

A meta-analysis of the syntactic bootstrapping phenomenon in word learning

Anjie Cao<sup>1</sup> & Molly Y. Lewis<sup>2</sup>

<sup>1</sup> Department of Psychology, Stanford University

<sup>2</sup> Department of Psychology, Carnegie Mellon University

Author Note

okie

Correspondence concerning this article should be addressed to Anjie Cao, what is this. E-mail: [acao@andrew.cmu.edu](mailto:acao@andrew.cmu.edu)

Abstract

so abstract!

*Keywords:* language acquisition, syntactic bootstrapping, meta-analysis

Word count:

A meta-analysis of the syntactic bootstrapping phenomenon in word learning

## Introductions

Children build up their vocabularies at a dazzling speed. Toward the end of the second year, they were estimated to have a productive vocabulary size of around 300 (Fenson et al., 1994). But children's impressive word-learning abilities do not apply equally to words of all kinds. Verbs, for example, constitute a special challenge for children. Numerous observational studies have shown that verbs are often learned later than nouns (Goldin-Meadow, Seligman, & Gelman, 1976; Lieven, Pine, & Barnes, 1992; Longobardi, Spataro, Putnick, & Bornstein, 2017; Nelson, 1973; Nelson, Hampson, & Shaw, 1993), and laboratory experiments also found that children's verb learning is more slowly and difficult than noun learning (Abbot-Smith, Imai, Durrant, & Nurmsoo, 2017; Childers, Heard, Ring, Pai, & Sallquist, 2012; Childers & Tomasello, 2002, 2006; Gentner, 1978; Imai et al., 2008; Oviatt, 1980; Schwartz & Leonard, 1984). Some scholars have argued that the challenge of verbs is language-dependent (Choi & Gopnik, 1995; Tardif, 1996; Tardif, Gelman, & Xu, 1999). Nevertheless, the cross-linguistic findings are mixed, with others found consistent patterns across languages (Au, Dapretto, & Song, 1994; Bornstein et al., 2004; Kim, McGregor, & Thompson, 2000; Papaeliou & Rescorla, 2011). A more recent large scale, cross-linguistic corpus analysis has shown that across 23 different languages, verbs and other predicates, when compared with nouns, constitute a smaller proportion of children's early production vocabularies, though there are more cross-linguistic differences in early comprehension vocabularies (Frank, Braginsky, Marchman, & Yurovsky, 2019). Together, these findings suggest that at least certain aspects of the challenge associated with verb learning are universal across languages. Some have hypothesized that it is due to the nature of verbs' references: Unlike nouns, the references of verbs are ephemeral and ever-changing. As a consequence, to learn verbs, children need to rely less on the unreliable extralinguistic information and rely more on the information within the verbs' linguistic

context, such as syntax (Gentner & Boroditsky, 2001, p. @gentner2006verbs).

The idea that syntactic information can shape children's interpretations of novel words was not new. As early as the late 1950s, a classic study by Brown (1957) has shown that 3- to 5-year-old children would choose to map a novel word to an action if the word is introduced in a verb context (e.g. "Do you know what it means to sib?"), and map to an object if introduced in a noun context (e.g. "Do you know what a sib is?"). However, this early finding only provided evidence for older children's abilities to discriminate between syntactic categories. It remained unknown how early the abilities emerge and how detailed the underlying representations for the verbs are. In the late 1980s and early 1990s, a hypothesis known as "Syntactic Bootstrapping" revived the interests in the early interaction between syntax and verbs' semantics during language development (Gleitman, 1990; Landau & Gleitman, 1985). The basic idea of Syntactic Bootstrapping is that the structures of the sentences were potent cues for children to interpret the meanings of the verbs. This idea soon received empirical support from a seminal study by Naigles (1990). By using the Intermodal Preferential Looking Paradigm (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987), Naigles (1990) provided the first evidence for 25-month-olds children's abilities to incorporate syntactic cues in verbs learning.

In this seminal study, the children were tested on four different novel verbs. For each novel verb, there was a training phase and a testing phase. During the training phase, the children would hear either a series of transitive sentences or intransitive sentences, depending on the conditions they were assigned to. Each sentence grammatically contained the novel verb (e.g., for transitive: "The duck is gorpig the bunny"; for intransitive: "The duck and the bunny are gorpig".) The children were also exposed to the visual stimuli depicting a two-actor action while listening to these sentences. The two actors, one in a bunny suit and the other in a duck suit, would perform two novel actions simultaneously. They would move their arms synchronously while the one pushed the other bending

forward. After the training phase, the children would be prompted to look for the action (e.g. “Where’s the gorping now?”). The children faced two screens side-by-side, one would show the bunny and duck performing non-causal synchronous arm movements, while the other would show the causal pushing-and-bending action. The children’s fixation time toward each of the screens was recorded. A longer looking time for either of the screen indicated a match between the children’s conceptual representations of the novel verbs and the action on the screens. Naigles (1990) found that children assigned to the transitive condition looked longer at the causal action than the non-causal action, whereas the children in the intransitive condition showed the opposite pattern. This indicated that at this age children can interpret the meanings of novel verbs as informed by the syntactic structures in a surprisingly detailed way.

This paradigm soon became a standard in studying Syntactic Bootstrapping in young children. Numerous experiments have adopted and extended this paradigm to explore different facets of Syntactic Bootstrapping. Many variations of this paradigm were developed: using pointing instead of looking as the behavioral responses measured (e.g., Arunachalam & Waxman, 2010; Kline, Snedeker, & Schulz, 2017; Rowland & Noble, 2010); using human actors as the protagonists of the actions instead of the humans in animal suits (e.g., Gertner & Fisher, 2012; Bungler & Lidz, 2006; Messenger, Yuan, & Fisher, 2015); adding a practice phase or a character-identification phase to reduce the task demands for young children (Gertner & Fisher, 2012; Scott, Chu, & Schulz, 2017; Scott & Fisher, 2009). The changes can potentially bring unintended effects on the learning outcomes: studies differ significantly in the amount of training experience provided to the participants. Some studies, similar to the original Naigles (1990) design, distributed the training experience across multiple different novel verbs (e.g., He et al., 2020a; Arunachalam, 2013; Naigles & Kako, 1993; Naigles, 1996). Others, by the contrary, focused on training and testing on one novel verb throughout the experiment (e.g., Jin, 2015; Kline et al., 2017; Messenger et al., 2015). As a result, children may have heard the novel verb repeated from as little as 3

times (e.g. Naigles, 1990, 1996; Naigles & Kako, 1993), to as many as 27 times (e.g., Arunachalam, 2013) before they were tested. There are also changes motivated by theoretical considerations. For instance, whether the syntactic cues alone were sufficient to support verbs learning has been studied by altering the visual stimuli during the training phase. Rather than seeing the visual stimuli representing the relevant actions, the children in some training phases would see two actors talking to each other or one actor talking over the phone (e.g. Yuan & Fisher, 2009). The absence of extralinguistic cues did not prevent the children from interpreting the novel verbs correctly, which further suggests the significance of the syntactic cues in verb learning at an early age.

The changes in paradigms have also led to some inconsistent findings. For example, The results have been mixed on whether the children have a transitivity bias, that is, children are more likely to match the correct scenes when the novel verbs are embedded in the transitive frames. Some experiments found that children who were exposed to the transitive sentences would show a larger preference to the matching scenes, whereas children exposed to the intransitive sentences looked by chance or showed a looking pattern similar to the control group who were not provided with any relevant syntactic cues (Arunachalam & Waxman, 2010; Yuan, Fisher, & Snedeker, 2012). This transitivity advantage can be attributed to three factors: the clarity, the experience, and the perceptual saliency of the stimuli. First, under certain contexts, the transitive sentences have less ambiguity than their intransitive counterparts. If one sees two scenes: in one a girl patting a boy, in the other a girl and a boy jumping side-by-side, then upon hearing “the girl is gorping the boy”, only the scene with patting action is a plausible interpretation. In contrast, if one hears the intransitive sentence, “the girl is gorping”, the verb can be interpreted as being consistent with both the causative scene and the non-causative scene. Second, children may have more experience with transitive sentences than intransitive sentences. A corpus analysis on parental utterances for 1- to 6-year-old children found that transitive sentences make up approximately 24.36% of all utterances, whereas intransitive

sentences occupy only 17.24% (Laakso & Smith, 2007). Thus the familiarity with transitive sentences can potentially improve children's understanding of transitive sentences in the lab. Last but not least, the perceptual saliency of the visual scenes may have contributed to the asymmetrical performance in the transitive conditions and the transitive conditions. Children were found to have a baseline preference for two actors event over one actor event (e.g., Yuan et al., 2012), and for synchronous movement over causative movement (e.g., Naigles & Kako, 1993). These preferences can make detecting an effect in the transitive condition easier than in the intransitive condition.

However, other studies have found the opposite pattern. In the second and the third experiment in Naigles and Kako (1993), it was children in the transitive conditions who showed no preference for neither of the events, whereas children in the intransitive conditions looked longer at the matching event. Similarly, Bunger and Lidz (2004) also showed that children exposed to the intransitive sentences tend to look longer at the matching scene but those exposed to the transitive sentences look by chance. There were two proposed explanations for this reversed effect. First, transitive verbs can be more ambiguous under certain contexts due to children's sensitivity to the internal structure of the event. Research has shown that children under 1 year of age already possess impressive abilities to detect the internal structure of an event. Following the structure, they can parse the event into subcomponents (Hespos, Grossman, & Saylor, 2010; Hespos, Saylor, & Grossman, 2009; Stahl, Romberg, Roseberry, Golinkoff, & Hirsh-Pasek, 2014; for a recent review, see Levine, Buchsbaum, Hirsh-Pasek, & Golinkoff, 2019). On the one hand, successful parsing underlies successful verb learning (Friend & Pace, 2011). But on the other hand, this sensitivity to the internal structure of the event can also lead to challenges for learning transitive verbs. As Bunger and Lidz (2004) pointed out, the transitive verbs can denote both the means of the action and the results of the action. In comparison, the possible meanings of intransitive verbs are more constrained to the results of the action. The wider range of possible meanings could make it more difficult for the children to

correctly match the sentences and the scenes. The second possibility for this asymmetry is the processing demands brought by the words in the argument positions. Lidz, Bunker, Leddon, Baier, and Waxman (2009) noted that the minimal demand from a simple transitive sentence is for children to map at least two nouns, the subject and the object to the corresponding agent and the patient. But for an intransitive sentence, the demand can be as low as only mapping only the subject of the sentence to the agent of the scene. In consequence, learning the novel verbs in the intransitive sentences can be easier than in the transitive sentences.

Besides the number of words in the verbs' arguments, the semantic content is also considered to be a key moderator for the learning process. Some scholars have proposed that semantically rich contexts are beneficial for verb learning, because the semantic content of the surrounding nouns can scaffold the interpretation of the verbs, together with the syntactic structure (Arunachalam & Waxman, 2011, 2015; Fisher, Hall, Rakowitz, & Gleitman, 1994; Gillette, Gleitman, Gleitman, & Lederer, 1999; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005; Imai et al., 2008; Piccin & Waxman, 2007). But the empirical support for this view is mixed. In a recent study, He et al. (2020b) manipulated the amount of semantic content in the subject argument. Some preschoolers were provided with more information about the subject ("The tall girl is fezzing"), and others less ("The girl is fezzing"). Contrary to the semantic-scaffolding view, He et al. (2020b) found that the children learned better when the nouns were not modified, indicating that the children's limited information processing abilities may have impeded them from utilizing the semantic content. This pattern also persists when the number of words in the agent argument is controlled for. Pronouns have less semantic content than concrete nouns. Multiple studies have found that preschool children learn verbs better when the verbs co-occur with pronouns instead of nouns. They are not only more likely to correctly identify the scenes corresponding to the linguistic stimuli but also are more likely to generalize the syntactic frames to new verbs (Childers & Tomasello, 2001; Lidz et al., 2009).



To further shed light on these conflicting findings, we decided to conduct a meta-analysis on syntactic bootstrapping literature. Our first goal is to estimate the robustness of this phenomenon by calculating the meta-analytic effect size. Knowing the effect size is consequential to theory building (Bergmann et al., 2018; Lewis, 2016). Many past meta-analyses have tapped into other domains of early language development such as phonotactic learning (Cristia, 2018), word segmentation (Bergmann & Cristia, 2016), and mutual exclusivity (Lewis, Cristiano, Lake, Kwan, & Frank, 2020). To our knowledge, this is the first meta-analysis examining the development of syntactic abilities in early childhood and its influence on word learning. Our second goal is to examine the potential moderators of the effect. Understanding the moderators can be valuable for both theoretical reasons and practical reasons. Due to the limited sample sizes, individual infant experiments often have insufficient statistical power, which can make detecting relevant moderators difficult (Oakes, 2017). The variations among different testing procedures also made it difficult to reconcile the conflicting findings, because the differences in the observed effect can be caused by a combination of methodological factors, rather than reflecting the differences in the measured constructs. The meta-analytic approach has the opportunity to reveal some of the underlying interactions between different factors, which can inform future researchers about their experimental designs. Last but not least, our fourth goal is to evaluate how the effect changes with age. Most of the syntactic-bootstrapping studies used a cross-sectional design. Despite the virtue of being rigorously controlled, the “time-slice” nature of this design made it difficult to reconstruct a continuous developmental trajectory. The meta-analysis can provide unique insights into how children’s abilities to incorporate syntactic information into verbs learning to develop in the first three years of life.

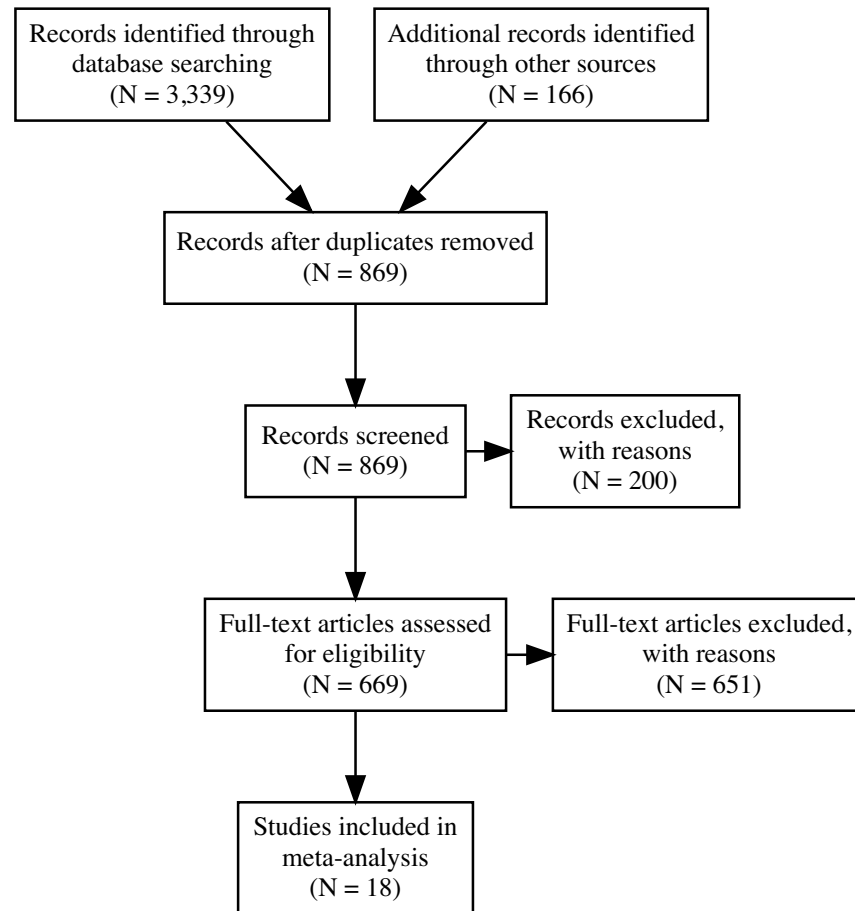


Figure 1. PRISMA plot showing...

## Methods

### Literature Search

We conducted our literature search following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist (PRISMA; Moher, Liberati, Tetzlaff, Altman, & Prisma Group, 2009). We identified relevant papers by conducting a keyword search in Google Scholar with the phrase “Syntactic Bootstrapping” and a forward search on papers that cited the seminal paper (Naigles, 1990) (total records identified:  $N = 3339$ ; retrieved between May 2020 and July 2020; Figure 1). We screened the first 60 pages ( $N = 600$ ) of the keyword search results and the first 10 pages ( $N = 100$ ) of the forward search

results. The screening processes ended because we could no longer identify relevant, non-duplicate papers from consecutive pages. Additional papers were identified by consulting the references section of the most recent literature review ( $N = 155$ ) (Fisher, Jin, & Scott, 2020) and the experts in the field ( $N = 11$ ). In our final sample, we included published journal articles, conference proceedings, doctoral dissertations, and unpublished manuscripts. They will be collectively referred to as “papers” in the following sections. Each paper may include multiple experimental conditions, and thus provides multiple effect sizes for the final analysis.

We restricted our final sample to papers that satisfied the following conditions: First, the experimental paradigm involved a two-alternative forced-choice task, in which the participants were instructed to identify the scene that matched the linguistic stimuli. Second, the visual stimuli were two events displayed side-by-side on a computer monitor. The two events must include one depicting causative action (e.g. one agent causes the other to move), and one non-causative action (e.g. two agents move simultaneously but do not causally interact with each other). The media displayed can be either videos or animation clips. Third, the linguistic stimuli included at least one novel verb embedded in a syntactically informative frame. For example, “Look, it’s kradding!” embeds the novel verb in an intransitive syntactic frame that is informative about the meaning of the novel verb “kradding”, while “Look, Kradding!” does not and therefore would be excluded. Finally, we restricted our sample to English-speaking, typically-developing children.

Our final sample included data from 849 unique infants (Mean age: 758.77 Days), reported in 60 individual effect sizes from 18 individual papers.

## Data Entry

For each paper, we entered the paper metadata (e.g., citation), information to calculate effect sizes, and moderators. We entered a separate effect size for each experimental condition and age group in paper. Most papers therefore contained multiple effect sizes.

In order to calculate the effect sizes, we recorded the sample size of each condition, the group mean, and the across-participant standard deviation. The mean and standard deviation were obtained from one of the four ways: a) directly retrieved from the results section or the data-presenting tables ( $N = 37$ ); b) recovered from the plots by measuring the height of the bars and the error bars ( $N = 10$ ); c) contacted the original authors ( $N = 10$ ); d) imputed using values from studies with similar designs ( $N = 3$ ; Among the 3, missing standard deviation from (Hirsh-Pasek, Golinkoff, & Naigles, 1996) were imputed with values from (Naigles, 1990); missing values from (Arunachalam et al., 2013) were imputed with values from (Arunachalam, 2013)). We decided to impute the missing standard deviations instead of excluding the studies because past work has shown that using “borrowed values” is beneficial to the accuracy of meta-analysis (Furukawa, Barbui, Cipriani, Brambilla, & Watanabe, 2006).

For looking time studies, when the paper only reported the raw looking time in seconds, we calculated the proportion of correct response by dividing the mean looking time toward the matching scenes by the sum of mean looking time toward the matching scenes and the mean looking time toward the non-matching scenes (i.e., excluding the look away time from the denominator). The raw standard deviations were also converted to the corresponding values by being divided by the sum.

At the participant level, we coded infants’ mean age in days ( $N = 60$ ) and the median productive vocabulary measured by MacArthur-Bates Communicative

Development Inventories (CDI) Words and Sentences (Fenson et al., 2000) ( $N = 32$ ).

At experimental condition levels, we coded moderators that are theoretically relevant and methodologically relevant. The information was either retrieved from the methods section of the paper or obtained via contacting authors. First, for the theoretically relevant moderators, we coded features of the linguistic stimuli during both training and testing, including the sentence structure and the type of words used in the agent arguments. Sentence structure has two levels, transitive ( $N = 30$ ) and intransitive ( $N = 30$ ). The training linguistic stimuli were categorized as transitive if the novel verbs were embedded in a sentence with two or more noun arguments, and intransitive if the novel verbs were embedded in a sentence with one noun argument. We also coded the types of words used in the agent argument (Noun:  $N = 26$ ; Pronoun:  $N = 34$ ). If in one condition the linguistic stimuli contained at least one occurrence of pronouns in the agent argument, then the condition would be coded as “pronoun”. Those with no pronoun throughout were coded as “noun”. We also coded word types used in the patient argument following the same rule (Noun:  $N = 24$ ; Pronoun:  $N = 6$ ). Because this only applies to studies with transitive structure ( $N = 30$ ), we only presented the exploratory analysis in the SI.

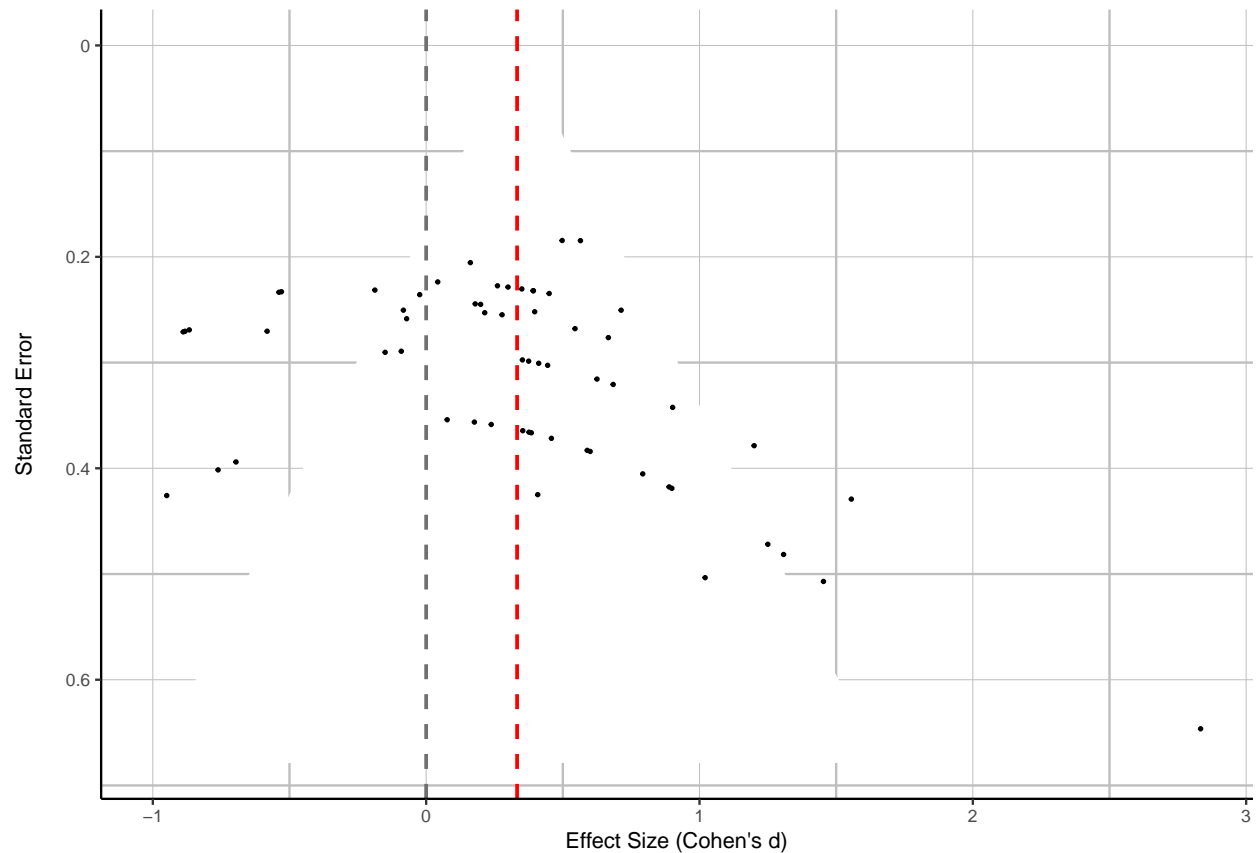
For methodologically relevant moderators, we coded the types of media used (video or animation) and the types of protagonists in the events (person or non-person). 52 conditions used video recordings of human actors ( $N = 42$ ) or human actors in animal suits ( $N = 10$ ). The other 8 conditions used clips of animation with non-person figures as the protagonists of the events. We also coded how the onset of the linguistic stimuli aligned with the visual stimuli. The procedure was coded as “simultaneous” if the very first training sentence was presented along with the visual stimuli depicting relevant action ( $N = 5$ ). It was coded as “immediately after” if it was presented along with an attention-getter or a blank screen, immediately followed by the relevant action ( $N = 18$ ). Finally, some experimental conditions first presented the linguistic stimuli paired with

irrelevant visual scenes (e.g. a person on the phone talking). The relevant visual stimuli were not shown until the training phase is over. For experimental conditions using this alignment, they were coded as “asynchronous”(N = 37).

We also coded aspects of experimental procedures. First, the type of responses elicited from the participants: whether the infants were explicitly prompted to point or their eye gaze duration were measured as they heard the linguistic stimuli (Pointing: N = 2; Looking: N = 58). Second, two characteristics of the experimental procedures were coded as categorical variables: the inclusion of practice phase (Yes: N = 36 ; No: N = 24), the inclusion of character-identification phase (Yes: N = 16; No: N = 44). Third, the distribution of the training and the testing trials (Mass: N = 28; Distributed: N = 32). A procedure was categorized as “mass” if and only if the infants were trained exclusively on one novel verb and tested on the very same verb. It was “distributed” if the infants were given multiple train and test pairs on multiple novel verbs. To better characterize the experience infants had prior to testing, we also coded how many train-test pair the infants were given (for mass procedure it was always 1), how many trials during the test phase infants were given, how many times the visual stimuli showing the relevant actions were presented, and how many times each novel verb was spoken in a syntactically-informative way.

## Data Preprocessing

We calculated the Cohen’s d effect size for each experimental condition. Because we only included studies that used a two alternative forced-choice test method, we compared the mean proportion of correct response against chance level (0.5) for the group. Although some papers collected a baseline measure (N = 6; e.g. (Arunachalam & Dennis, 2019; Arunachalam et al., 2013; Naigles, Kelty, Jaffery, & Fein, 2011)), most papers did not. In order to calculate a consistent effect size measure across all condition, we used chance level



*Figure 2.* Funnel plots. Each point corresponds to an effect size estimate. The gray dashed line represents the effect size of zero, the red dashed line represents the mean effect size, and the white funnel represents a 95% confidence interval around this mean.

as a baseline for calculating effect size for all conditions (see SI for comparison of effect sizes using reported baseline versus chance).

All the effect sizes and the coded variables were then analyzed with the metafor package in R (Viechtbauer, 2010).

## Results

### Publication Bias

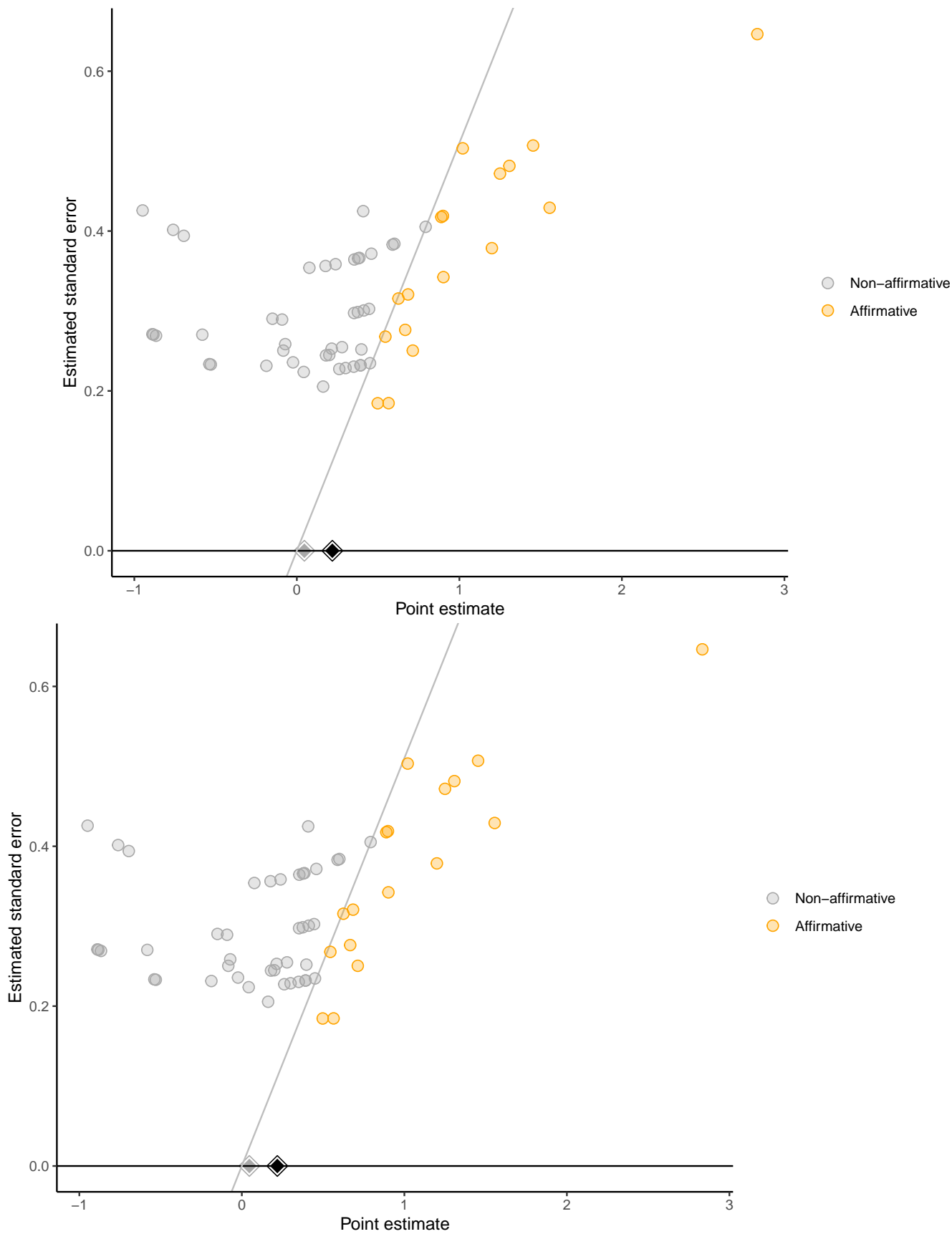
Publication bias refers to the tendency that journals would selectively publish on positive findings. This practice would inflate the real effect size and threaten the evidential

value of the literature. In this section, we presented two analyses on publication bias, one using the traditional funnel plot approach combined with Egger’s regression test, the other using a newly established method known as publication bias sensitivity analysis (Mathur & VanderWeele, 2020). We found evidence for publication bias, but the bias alone is not sufficient to explain the meta-analytic effect size.

Upon first examination, the shape of the funnel plot is asymmetrical (Figure 2). We formally tested on the asymmetry by conducting Egger’s regression test using a multi-level model. The quantitative result confirms that the asymmetry in the funnel plot is statistically significant ( $z = 3.97$ ;  $p < .0001$ ). However, this does not necessarily imply publication bias. The classical publication bias analysis methods such as Egger’s regression test have assumptions that were often not met. First, Egger’s regression test assumes homogeneous effects. But the heterogeneity test on our dataset has significant results ( $Q = 196.07$ ;  $p < .001$ ), which violated this assumption. Second, the classical methods assume that the selection bias operates on the effect size. But in reality the p-value from statistics test was often the selection criterion.

We addressed these concerns by supplementing our analysis with a sensitivity analysis method introduced by Marthur and VaderWeele (2020). Unlike its predecessors, this method did not have the homogeneity assumption nor the effect-size selection assumption. Instead, it estimated the severity of publication bias by considering how much more likely the “affirmative studies” will be published than the “non-affirmative studies”. Furthermore, it also calculated the severity of publication bias needed in the literature to “explain-away” the meta-analytic effect size.





We conducted the sensitivity analysis by using the PublicationBias provided by Marthur and VaderWeele (2020). We found that [NOT QUITE SURE HOW TO INTERPRET THE ETA STUFF]. Moreover, the analysis also suggested that it is impossible to attenuate the meta-analytic effect size to zero by hypothetically increasing the severity of publication bias (REPORT SOME TEST?). This result can also be seen in the modified funnel plot (Figure @ref(fig:moderated\_funnel)). In this modified funnel plot, the x-axis represents the point estimate of Cohen's d. The y-axis represents the estimated standard error. The gray line represents studies with a p-value of 0.05. To the left of the line are the non-affirmative studies (represented by gray dots), and to the right are the affirmative studies (represented orange dots). The diamonds near the x-axis are meta-analytic effect sizes. Note that the black one is pooling from all studies, whereas the gray one is only pooling from the non-affirmative studies. While the gray diamond is closer to the significance line than the black diamond, it does not cross the gray line. In other words, even in the scenarios with the worst publication bias, where we only consider the non-affirmative studies results, the meta-analytic effect size is still above zero.

In conclusion, we found moderate evidence for publication bias, but the bias alone is not sufficient to explain away the observed effect. Some of the observed bias using the classical methods might be attributed to the heterogeneity in the data. In the following sections, we analyzed theoretical factors and methodological factors that were contributing to the heterogeneity.

## Preliminary Analysis

The final analysis included 60 independent effect sizes from 18 papers and contained testing results from 730 unique infants. The weighted mean effect size in Cohen's d was 0.25 [0.05, 0.44], which was significantly different from 0 ( $z = 2.46$ ;  $p = 0.01$ ). The range of effects sizes across different studies was shown in the forest plot (Figure 3).

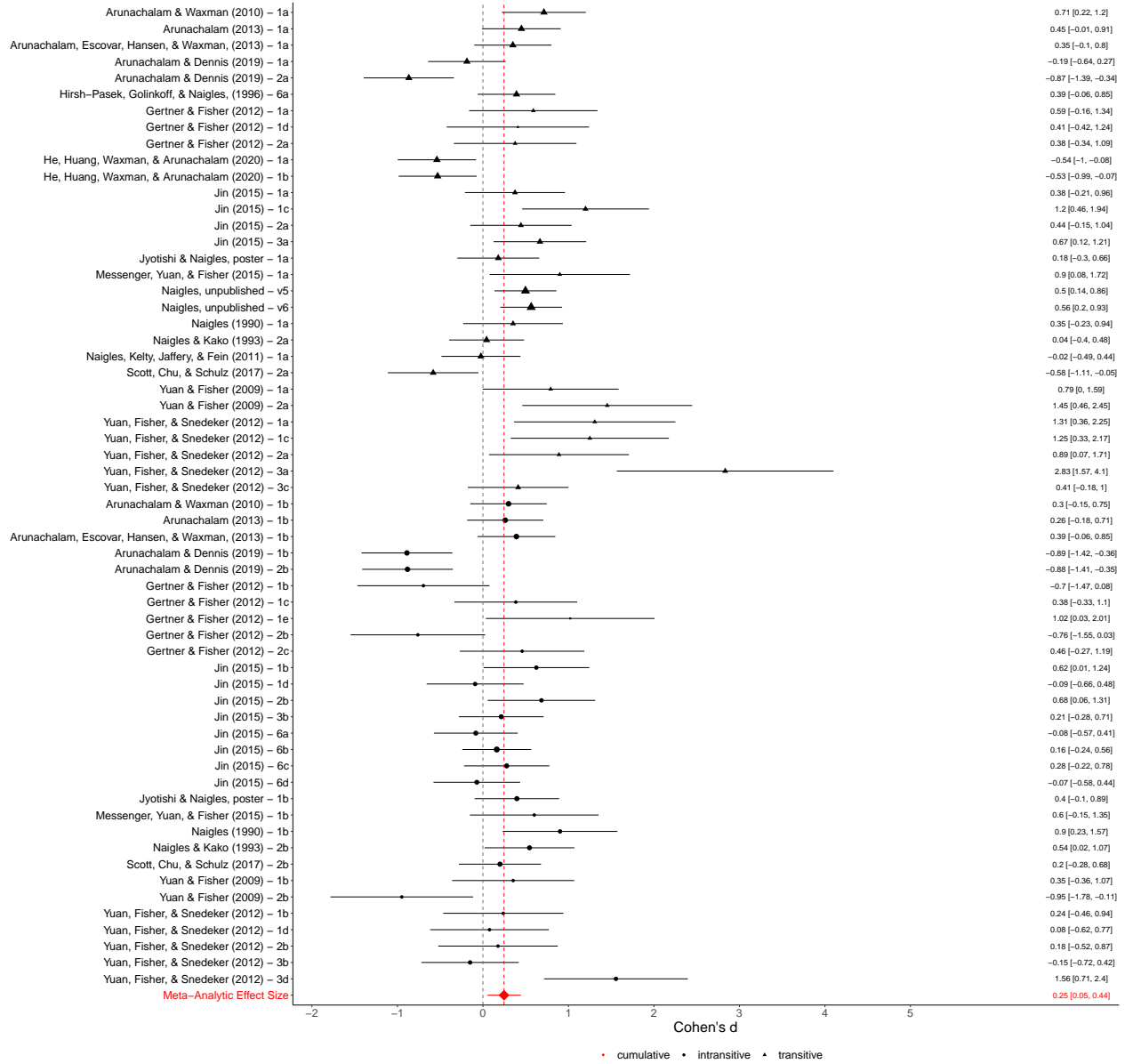


Figure 3. wow

## Theoretical Moderators

**Role of experience.** Children do not perform better in the syntactic bootstrapping task as they get more experience. Although age and median productive vocabulary size correlated with each other ( $r(30) = 0.85$ ,  $p < 0.0001$ ), neither one was a significant predictor of the effect size in the mixed-effect linear regression model (Age:  $\beta = -0.36$  [-0.84, 0.12],  $SE = 0.25$ ,  $z = -1.46$ ,  $p = 0.14$ ; Vocabulary Size:  $\beta = -0.01$  [-0.02, 0],

$SE = 0.01$ ,  $z = -1.74$ ,  $p = 0.08$ ).

**Role of sentence structure.** How would the verbs' syntactic environment and the semantic environment influence the learning outcome? To answer this question, we considered two moderators: the sentence structure and the types of words in the agent argument. The moderator sentence structure has two levels: transitive sentence and intransitive sentence. We fit a mixed effect model with the sentence structure as the single predictor. Our model suggests that sentence structure is a significant predictor ( $\beta = 0.27$  [0.04, 0.5],  $SE = 0.12$ ,  $z = 2.33$ ,  $p = 0.02$ ). Children tested in the transitive conditions ( $M = 0.49$ ,  $SD = 0.72$ ) tend to have larger effect size than those tested under the intransitive conditions ( $M = 0.17$ ,  $SD = 0.58$ ).

**Role of word types.** Like sentence structure, the moderator agent argument structure also has two levels: pronoun and noun. Using a similar model, we found that the agent argument structure is not a significant predictor of the effect size ( $\beta = 0.14$  [-0.23, 0.51],  $SE = 0.19$ ,  $z = 0.75$ ,  $p = 0.46$ ).

**Theoretical Moderators Model.** Finally, we also fit a mixed-effect model that includes all the moderators discussed above. Becasue there is no a priori reason to believe that these factors would interact with each other, we did not include interaction terms. In this model, the intercept is no longer significant ( $\beta = -0.31$  [-1.49, 0.88],  $SE = 0.61$ ,  $z = -0.50$ ,  $p = 0.61$ ). This suggests that the moderators can explain away some observed learning effects. The results for other predictors are consistent with the single-moderator models. Sentence structure is a significant predictor ( $\beta = 0.66$  [0.27, 1.05],  $SE = 0.20$ ,  $z = -0.50$ ,  $p = 0.61$ ), whereas agent argument type ( $\beta = 0.17$  [-0.47, 0.81],  $SE = 0.32$ ,  $z = 0.52$ ,  $p = 0.37$ ), mean age ( $\beta = 0$  [0, 0],  $SE = 0$ ,  $z = 0.90$ ,  $p = 0.37$ ), and median productive vocabulary size is not ( $\beta = -0.01$  [-0.03, 0.01],  $SE = 0.01$ ,  $z = -1.40$ ,  $p = 0.16$ ). In summary, we found that sentence structure is a signifnicant predictor of the effect size. Children tested with the transitive sentences performed better than those tested with intransitive sentences.

The semantic content of the words in agent argument, that is, whether the words are nouns or pronouns, was not found to be a significant predictor. Neither age nor the median productive vocabulary size as measured by CDI was a significant predictor of the effect size.

### Methodological Moderators

Motivated by the original Naigles (1990) study, a majority of the research in Syntactic Bootstrapping drew heavily on the Intermodal Preferential Looking Paradigm Golinkoff et al. (1987). Each of these later study designs often involved some study-specific adaptations of the paradigms. In this section, we consider how the various methodological changes may or may not have moderated the infants' learning outcome.

**Role of testing phases.** Some studies added additional phases such as character identification phase or practice phase to help infants get used to the testing set-up. We found that neither the presence of the character identification phase ( $\beta = 0.2$  [-0.23, 0.63],  $SE = 0.22$ ,  $z = 0.90$ ,  $p = 0.37$ ) nor the practice phase ( $\beta = -0.21$  [-0.5, 0.08],  $SE = 0.15$ ,  $z = -1.43$ ,  $p = 0.15$ ) predicts the effect size.

**Role of synchronicity..** Studies fall under three categories under the synchronicity between linguistic stimuli and visual stimuli: asynchronous, immediate-after and simultaneous. Asynchronous studies present the linguistic stimuli before the relevant visual stimuli. Immediate-after studies show the linguistic stimuli while the infants are exposed to attention-getter or blank-screen, and the visual stimuli were to follow immediately after. Simultaneous studies present the visual stimuli and the linguistic stimuli together. Our model result suggested that infants' learning outcomes were not influenced by the synchronicity between the stimuli. ( $\beta = 0.1$  [-0.24, 0.43],  $SE = 0.17$ ,  $z = 0.57$ ,  $p = 0.57$ ).

**Role of the testing procedure structure.** Last but not least, the structure of the testing procedure, mass testing or distributed testing, did not predict the effect size in

a statistically meaningful way ( $\beta = 0.35$  [-0.06, 0.77],  $SE = 0.21$ ,  $z = 1.67$ ,  $p = 0.1$ ).

Infants performed similarly in studies that have multiple train-test pair and single train-test pair. Furthermore, contrary to common beliefs, the number of exposures per novel verb the infants received prior to testing does not predict the effect sizes either ( $\beta = 0.01$  [-0.02, 0.03],  $SE = 0.01$ ,  $z = 0.44$ ,  $p = 0.66$ ).

**Methodological Moderators Model.** We fit a model that included all the theoretical moderators considered above: the presence of character identification phase, the presence of practice phase, the synchronicity between linguistic stimuli and visual stimuli, the structure of testing procedure and the number of repetitions per novel verb prior to the test. Like the methodological moderators models, the intercept became insignificant ( $\beta = -0.2$  [-0.81, 0.41],  $SE = 0.31$ ,  $z = -0.65$ ,  $p = 0.52$ ). Most of the moderators were still insignificant, with the exception of testing procedure structure. This moderator became marginally significant in the mega-model, suggesting that children tested in procedure with only one train-test pair performed better than those tested in procedure with multiple train-test pairs ( $\beta = 0.48$  [0.03, 0.93],  $SE = 0.23$ ,  $z = 2.07$ ,  $p = 0.04$ ). The differences between predictors' performance in single-moderator model and mega-model can be seen in figure(??).

## General Discussion

- Explanations for lack of age effect: variability in design as a function of age, change in input, etc. [Meta point: measuring developmental change requires constancy in task]
- compare to other effect sizes in language acquisition
- Transitivity effect is robust - implications for this? Cross-linguistic implications?

## References

- Abbot-Smith, K., Imai, M., Durrant, S., & Nurmsoo, E. (2017). The role of timing and prototypical causality on how preschoolers fast-map novel verb meanings. *First Language*, 37(2), 186–204.
- \* Arunachalam, S. (2013). Two-year-olds can begin to acquire verb meanings in socially impoverished contexts. *Cognition*, 129(3), 569–573.
- Arunachalam, S., & Dennis, S. (2019). Semantic detail in the developing verb lexicon: An extension of naigles and kako (1993). *Developmental Science*, 22(1), e12697.
- \* Arunachalam, S., Escovar, E., Hansen, M. A., & Waxman, S. R. (2013). Out of sight, but not out of mind: 21-month-olds use syntactic information to learn verbs even in the absence of a corresponding event. *Language and Cognitive Processes*, 28(4), 417–425.
- \* Arunachalam, S., & Waxman, S. R. (2010). Meaning from syntax: Evidence from 2-year-olds. *Cognition*, 114(3), 442–446.
- Arunachalam, S., & Waxman, S. R. (2011). Grammatical form and semantic context in verb learning. *Language Learning and Development*, 7(3), 169–184.
- Arunachalam, S., & Waxman, S. R. (2015). Let’s see a boy and a balloon: Argument labels and syntactic frame in verb learning. *Language Acquisition*, 22(2), 117–131.
- Au, T. K.-F., Dapretto, M., & Song, Y.-K. (1994). Input vs constraints: Early word acquisition in korean and english. *Journal of Memory and Language*, 33(5), 567–582.

- Bergmann, C., & Cristia, A. (2016). Development of infants' segmentation of words from native speech: A meta-analytic approach. *Developmental Science*, 19(6), 901–917.
- Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development*, 89(6), 1996–2009.
- Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S.-Y., Pascual, L., ... Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, dutch, french, hebrew, italian, korean, and american english. *Child Development*, 75(4), 1115–1139.
- Brown, R. W. (1957). Linguistic determinism and the part of speech. *The Journal of Abnormal and Social Psychology*, 55(1), 1.
- \* Bunger, A., & Lidz, J. (2004). Syntactic bootstrapping and the internal structure of causative events. In *Proceedings of the 28th annual boston university conference on language development* (Vol. 28, pp. 74–85). Cascadilla Press Somerville, MA.
- \* Bunger, A., & Lidz, J. (2006). Constrained flexibility in the acquisition of causative verbs. In *Proceedings of the annual boston university conference on language development* (Vol. 30, pp. 60–71).
- Childers, J. B., Heard, M. E., Ring, K., Pai, A., & Sallquist, J. (2012). Children use different cues to guide noun and verb extensions. *Language Learning and Development*, 8(3), 233–254.
- Childers, J. B., & Tomasello, M. (2001). The role of pronouns in young children's



- acquisition of the english transitive construction. *Developmental Psychology*, 37(6), 739.
- 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.
- Childers, J. B., & Tomasello, M. (2006). 12 are nouns easier to learn than verbs? Three experimental studies. *Action Meets Word: How Children Learn Verbs*, 311.
- Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in korean: A cross-linguistic study. *Journal of Child Language*, 22(3), 497–529.
- Cristia, A. (2018). Can infants learn phonology in the lab? A meta-analytic answer. *Cognition*, 170, 312–327.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, i–185.
- Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the macarthur communicative development inventories. *Applied Psycholinguistics*, 21(1), 95–116.
- Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*, 92(1-4), 333–375.
- Fisher, C., Jin, K. S., & Scott, R. M. (2020). The developmental origins of syntactic bootstrapping. *Topics in Cognitive Science*, 12(1), 48–77.
- Frank, M., Braginsky, M., Marchman, V., & Yurovsky, D. (2019). Variability and

- consistency in early language learning: The wordbank project. MIT Press  
Cambridge, MA.
- Friend, M., & Pace, A. (2011). Beyond event segmentation: Spatial-and  
social-cognitive processes in verb-to-action mapping. *Developmental Psychology*,  
47(3), 867.
- Furukawa, T. A., Barbui, C., Cipriani, A., Brambilla, P., & Watanabe, N. (2006).  
Imputing missing standard deviations in meta-analyses can provide accurate  
results. *Journal of Clinical Epidemiology*, 59(1), 7–10.
- Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child  
Development*, 988–998.
- Gentner, D. (2006). Why verbs are hard to learn. *Action Meets Word: How  
Children Learn Verbs*, 544–564.
- Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word  
learning. *Language Acquisition and Conceptual Development*, 3, 215–256.
- \* Gertner, Y., & Fisher, C. (2012). Predicted errors in children’s early sentence  
comprehension. *Cognition*, 124(1), 85–94.
- Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations  
of vocabulary learning. *Cognition*, 73(2), 135–176.
- Gleitman, L. (1990). The structural sources of verb meanings. *Language  
Acquisition*, 1(1), 3–55.
- Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005).  
Hard words. *Language Learning and Development*, 1(1), 23–64.

- Goldin-Meadow, S., Seligman, M. E., & Gelman, R. (1976). Language in the two-year old. *Cognition*, 4(2), 189–202.
- Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, 14(1), 23–45.
- \* He, A. X., Huang, S., Waxman, S., & Arunachalam, S. (2020a). Two-year-olds consolidate verb meanings during a nap. *Cognition*, 198, 104205.
- He, A. X., Kon, M., & Arunachalam, S. (2020b). Linguistic context in verb learning: Less is sometimes more. *Language Learning and Development*, 16(1), 22–42.
- Hespos, S. J., Grossman, S. R., & Saylor, M. M. (2010). Infants’ ability to parse continuous actions: Further evidence. *Neural Networks*, 23(8-9), 1026–1032.
- Hespos, S. J., Saylor, M. M., & Grossman, S. R. (2009). Infants’ ability to parse continuous actions. *Developmental Psychology*, 45(2), 575.
- \* Hirsh-Pasek, K., Golinkoff, R. M., & Naigles, L. (1996). Young children’s use of syntactic frames to derive meaning. *The Origins of Grammar: Evidence from Early Language Comprehension*, 123–158.
- Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R. M., & Shigematsu, J. (2008). Novel noun and verb learning in chinese-, english-, and japanese-speaking children. *Child Development*, 79(4), 979–1000.
- \* Jin, K. S. (2015). *The role of syntactic and discourse information in verb learning* (PhD thesis). University of Illinois at Urbana-Champaign.
- Kim, M., McGregor, K. K., & Thompson, C. K. (2000). Early lexical development in english-and korean-speaking children: Language-general and language-specific

- patterns. *Journal of Child Language*, 27(2), 225–254.
- \* Kline, M., Snedeker, J., & Schulz, L. (2017). Linking language and events: Spatiotemporal cues drive children’s expectations about the meanings of novel transitive verbs. *Language Learning and Development*, 13(1), 1–23.
- Laakso, A., & Smith, L. B. (2007). Pronouns and verbs in adult speech to children: A corpus analysis. *Journal of Child Language*, 34(4), 725.
- Landau, B., & Gleitman, L. R. (1985). Language and experience: Evidence from the blind child. Cambridge, MA, US: Harvard University Press.
- Levine, D., Buchsbaum, D., Hirsh-Pasek, K., & Golinkoff, R. M. (2019). Finding events in a continuous world: A developmental account. *Developmental Psychobiology*, 61(3), 376–389.
- Lewis, B., M. (2016). A quantitative synthesis of early language acquisition using meta-analysis. arXiv.
- Lewis, M., Cristiano, V., Lake, B. M., Kwan, T., & Frank, M. C. (2020). The role of developmental change and linguistic experience in the mutual exclusivity effect. *Cognition*, 198, 104191.
- Lidz, J., Bunger, A., Leddon, E., Baier, R., & Waxman, S. (2009). When one cue is better than two: Lexical vs. Syntactic cues to verb learning. *Unpublished Manuscript*.
- Lieven, E. V., Pine, J. M., & Barnes, H. D. (1992). Individual differences in early vocabulary development: Redefining the referential-expressive distinction. *Journal of Child Language*, 19(2), 287–310.
- Longobardi, E., Spataro, P., Putnick, D. L., & Bornstein, M. H. (2017). Do early

- noun and verb production predict later verb and noun production? Theoretical implications. *Journal of Child Language*, 44(2), 480.
- Mathur, M., & VanderWeele, T. (2020). Sensitivity analysis for publication bias in meta-analyses.
- \* Messenger, K., Yuan, S., & Fisher, C. (2015). Learning verb syntax via listening: New evidence from 22-month-olds. *Language Learning and Development*, 11(4), 356–368.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. *PLoS Med*, 6(7), e1000097.
- \* Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, 17(2), 357–374.
- \* Naigles, L. G., & Kako, E. T. (1993). First contact in verb acquisition: Defining a role for syntax. *Child Development*, 64(6), 1665–1687.
- \* Naigles, L. R. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, 58(2), 221–251.
- \* Naigles, L. R., Kelty, E., Jaffery, R., & Fein, D. (2011). Abstractness and continuity in the syntactic development of young children with autism. *Autism Research*, 4(6), 422–437.
- Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, 1–135.
- Nelson, K., Hampson, J., & Shaw, L. K. (1993). Nouns in early lexicons: Evidence, explanations and implications. *Journal of Child Language*, 20(1), 61–84.

- Oakes, L. M. (2017). Sample size, statistical power, and false conclusions in infant looking-time research. *Infancy*, 22(4), 436–469.
- Oviatt, S. L. (1980). The emerging ability to comprehend language: An experimental approach. *Child Development*, 97–106.
- Papaeliou, C. F., & Rescorla, L. (2011). Vocabulary development in greek children: A cross-linguistic comparison using the language development survey. *Journal of Child Language*, 38(4), 861.
- Piccin, T. B., & Waxman, S. R. (2007). Why nouns trump verbs in word learning: New evidence from children and adults in the human simulation paradigm. *Language Learning and Development*, 3(4), 295–323.
- \* Rowland, C. F., & Noble, C. L. (2010). The role of syntactic structure in children’s sentence comprehension: Evidence from the dative. *Language Learning and Development*, 7(1), 55–75.
- Schwartz, R. G., & Leonard, L. B. (1984). Words, objects, and actions in early lexical acquisition. *Journal of Speech, Language, and Hearing Research*, 27(1), 119–127.
- \* Scott, K., Chu, J., & Schulz, L. (2017). Lookit (part 2): Assessing the viability of online developmental research, results from three case studies. *Open Mind*, 1(1), 15–29.
- \* Scott, R. M., & Fisher, C. (2009). Two-year-olds use distributional cues to interpret transitivity-alternating verbs. *Language and Cognitive Processes*, 24(6), 777–803.
- Stahl, A. E., Romberg, A. R., Roseberry, S., Golinkoff, R. M., & Hirsh-Pasek, K.

- (2014). Infants segment continuous events using transitional probabilities. *Child Development*, 85(5), 1821–1826.
- Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from mandarin speakers' early vocabularies. *Developmental Psychology*, 32(3), 492.
- Tardif, T., Gelman, S. A., & Xu, F. (1999). Putting the “noun bias” in context: A comparison of english and mandarin. *Child Development*, 70(3), 620–635.
- Viechtbauer, W. (2010). Conducting meta-analyses in r with the metafor package. *Journal of Statistical Software*, 36(3), 1–48.
- \* Yuan, S., & Fisher, C. (2009). “Really? She blicked the baby?” Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*, 20(5), 619–626.
- \* Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child Development*, 83(4), 1382–1399.

## Appendix