–168. https://doi.org/10.1111/CDEP.12079

Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? the role of presentation timing and retrieval dynamics in generalization. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(1), 246–254. https://doi.org/10.1037/a0025260

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375–382. https://doi.org/10.1016/j.cognition.2013.02.015

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*(1), 163–167. https://doi.org/10.1016/j.cognition.2008.07.013

Weisleder, A., & Fernald, A. (2014). Social environments shape children's language experiences, strengthening language processing and building vocabulary. In & B. E. I. Arnon, M. Casillas, C. Kurumada (Ed.), *Language in Interaction. Studies in honor of Eve V. Clark* (pp. 29–49). Amsterdam: Benjamins.

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

Wickelgren, W. A. (1970). Multitrace Strength Theory. In D. A. Norman (Ed.), *Models of Human Memory.* (pp. 65–102). Elsevier Science.

Yoshida, H., & Smith, L. B. (2005). Linguistic Cues Enhance the Learning of Perceptual Cues: *Psychological Science*, *16*(2), 90–95. https://doi.org/10.1111/J.0956-7976.2005.00787.X

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262. https://doi.org/10.1016/j.cognition.2012.06.016

Yu, C., & Smith, L. B. (2016). Multiple Sensory-Motor Pathways Lead to Coordinated Visual Attention. *Cognitive Science*, *41*(S1), 5–31. https://doi.org/10.1111/cogs.12366

Yu, C., & Smith, L. B. (2017). Hand–Eye Coordination Predicts Joint Attention. *Child Development*, *88*(6), 2060–2078. https://doi.org/10.1111/cdev.12730

Yuan, L., Xu, T. L., Yu, C., & Smith, L. (2017). Seeing Is Not Enough for Sustained Visual Attention. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley.