Quantifying the syntactic bootstrapping effect in verb learning through meta-analysis

Anjie Cao[1] & Molly Y. Lewis[2]

[1] Department of Psychology, Stanford University

[2] Department of Psychology, Carnegie Mellon Unviversity

Author Note

okie

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

Abstract

so abstract!

Keywords: language acquisition, syntactic bootstrapping, meta-analysis

Word count:

Quantifying the syntactic bootstrapping effect in verb learning through meta-analysis

Introduction

In order to become fluent users of language, children must learn the names not only for concrete nouns, like “ball” and “apple”, but also verbs like “eat” and “throw”. Learning the meaning of a verb presents a particular challenge to learners because the perceptual information about a verb’s meaning is complex – it is ephemeral and temporally dynamic. How are children able to overcome this perceptual ambiguity in order to learn verb meanings? One of the central theoretical proposals is that children use syntactic information in a verb’s linguistic context to infer its meaning (Gentner, 2006; Gentner & Boroditsky, 2001), a phenomenon known as “syntactic bootstrapping” (Brown, 1957; Gleitman, 1990; Landau & Gleitman, 1985; Naigles, 1990). For example, if a child hears the sentence “Mom pilked the apple”, the child can use knowledge about the language’s syntax to infer that “pilked” refers to an event where “Mom” is an agent acting on a patient, “the apple”, thereby constraining the hypothesis space of possible referents. A large body of empirical literature over the past twenty years has experimentally tested children’s ability to use syntactic information to infer verb meanings. The goal of the current paper is to synthesize this literature quantitatively using meta-analytic methods in order to evaluate the evidential value of this literature and understand theoretical and methodological moderators of the strength of the effect. The earliest work on syntactic bootstrapping demonstrated that children are able to use coarse-grain syntactic information to constrain hypotheses about a word’s meaning. Broad syntactic categories, like noun and verb, are probabilistically linked to semantic categories, like concrete things and actions. Brown (1957) demonstrated that children are able to make use of this information by presenting children with sentences containing a novel word in different syntactic categories (e.g.,“Do you know what it means to sib?” (verb) vs. “Do you know what a sib is?” (noun)), and measuring whether children inferred the novel word to refer to an action or a concrete entity . Subsequent studies demonstrated that children are also able to use more nuanced syntactic information to infer a word's meaning. In particular, syntactic information provides information about the number of arguments a predicate takes. For example, verbs denoting a causative event (e.g. pushing) tend to take two noun arguments (a "pusher" and a "pushee"), whereas verbs denoting self-generated motion (e.g., waving) take only one noun argument (a "waver"). A large body of work has demonstrated that children are able to successfully map a predicate with two noun arguments (a transitive predicate) to a two-agent causative event, and a predicate with one noun argument (an intransitive predicate) to a non-causative event. In a canonical test of the syntactic bootstrapping phenomenon (Naigles, 1990), children are simultaneously presented with two visual stimuli and an auditory sentence (Fernald, Pinto, Swingley, Weinbergy, & McRoberts, 1998; Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987). One visual stimulus depicts two characters doing an unfamiliar causative action to each other (e.g., a novel pushing motion), and the other depicts one or more characters doing an unfamiliar non-causative action (e.g., a novel waving gesture). Children then hear a transitive (e.g., " "The duck is gorping the bunny") or intransitive sentence (e.g., "The duck and the bunny are gorping") with a novel verb, and then asked to find the novel verb ("Where's the gorping now?"). Children's fixation time or pointing selections are then measured. Evidence for syntactic bootstrapping is found when children look longer at the visual stimulus matching the syntactic structure of the sentence.

The body of empirical literature investigating syntactic bootstrapping has demonstrated that children are able to successfully use syntactic information to bootstrap verb meaning given a range of linguistic (e.g. sentences used differed in their predicate types and word types) and visual stimuli (e.g. visual stimuli ranged from using animations with simple geometrical shapes to videos of human actors), and under a range of methodological implementations (e.g. the inclusion of practice phase, character-identification phase and the number of training-testing pair). These studies provide a binary description of the effect ("there is an effect") with various samples of children under various condition, but do not quantify the *strength* of the effect. Quantifying the strength of the effect allows researchers to assess the degree to which a learning strategy can be used by young learners to acquire verb meanings. Further, quantifying the size of an effect allows for the quantitative comparison to other proposed learning mechanisms for inferring the meaning of a word in a local context. For example, children may be able to infer a word's meaning by using co-occurrence statistics between words and referents ("cross-situational learning"; Smith & Yu, 2008; Yu & Smith, 2007), or by relying on knowledge that each word in a language tends to only have one meaning ("mutual-exclusivity"; Lewis, Cristiano, Lake, Kwan, & Frank, 2020; Clark, 1987; Markman & Wachtel, 1988). While these various learning strategies, and others, could in principle be applied to verbs, there is reason to think that these strategies may be better suited for concrete nouns, where the perceptual information is more stable and less complex. Consistent with this prediction, verbs are typically learned later in development than concrete nouns (e.g. Gentner, 1978; Abbot-Smith, Imai, Durrant, & Nurmsoo, 2017; Childers, Heard, Ring, Pai, & Sallquist, 2012; Childers & Tomasello, 2002, p. @frank2019variability, 2006; Imai et al., 2008; Oviatt, 1980; Schwartz & Leonard, 1984). Understanding the relative strength of various learning strategies for mapping word forms to word meanings could shed light on the source of the developmental lag in verb learning.

In addition to uncertainty about the overall size of the syntactic bootstrapping effect, there are a number of open theoretical questions about the nature of the effect. First, how does the strength of the effect change across development? One possibility is that the ability to use syntactic bootstrapping to learn novel verbs is unlearned or available early on in development as soon as the relevant syntactic information is represented, and the ability does not strengthen with development. An alternative possibility is that the effect becomes stronger with maturation and experience. There are a range of reasons to think that the effect might strengthen. For example, with development, children's vocabulary size increases, making it more likely they will know the nouns in the linguistic input containing a novel verb. If the effect becomes stronger with development, this would suggest that syntactic bootstrapping might become an increasingly powerful learning mechanism for children as they learn more words. [learning to learn stuff] Prior work has examined the syntactic bootstrapping effect across different age groups (e.g. Jin, 2015; Gertner & Fisher, 2012) but it is unclear what the developmental trajectory of this effect looks like. Characterizing this developmental change is important to understanding the role that different word learning strategies play across the development.

Second, how robust is the effect to different predicate types? Syntactic bootstrapping is theorized to be a general learning mechanism that could in principle be applied to the range of syntactic structures in children's input. However, some experiments have found a syntactic bootstrapping effect for transitive sentences, but not intransitive sentences (Arunachalam & Waxman, 2010; Yuan, Fisher, & Snedeker, 2012), whereas others find the opposite pattern (Bunger & Lidz, 2004; Naigles & Kako, 1993). Plausible explanations for both patterns have been proposed. For example, the effect may be stronger for transitive sentences relative to intransitives, because intransitive sentences are relatively more flexible in their usage. Consider, for instance, the sentence "the girl is daxing." This sentence could describe either a scene in which a girl is doing a transitive action (e.g. patting a boy) or a scene in which the girl is doing an intransitive action (e.g. jumping). In contrast, the sentence "the girl is daxing the boy" could only describe a scene in which the girl is doing a transitive action (e.g. patting the boy). This asymmetry may lead children to treat the syntactic information in sentences with transitive novel verbs as a stronger cue to a word's meaning, relative to sentences with an intransitive novel verb. On the other hand, intransitive sentences have fewer processing demands relative to transitive nouns (Lidz, Bunger, Leddon, Baier, & Waxman, 2009), thereby potentially making syntactic bootstrapping easier for intransitive nouns.

The literature has also revealed conflicting findings about the robustness of syntactic bootstrapping to different types of noun phrase structures.In everyday discourse, people and objects are often referred to by pronouns (e.g., “she”) instead of descriptive nouns (“the girl”). There is some evidence in the literature that semantically rich, descriptive nouns are beneficial for verb learning because the semantic content of the surrounding nouns can help scaffold the interpretation of the verbs (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). Others, however, have found that pronouns reduce processing load, thereby facilitating syntactic bootstrapping (Childers & Tomasello, 2001; Lidz et al., 2009). In the case of both predicate types and noun phrase structures, it is difficult to evaluate the robustness of syntactic bootstrapping to variability in the linguistic input without a clear understanding of the empirical pattern

One challenge in addressing these theoretical questions from the existing literature is that people make a range of different methodological decisions. As a result, it is not clear if the differences are due to method, sampling error, or theoretical moderators of interests. For example, researchers would adapt the experimental procedures based on the participants’ age. Some studies included a practice phase or a character identification phase to familiarize the infants with the testing procedure and the agents involved in the events. However, doing so also lengthened the experiment and increased the task demands for infants because they need to stay focused for longer. Moreover, the onset of the linguistic stimuli and the visual stimuli differed across different designs. In some studies, the infants would see the events at the same time or soon after hearing the relevant sentences. In others, the sentences are accompanied by an irrelevant scene (e.g. a person on the phone talking or two people conversing). The lag between the linguistic stimuli and the visual stimuli is likely to increase the memory demand for the infants, and thus influences their performance at test. Finally, the structure of the experimental procedure also differs.

Studies differed in terms of how many training-testing pairs they included. In some designs, the infants were exclusively trained and tested on one single novel verb. In others, they were given multiple training phases and were tested on multiple different novel verbs. Infants differed in the amount of experience they had with the novel verbs prior to the testing. These various implementational differences would further complicate answers to the theoretical questions in the literature.

The meta-analytic method provides a powerful analytic tool for quantifying theoretically important effects, and has increasingly been used in the language acquisition literature (Bergmann et al., 2018). Meta-analysis involves quantifying the size of an effect in individual experiments with a standardized measure (such as Cohen's d), and then aggregating across studies statistically to estimate an overall effect size. Because a meta-analysis reflects estimates from many more participants than an individual study, the meta-analytic method has greater power to detect a true effect and precisely quantify it. The meta-analytic method also allows researchers to explore moderators of an effect. Detecting a moderator to a small effect requires large sample sizes, but infant experiments, like those investigating syntactic bootstrapping, typically have small sample sizes (Bergmann et al., 2018; Oakes, 2017). These small sample sizes mean that individual studies have low power to detect a true moderating effect, leading to literature with many null and conflicting effects for moderating influences. Meta-analysis addresses this limitation by providing a high-powered test of theoretically important moderators factors, as well as revealing variability in effect sizes due to methodological variability across studies.

The plan for the paper is as follows. We first describe our method for conducting a meta-analysis of the syntactic bootstrapping effect. We then assess the evidential value of this literature by evaluating the influence of publication bias. Next, we characterize the size of the effect with respect to other word learning phenomena. We then examine the moderating influence of experience, sentence structure, and noun phrase structure on the syntactic bootstrapping effect, and the relationship between theoretical moderators and methodological variability in our meta-analytic data. We find some evidence for publication bias in the syntactic bootstrapping literature, but that there is nevertheless a small syntactic bootstrapping effect. The effect is larger for transitive sentences, relative to intransitive sentences, but does not appear to be influenced by experience, noun phrase structure, and a range of methodological factors. In the General Discussion, we discuss the implications of our findings for theories of verb learning.

Method

Literature Search

We conducted a literature search of the syntactic bootstrapping literature 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 = 3,339; retrieved between May 2020 and July 2020; Figure 1). We screened for relevance the abstracts of the first 60 pages of the keyword search results (N = 600) and the first 10 pages of the forward search results (N = 100). 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 experts in the field (N = 11). Our sample included published journal articles, conference proceedings, doctoral dissertations, and unpublished manuscripts. These 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 criteria: First, the experimental paradigm involved a two-alternative forced-choice task in which 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 included 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). We included studies with either videos of live actors or animated 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". In contrast, "Look, kradding!" does not provide informative syntactic information. Finally, we restricted our sample to studies with English-speaking, typically-developing children. Papers that satisfied these constraints reflected a range of methodological implementations which we examine analytically below (see moderators). Papers included in our meta-analysis are marked with an asterisks in the references.

Our final sample included 18 individual papers, corresponding to 60 individual effect sizes and 730 unique infants (Mean age: 24 Months 28 Days).

Data Entry

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

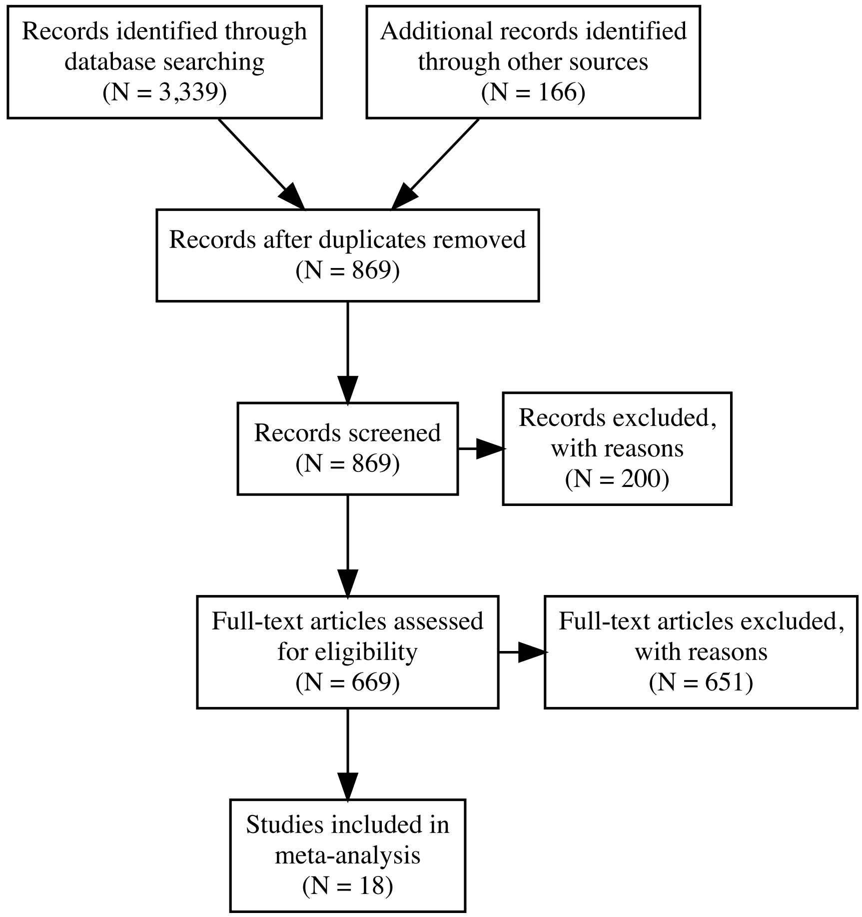


Figure 1. PRISMA plot showing the paper selection process.

Calculating individual effect sizes. In order to calculate each effect size, 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 methods: a) retrieved from the results section or the data-presenting tables (N = 37); b) recovered from the plots (N =10); c) contacted the original authors (N = 10); d) imputed using values from studies with similar designs (N = 3, Hirsh-Pasek, Golinkoff, and Naigles (1996); the missing SD values were imputed from Naigles (1990)). Previous work suggests using imputed values from highly similiar studies improves the accuracy of effect size estimates (Furukawa, Barbui, Cipriani, Brambilla, & Watanabe, 2006). The reported results do not qualitatively change when conditions from Hirsh-Pasek et al. (1996) are excluded from our sample (see SI).

Using the raw coded data, we calculated an effect size estimate for each condition as Cohen's *d*. Cohen's *d* was calculated as the differemce between the proportion correct responses and chance (.5), divided by a pooled estimate of variance (see SI for example calculation). Note that we assume baseline performance to be .5 in all cases, even when an empirical baseline was reported (N = 6 conditions). This analytical decision was made in order to standardize the effect size estimate across all conditions in our sample, most of which did not report an empirical baseline.

Moderators. For each effect size in our sample, we coded several theoretical and methodological variables. The information was either retrieved from the methods section of the paper or obtained by contacting authors.

Four theoretical variables were coded: participant age, participant vocabulary size, predicate type, and noun phrase type. Participant age was entered in mean age in days (N = 60). Vocabulary size was recorded as the median productive vocabulary measured by MacArthur-Bates Communicative Development Inventories (CDI) Words and Sentences (Fenson, 2000; N = 32). Predicate type was coded as either transitive (N = 30) or intransitive (N = 30). Noun phrase type encoded information about the agent verb argument of the sentence stimulus. The agent of the sentence was coded as either as being either noun (e.g., "the girl"; N = 26) or pronoun ("she"; N = 34). A condition was coded as "pronoun" if it contained at least one instance of a pronoun refering to the agent.

In addition to the theoretical variables, we coded a range of methodological variables that varied across the studies in our sample. First, we coded whether the paradigm included a practice trial prior to the testing phase. A study was coded as having a practice trial if there was at least one trial in which children were presented with a familiar verb and asked to identify a familiar action (e.g. "Find jumping"; N with practice phase = 36). Second, we coded whether or not the paradigm involved trials in which children were prompted to identify the nouns in the testing events (e.g., "Where's the bunny?"; N with character identifcation phase = 16). Third, we coded whether the linguistc and visual stimuli were presented synchronously with each other ("Stimuli Synchronicity"). An experimental condition was coded as "asynchronous" if the linguistic stimulus was first paired with an irrelevant visual scene (e.g. a person on the phone talking), and the matching visual stimulus was not shown until the training phase is over ($N = 37$); a condition was coded as "simultaneous" if the very first training sentence was presented along with the visual stimuli depicting relevant action or along with an attention-getter or a blank screen, immediately followed by the relevant action ($N = 5$). Fourth, we coded coded the temporal distribution of the training and the testing trials (Mass: $N = 28$; Distributed: $N = 32$). A procedure was categorized as "mass" if participants were trained exclusively on one novel verb and tested on the same verb. Finally, we coded how many times each novel verb was spoken in a syntactically-informative way during training.[1]

Analytic Approach

We analyzed the data using multi-level random effect models implemented in the metafor package in R (Viechtbauer, 2010). The random effect stucture included groupings by paper and participant group in order to account for the clustering of effect sizes in our sample. The sensitivity analysis was conducted using the PublicationBias package in R (Mathur & VanderWeele, 2020). Moderator variables were included as additive fixed effects. All estimate ranges correspond to 95% confidence intervals, unless otherwise noted. Data and analysis scripts are available at XXXX, and the dataset can be interactively explored at XXXX.

