Quantifying the syntactic bootstrapping effect in verb learning: A meta-analytic synthesis

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

[1] Department of Psychology, Stanford University

[2] Department of Psychology, Carnegie Mellon Unviversity

Abstract

How do children infer the meaning of a novel verb? One prominent proposal is that children rely on syntactic information in the linguistic context, a phenomenon known as "syntactic bootstrapping" (Naigles, 1990). For example, given the sentence "The bunny is gorping the duck", a child could use knowledge of English syntactic roles to infer that "gorping" refers to an action where the bunny is acting in some way on a duck. Here, we examine the strength of the syntactic bootstrapping effect, its developmental trajectory and generalizability using meta-analytic methods. Despite its theoretical prominence, we find that the syntactic bootstrapping is relatively small (d = .24), though significantly different from zero. The effect is comparable in size to cross-situational learning and sound symbolism but smaller than mutual-exclusivity and gaze-following. Further, we find that the effect does not strengthen with development, and is present only for studies using transitive sentences. An examination of a range of methodological factors suggests that the effect is not strongly influenced by methodological implementation. In the General Discussion, we discuss implications of our findings for theories of verb learning and make recommendations for future research.

Keywords: language acquisition, syntactic bootstrapping, meta-analysis

Word count:

Quantifying the syntactic bootstrapping effect in verb learning: A meta-analytic synthesis

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 temporally dynamic and relational (Gentner, 2006). How are children able to overcome these complexities 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 meaning, 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”, thus significantly constraining the hypothesis space of possible meanings for the novel verb. A large body of empirical literature over the past thirty 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 in word learning 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 refered to an action or a concrete entity. Subsequent studies demonstrated that children are also able to use more nuanced syntactic information to infer the meaning of novel words. In particular, the number of arguments a predicate takes can be used to infer what that predicate refers to. 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 one-agent non-causative event.

The ability to use syntactic information to infer a novel verb meaning is typically tested in a paradigm where children are simultaneously presented with two visual stimuli and an auditory sentence (Fernald, Pinto, Swingley, Weinbergy, & McRoberts, 1998; Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987). Canonically (e.g., Naigles, 1990), 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") containing a novel verb, and then are asked to find the corresponding scene ("Where's gorping now?"). Children's fixation time or pointing selections are measured. Evidence that children are able to "bootstrap" verb meanings from syntax is found when they look longer at the visual stimulus matching the syntactic structure of the sentence. These studies have found that children are able to use syntactic information to infer verb meaning given a range of linguistic and visual stimuli, and in a variety of experimental manipulations.

Importantly, however, while these studies provide a binary description of the effect ("children can bootstrap verb meanings from syntax"), they 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, it 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. Bates et al., 1994; Frank, Braginsky, Marchman, & Yurovsky, in press). 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 strength 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 (Fisher, Jin, & Scott, 2020; Gleitman, 1990). 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. 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 stages of language acquisition.

Second, how robust is the effect to different syntactic structures? 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 a transitive novel verb as a stronger cue to a verb'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. Therefore, although syntactic bootstrapping is thought to be a general verb learning mechanism, it is an open question the degree to which children able to learn verb meanings from different types of syntactic structures.

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 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, 2015; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005). 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 is the range of methodological implementations of tests of the syntactic bootstrapping phenomenon across the literature. The presence of methodological variability means that it is not clear if observed differences in outcomes are due to methodological variability, sampling error, or theoretical moderators of interests. For example, across studies, there is variability in the relative onset of the linguistic and visual stimuli. In some studies, children see the events at the same time or soon after hearing the relevant sentences (Gertner & Fisher, 2012; Naigles, 1990), while in others, the sentences are accompanied by an irrelevant scene (e.g. a person on the phone talking or two people conversing) followed by the target visual stimuli (Arunachalam & Waxman, 2010; Yuan & Fisher, 2009). The lag between the linguistic stimuli and the visual stimuli likely increases the memory demands of the task, and could influence children's ability to identify the correct referent. Critically, if this methodological difference co-varies with the age of children tested, it may be difficult to draw conclusions about the developmental trajectory of the effect across studies.