Results

Figure 2 shows effect size estimates for all conditions in our sample. The weighted mean effect size was 0.22 [0.01, 0.43], which significantly differed from 0 (Z = 2.06; p = 0.04). There was evidence for considerable heterogenity in effect sizes across our sample (Q = 196.07; p <.001), meaning that there is unexplained variance in effect sizes aross studies.

Evidential value of the syntactic bootstrapping literature

We first evaluated the evidential value of the literature by assessing the evidence for publication bias. Publication bias refers to the tendency to selectively publish positive findings. The intuition underlying these analyses is that a meta-analysis includes both studies that have statistically significant effect sizes for the target phenomenon and studies those do not. Critically, due to publication pressures, there are likely some studies that have been conducted that are not stastically significant but are not part of our meta-analysis. The absence of these studies from the meta-analysis leads to a meta-analytic estimate that over-estimates the true effect size, and threatens the evidential value of the literature. We present two analyses of publication bias: a classic funnel plot analysis, and a sensitivity analysis that assumes a more plausible model of publication bias.

Figure 3 presents the funnel plot for the effect sizes in our sample. A funnel plot shows estimates of effect size variance (plotted with large values lower on the axis) as a function of the magnitude of the effect size (Egger, Smith, Schneider, & Minder, 1997). Under a model of publication bias in which researchers decide whether or not to publish a study based on the magnitude of its effect size (larger effect sizes being more likely), effect size estimates should fall symmetrically around the grand effect size estimate. Evidence of asymmetry around the grand mean, particularly more large, positive effect sizes, would

[1]See SI for additional methodological moderators.

suggest that the literature reflects a biased sample of studies. A formal test of asymmetry in our sample revealed evidence for asymmetry (Egger's test: Z = -3.55 and 4.62; $p <$ .0001).

The funnel plot analysis provides some evidence for publication bias, but the interpretation of this analysis is limited by the fact that it assumes a relatively implausible model of how researchers decide which studies to make public: the criteria for publishing a study in a journal is typically not the *size* of the effect, as assumed by the funnel plot analysis, but rather whether or not the p-value of the hypothesis test for that effect is below some threshold (usually .05). We therefore conducted a second analysis of publication bias, called a sensitivity analysis (Mathur & VanderWeele, 2020), which assumes that the decision to publish results is determined by the size of the p-value, rather than the magnitude of the effect size.

The goal of the sensitivity analysis is determine how sensitive the meta-analytic effect size is to "missing" non-significant studies. Critically, because the degree of publication is not known (i.e. hthe degree to which significant results are more likely to be published, relative to insignificant results), the sensitivity analysis assumes a worst-case publication bias scenario and estimates the meta-analytic effect size under this scenario. The worst-case scenario assumed by the model is that significant studies are infinitely more likely to be published than non-significant studies.[2] A meta-analytic effect size under this scenario can be estimated by analyzing only those studies with significant effect size estimates.

Conducting this sensitivity analysis on our data reveals that no amount of publication bias could attentuate the point estimate of the effect size to 0. Nevertheless, the worse-case scenario appreciably attenuates the meta-analytic effect size, and the attenuated effect size estimate includes 0 in its 95% confidence interval (0.08 [-0.1, 0.26];

[2]Technically, the model assumes studies with effect sizes that are statistically significant ($p < .05$) *and* greater than zero are infintely more likey to be published. See Mathur and VanderWeele (2020) for additional details.

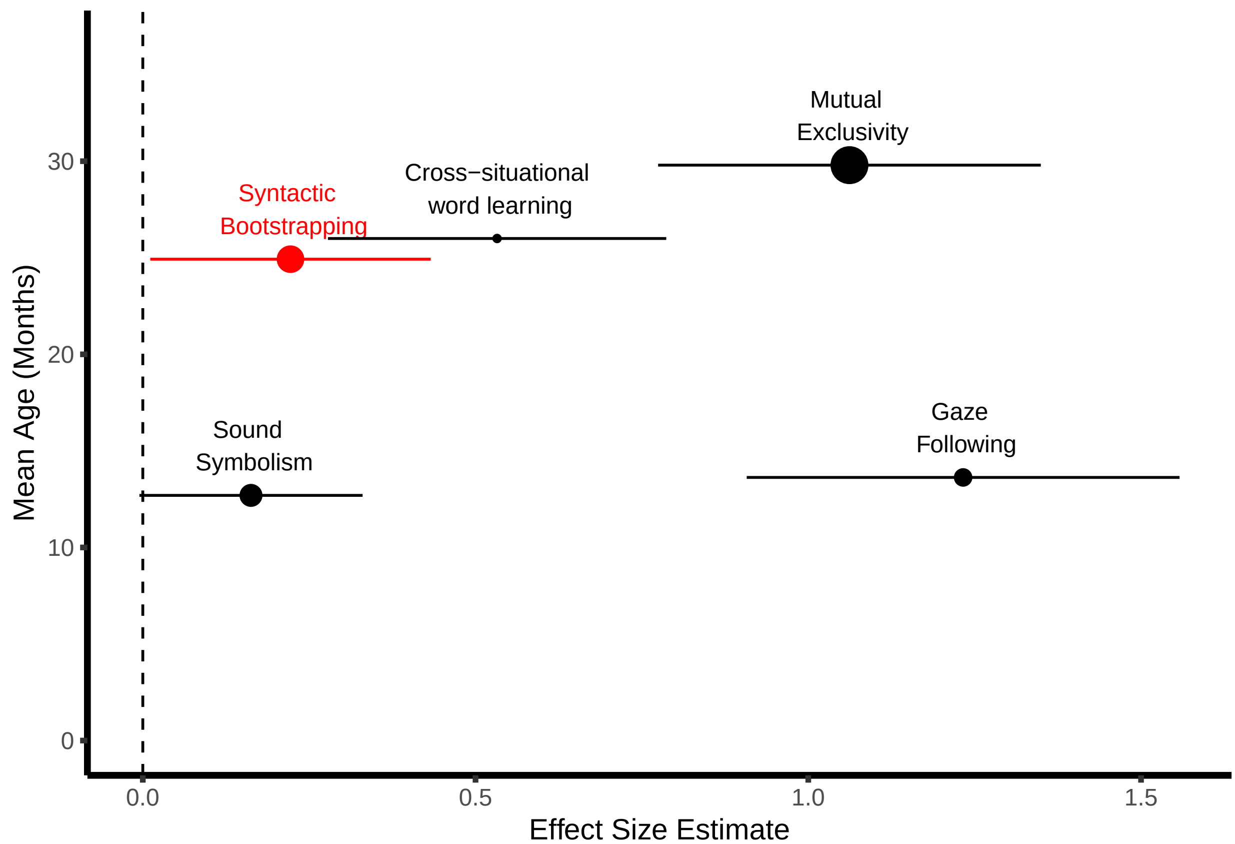
see SI for additional details).

In sum, across two types of analyses, we find some evidence for publication bias in the synactic bootstrapping literature, but even under worst-case scenarios publication bias was not enough to fully attenuate the meta-analytic point estimate to 0. Further, some of the publication bias observed in the funnel plot analysis may be due to heterogeneity in the data. In the following sections, we analyze theoretical and methodological moderators that could be contributing to this heterogeneity, though we emphasize that the likely presence of publication bias implies that these moderators should be interpreted with caution.

Relating the syntactic bootstrapping effect to other word learning stategies

```
## Getting raw MetaLab data from Google Sheets for dataset: Mutual exclusivity
## Getting raw MetaLab data from Google Sheets for dataset: Sound symbolism
## Getting raw MetaLab data from Google Sheets for dataset: Cross-situational word learning
## Getting raw MetaLab data from Google Sheets for dataset: Gaze following
## Getting raw MetaLab data from Google Sheets for dataset: Mutual exclusivity
## Getting raw MetaLab data from Google Sheets for dataset: Sound symbolism
## Getting raw MetaLab data from Google Sheets for dataset: Cross-situational word learning
## Getting raw MetaLab data from Google Sheets for dataset: Gaze following
## Getting raw MetaLab data from Google Sheets for dataset: Mutual exclusivity
## Getting raw MetaLab data from Google Sheets for dataset: Sound symbolism
## Getting raw MetaLab data from Google Sheets for dataset: Cross-situational word learning
## Getting raw MetaLab data from Google Sheets for dataset: Gaze following
## Getting raw MetaLab data from Google Sheets for dataset: Mutual exclusivity
## Getting raw MetaLab data from Google Sheets for dataset: Sound symbolism
## Getting raw MetaLab data from Google Sheets for dataset: Cross-situational word learning
## Getting raw MetaLab data from Google Sheets for dataset: Gaze following
```

```
## Getting raw MetaLab data from Google Sheets for dataset: Mutual exclusivity
## Getting raw MetaLab data from Google Sheets for dataset: Sound symbolism
## Getting raw MetaLab data from Google Sheets for dataset: Cross-situational word learning
## Getting raw MetaLab data from Google Sheets for dataset: Gaze following
```



In order to learn words, children must figure out the mapping relationships between the words and their references. Previous research has proposed numerous biases and learning strategies by which children can solve the mapping problem. Here we relate the syntactic bootstrapping effect to other word learning stratgies. We found its effect size is relatively small compared to the other phenomena in children under 36 months of age. [the other goes into the caption?]

Theoretical Moderators

We next asked whether the overall effect size estimate was moderated by our theoretical moderators of interest: development-related moderators (vocabulary and age), and sentence structure moderators (predicate and noun phrase types).

Development. Does the syntactic bootstrapping effect get stronger across development? We examined two measures of developmental change: age and vocabulary size. These two measures were strongly correlated with each other ($r(30) = 0.85$, p: $<$ 0.0001). Neither age ($\beta = 0$ [0, 0], $SE = 0$, $z =$ -1.64, $p = 0.1$; Fig. 5; 4), nor vocabulary size ($\beta =$ -0.01 [-0.02, 0], $SE = 0$, $z =$ -1.93, $p = 0.05$) significantly moderated the effect.

Sentence structure. We next asked how properties of the sentence struture that children heard influenced the strength of the syntactic bootstrapping effect. Predicate type (transitive vs. intransitive) significantly was a significant moderator, ($\beta = 0.24$ [0.02, 0.46], $SE = 0.11$, $z = 2.13$, $p = 0.03$), with the effect being larger for transitive conditions ($M = 0.49$, $SD = 0.72$) relative to intransitive conditions ($M = 0.17$, $SD = 0.58$). In contrast, there was no effet of agent argument type (pronoun vs. noun; $\beta = 0.14$ [-0.26, 0.53], $SE = 0.20$, $z = 0.69$, $p = 0.49$).

To compare the effects of all theoretical moderators, we fit an additive model with all theoretical variables as fixed effects. We excluded vocabulary size from the additive model since it was highly correlated with age, and only available for a subset of conditions ($N = 32$). 4 shows estimates for each of the single-predictor models along with the additive linear model. The additive model revelaed estimates that were highly comproable to the single-predictor model.

In summary, we found that predicate type is a significant predictor of the effect size. Conditions with transitive sentences were associated with larger effect sizes than those tested with intransitive sentences. No other theoretical variable significantly moderated the effect.

Methodological Moderators

One limiting factor in interpetting the moderating role of theoretical vairables on the syntactic bootstrapping effect is that there was variability across studies in the exact method used in testing children. It is possible that this methodological variability conceals true underlying moderating influences. For example, if researchers adapt their method to the age of the children they are targeting, this might conceal developmental change in the strength of the effect (Bergmann et al., 2018).

We next asked whether the five different methodological variables (practice phase, sentence repetitions, character identifcation phase, synchronicity, testing procedure structure) moderated the syntactic bootstrapping effect. None of these methodological variables were significant moderators of the effect in a single predictor model (Fig. 4; see SI for exact estimates). In an additive linear model with all five methodological predictors, there was a significant effect of testing procedure structure (β = 0.47 [-0.06, 0.99], SE = 0.27, z = 1.74, p = 0.08), with distributed testing designs tending to have larger effect sizes than mass designs. This finding suggests that children tested in a procedure with only one train-test pair performed better than those tested in procedure with multiple train-test pairs. Finally, we asked how these methodological moderators related to our theoretical moderators of interest. Controlling for methodological variables did not qualitiatively change the role of any of the theoretical moderators (see SI) . Taken together, these analyses suggest that methodological variables do not play a large influencing role on the size of the syntactic bootstrapping effect.

General Discussion

- interpret with caution - small es, and there's publication bias!
- 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]
- Transitivity effect is robust - implications for this? Cross-linguistic implications? Text from old intro about multiple possible explanations of this

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 et al., 2017; Childers et al., 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 et al., unpublished). 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 et al., 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 gorping the bunny"; for intransitive: "The duck and the bunny are gorping".) 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; Bunger & 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 et al., 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 et al. (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 et al., 1994; Gillette et al., 1999; Gleitman et al., 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 et al., 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.

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., & 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.

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. *Mechanisms of Language Acquisition. Hillsdale, NJ: Erlbaum*.

Cristia, A. (2018). Can infants learn phonology in the lab? A meta-analytic answer. *Cognition*, *170*, 312–327.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, *315*(7109), 629–634.

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.

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228–231.

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. (unpublished). Variability and consistency in early language learning: The wordbank project.

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.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157.

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.

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.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

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.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420.

* 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.

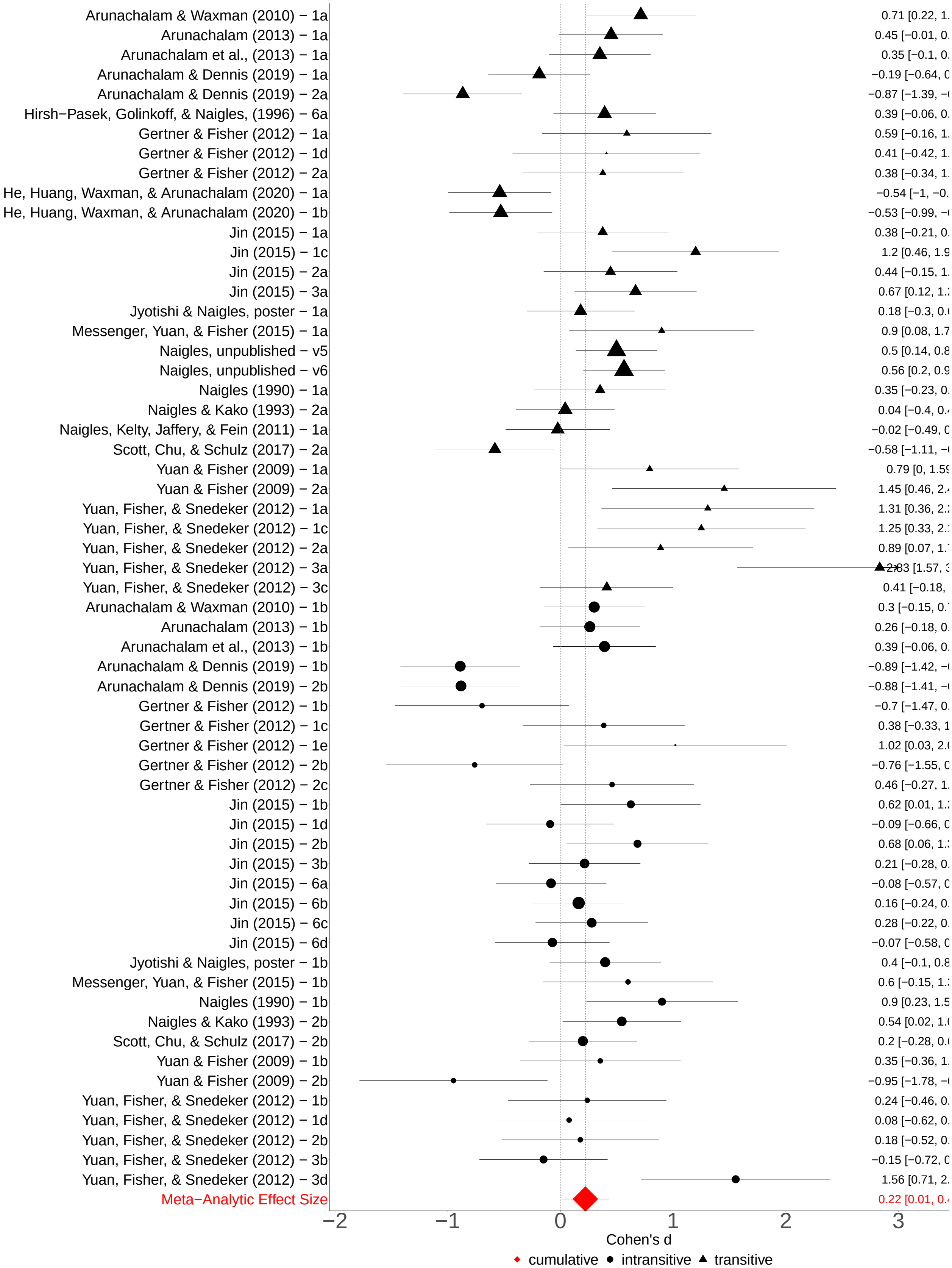

Arunachalam & Waxman (2010) − 1a
Arunachalam (2013) − 1a
Arunachalam et al., (2013) − 1a
Arunachalam & Dennis (2019) − 1a
Arunachalam & Dennis (2019) − 2a
Hirsh−Pasek, Golinkoff, & Naigles, (1996) − 6a
Gertner & Fisher (2012) − 1a
Gertner & Fisher (2012) − 1d
Gertner & Fisher (2012) − 2a
He, Huang, Waxman, & Arunachalam (2020) − 1a
He, Huang, Waxman, & Arunachalam (2020) − 1b
Jin (2015) − 1a
Jin (2015) − 1c
Jin (2015) − 2a
Jin (2015) − 3a
Jyotishi & Naigles, poster − 1a
Messenger, Yuan, & Fisher (2015) − 1a
Naigles, unpublished − v5
Naigles, unpublished − v6
Naigles (1990) − 1a
Naigles & Kako (1993) − 2a
Naigles, Kelty, Jaffery, & Fein (2011) − 1a
Scott, Chu, & Schulz (2017) − 2a
Yuan & Fisher (2009) − 1a
Yuan & Fisher (2009) − 2a
Yuan, Fisher, & Snedeker (2012) − 1a
Yuan, Fisher, & Snedeker (2012) − 1c
Yuan, Fisher, & Snedeker (2012) − 2a
Yuan, Fisher, & Snedeker (2012) − 3a
Yuan, Fisher, & Snedeker (2012) − 3c
Arunachalam & Waxman (2010) − 1b
Arunachalam (2013) − 1b
Arunachalam et al., (2013) − 1b
Arunachalam & Dennis (2019) − 1b
Arunachalam & Dennis (2019) − 2b
Gertner & Fisher (2012) − 1b
Gertner & Fisher (2012) − 1c
Gertner & Fisher (2012) − 1e
Gertner & Fisher (2012) − 2b
Gertner & Fisher (2012) − 2c
Jin (2015) − 1b
Jin (2015) − 1d
Jin (2015) − 2b
Jin (2015) − 3b
Jin (2015) − 6a
Jin (2015) − 6b
Jin (2015) − 6c
Jin (2015) − 6d
Jyotishi & Naigles, poster − 1b
Messenger, Yuan, & Fisher (2015) − 1b
Naigles (1990) − 1b
Naigles & Kako (1993) − 2b
Scott, Chu, & Schulz (2017) − 2b
Yuan & Fisher (2009) − 1b
Yuan & Fisher (2009) − 2b
Yuan, Fisher, & Snedeker (2012) − 1b
Yuan, Fisher, & Snedeker (2012) − 1d
Yuan, Fisher, & Snedeker (2012) − 2b
Yuan, Fisher, & Snedeker (2012) − 3b
Yuan, Fisher, & Snedeker (2012) − 3d
Meta−Analytic Effect Size
0.71 [0.22, 1.
0.45 [−0.01, 0.
0.35 [−0.1, 0.
−0.19 [−0.64, 0
−0.87 [−1.39, −
0.39 [−0.06, 0.
0.59 [−0.16, 1.
0.41 [−0.42, 1.
0.38 [−0.34, 1.
−0.54 [−1, −0.
−0.53 [−0.99, −
0.38 [−0.21, 0.
1.2 [0.46, 1.9
0.44 [−0.15, 1.
0.67 [0.12, 1.
0.18 [−0.3, 0.
0.9 [0.08, 1.7
0.5 [0.14, 0.8
0.56 [0.2, 0.9
0.35 [−0.23, 0.
0.04 [−0.4, 0.
−0.02 [−0.49, 0
−0.58 [−1.11, −
0.79 [0, 1.5
1.45 [0.46, 2.
1.31 [0.36, 2.
1.25 [0.33, 2.
0.89 [0.07, 1.
2.83 [1.57, 3
0.41 [−0.18,
0.3 [−0.15, 0.
0.26 [−0.18, 0.
0.39 [−0.06, 0.
−0.89 [−1.42, −
−0.88 [−1.41, −
−0.7 [−1.47, 0.
0.38 [−0.33, 1
1.02 [0.03, 2.
−0.76 [−1.55, 0
0.46 [−0.27, 1.
0.62 [0.01, 1.
−0.09 [−0.66, 0
0.68 [0.06, 1.
0.21 [−0.28, 0.
−0.08 [−0.57, 0
0.16 [−0.24, 0.
0.28 [−0.22, 0.
−0.07 [−0.58, 0
0.4 [−0.1, 0.8
0.6 [−0.15, 1.
0.9 [0.23, 1.5
0.54 [0.02, 1.
0.2 [−0.28, 0.
0.35 [−0.36, 1.
−0.95 [−1.78, −
0.24 [−0.46, 0.
0.08 [−0.62, 0.
0.18 [−0.52, 0.
−0.15 [−0.72, 0
1.56 [0.71, 2.
0.22 [0.01, 0.
−2
−1
0
1
2
3
Cohen's d
cumulative
intransitive
transitive

Figure 2. wow

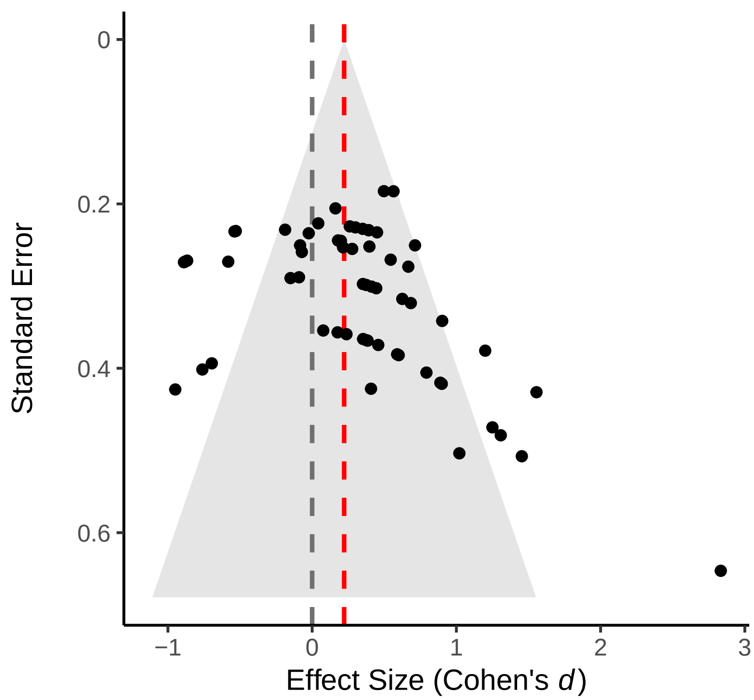


Figure 3. Funnel plot showing the standard error of each effect size estimate in our meta-analysis as a function of the magnitude of that effect size. The gray and red vertical dashed lines correspond to an effect size of zero and the meta-analytic effect size estimate, respectively. The grey funnel represents a 95% confidence interval around meta-analytic estimate. In the absence of publication bias, effect size estimates should by symmetrically distributed around the red line.

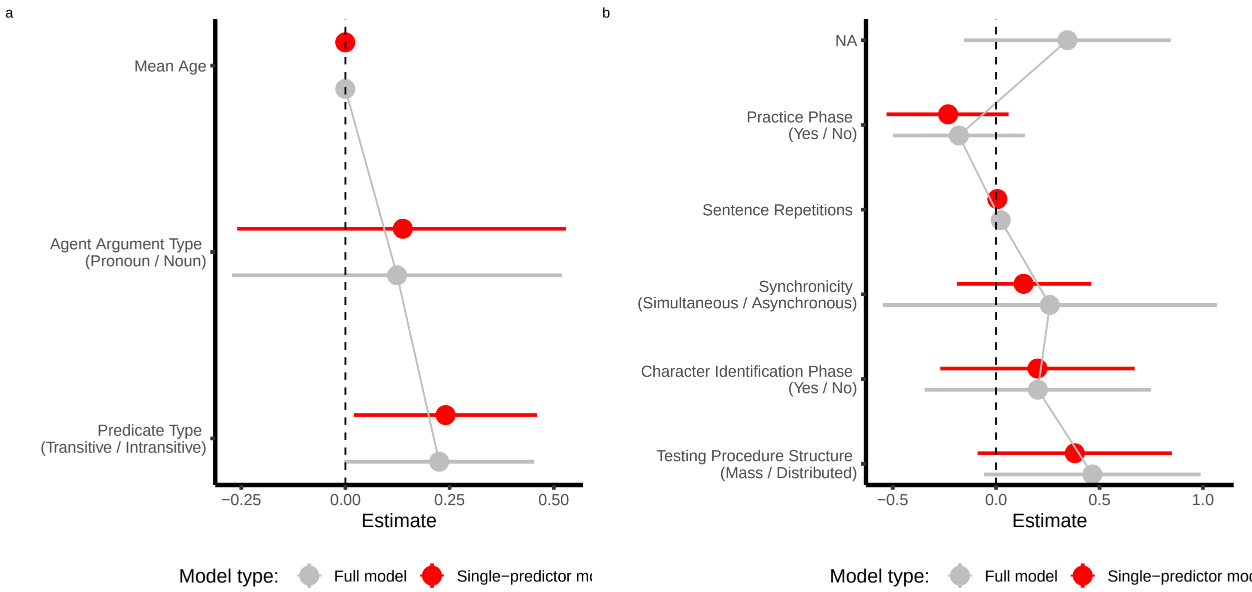


Figure 4. In these figures we presented moderator estimates from two types of model: the red dots represent moderator estimates from single-predictor model, and the gray dots connected with each other represent moderator estimates from the full model including all the moderators from the group. For categorical variables, the dots represent the level that appears first in the parentheses. We found that each theoretical moderator and methodological moderator has similar estimates in the single-predictor models and the full models. (a) Predicate type is a significant predictor for effect size. In particular, transitive sentence has a positive effect on the effect size relative to intransitive sentence. Note that in this plot we did not include median productive vocabulary size. This is because only a subset of studies reported the information. (b) Testing structure and sentence reptitions marginally predict..

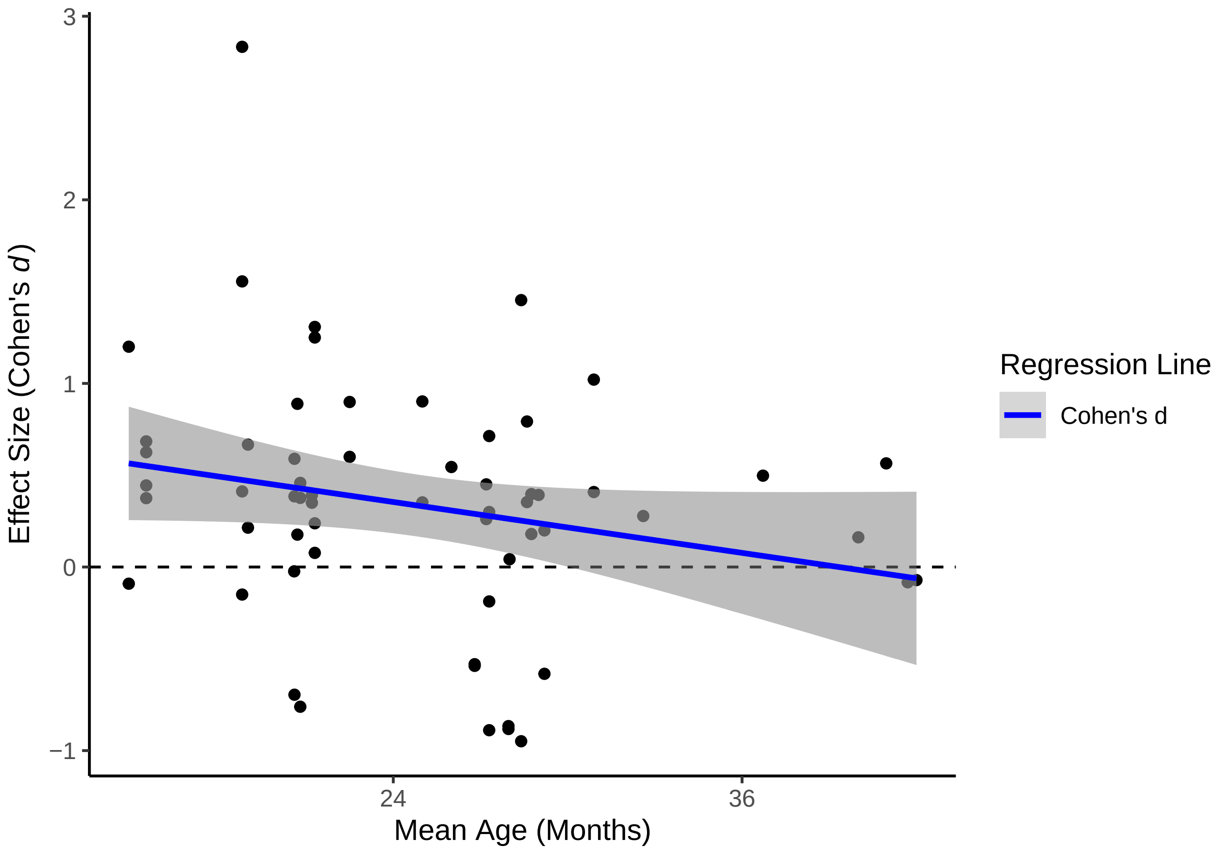


Figure 5. Effect size (Cohen's d) as a function of age in months. The dash line corresponds to effect size equals zero, the solid blue regression line is a simple linear model fit, and the shaded area corresponds to the standard error of the fitted model. Each dot represents one calcaulated effect size.