The meta-analytic method provides a powerful analytic tool for quantifying theoretically important effects, and has increasingly been applied in the language acquisition literature (Bergmann et al., 2018; Cristia, 2018; Lewis et al., 2020; Rabagliati, Ferguson, & Lew-Williams, 2019). 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 experiments statistically to estimate an overall effect size. Because a meta-analysis reflects estimates from many more participants than any 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, potentially leading to a literature with many null and conflicting effects for moderating influences. Meta-analysis addresses this limitation by providing a higher-powered test of theoretically important 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, predicate type, 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 through 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 the abstracts for relevance 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 et al., 2020) and experts in the field (N = 11). Our sample included published journal articles, conference proceedings, doctoral dissertations, and unpublished manuscripts. We refer to these collectively as "papers" in the following sections.

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. One event depicted a causative action (e.g., one agent causes the other to move), and the other a 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

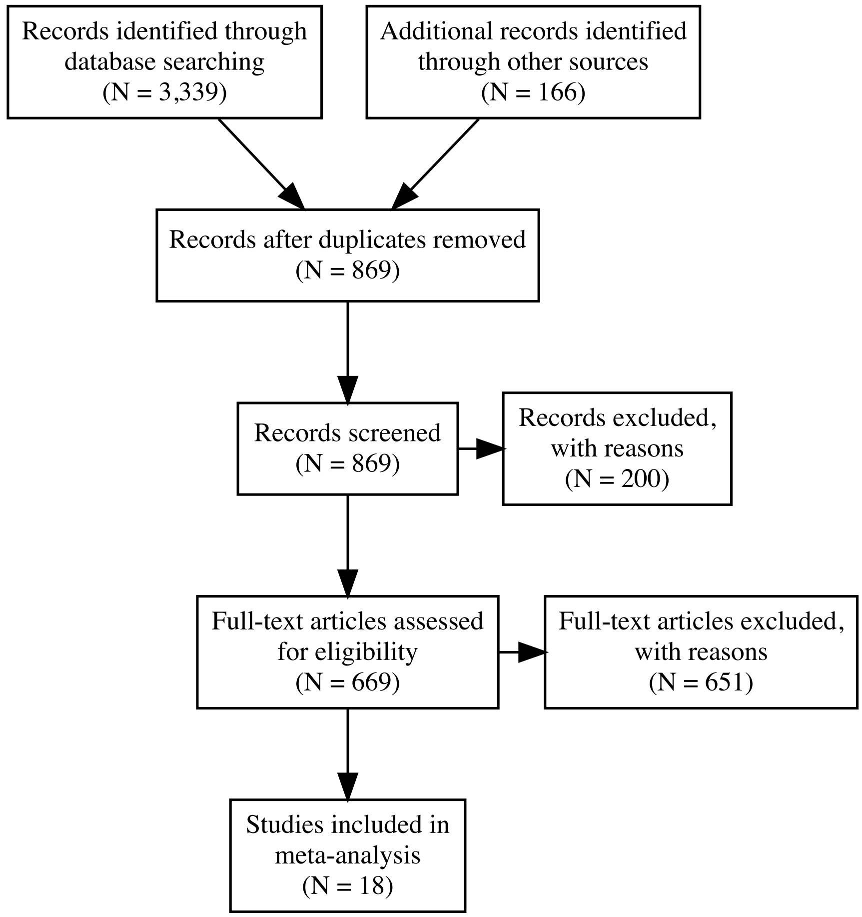

Figure 1. PRISMA plot showing literature review process. Values indicate number of papers at each stage of the review review process. Our meta-analysis included a final sample of 18 papers.

"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 systematically below (see Moderators section). Our final sample included 18 papers, indicated by an asterisks in the references.

Data Entry

For each paper, we entered metadata about the paper (e.g., citation), information to calculate effect sizes, and information about moderators. We entered a separate effect size for each experimental manipulation and age group per paper ("condition"). Most papers therefore contained multiple conditions, corresponding to multiple effect sizes in our meta-analysis. Our final sample included 60 conditions.

Calculating individual effect sizes. For each condition, we recorded the sample size, the mean proportion correct responses, and the across-participant standard deviation of proportion correct responses. The mean and standard deviation were obtained from one of four sources: a) text or tables in the results section (N = 37); b) plots (N =10); c) crrespondence with the original authors (N = 12); d) imputation using values from studies with similar designs (N = 1, Hirsh-Pasek, Golinkoff, and Naigles (1996); the missing standard deviation values were imputed from Naigles (1990)). Previous work suggests using imputed values from highly similar 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, Sec.4).

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 difference between the proportion correct responses and chance (.5), divided by a pooled estimate of variance (see SI, Sec.1 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 retrieved either from the methods section of the paper or 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 months (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 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 referring to the agent.

In addition to the theoretical variables, we coded a range of methodological variables that varied across the studies in our sample and for which there was independent reason to think they could influence the size of the effect. 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 identification phase = 16). Third, we coded whether the linguistic 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 the relevant action or along with an attention-getter or a blank screen, immediately followed by the relevant action ($N = 23$). The synchronicity between the linguistic stimuli and the visual stimuli could potentially scaffold the mapping process. Fourth, we coded the temporal distribution of the training and the testing trials (Mass: $N = 28$; Distributed: $N = 32$). The temporal distribution is linked to the amount of learning experience children have prior to the test. A procedure was categorized as "mass" if participants were trained exclusively on one novel verb and tested on the same verb, and "distributed" if they were trained and tested on multiple novel verbs. 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 structure 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

Our final sample of 60 conditions reflected 730 unique infants (mean age: 24 months; 28 days), with a mean sample size of 14.15 (SD = 6.16) children per condition. Figure 2 shows effect size estimates for all conditions. The weighted mean effect size was 0.24 [0.03, 0.44], which significantly differed from 0 ($Z = 2.27$; $p = 0.02$). There was evidence for considerable heterogeneity in effect sizes across our sample ($Q = 196.07$; $p < .001$), meaning that there is unexplained variance in effect sizes across studies.

Evidential value of the syntactic bootstrapping literature

We first evaluated the evidential value of the literature by assessing the evidence for publication bias. The intuition underlying these analyses is that, due to random variation, a literature should be expected to contain studies both with and without statistically

[1] See SI, Sec.6 for additional methodological moderators. These additional moderators overlap substantially with the target moderators of interest presented in the Main Text.

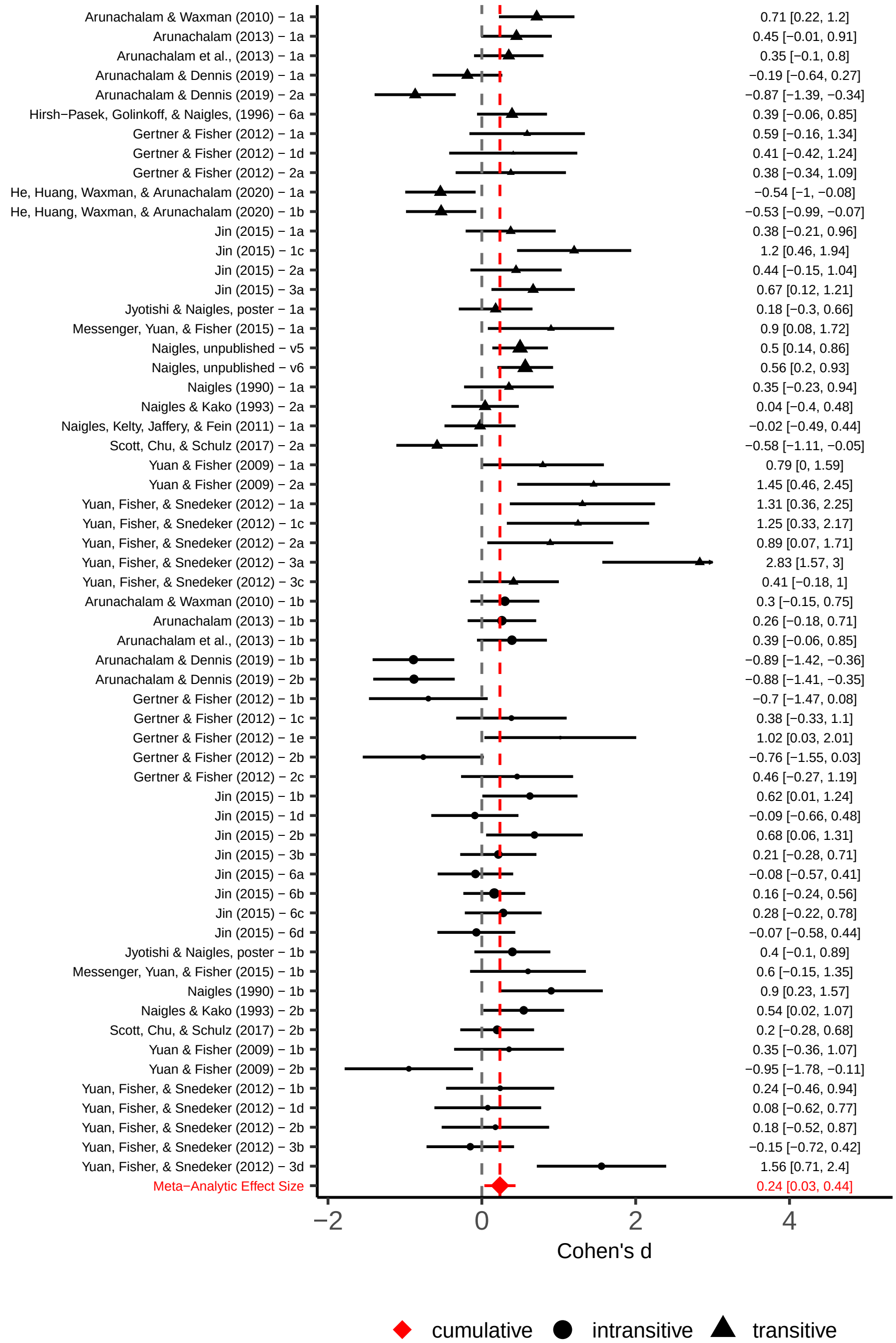

Figure 2. Forest plot showing ~~each of~~ the individual effect size included in the meta-analysis. Black circles and triangles correspond to individual conditions with transitive sentences and intransitive sentences, respectively. Point size corresponds to study weights, and horizontal error bars show 95% confidence intervals. The red diamond at the bottom is the meta-analytic effect size aggregated across all conditions included in the literature.

significant effect sizes for the target phenomenon. Critically, however, publication pressures may lead researchers to be more likely to publish findings with statistically significant

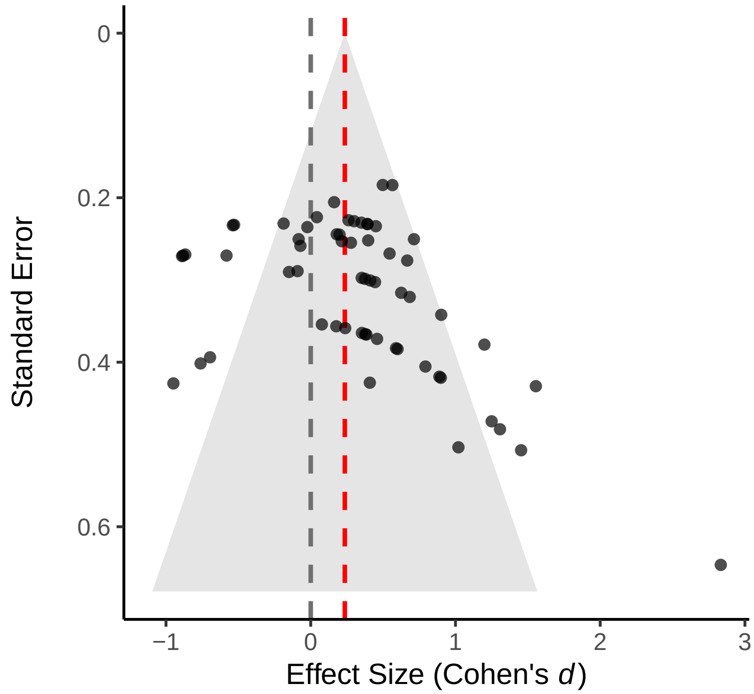

Figure 3. Funnel plot showing the standard error of each effect size estimate in our meta-analysis as a function of the magnitude of that effect size. The gray and red vertical dashed lines correspond to an effect size of zero and the meta-analytic effect size estimate, respectively. The grey funnel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

results, resulting in a biased literature. The absence of these studies from the meta-analysis yields a meta-analytic estimate that over-estimates the true effect size, and threatens the evidential value of the literature. We present two analyses that assess publication bias in the syntactic bootstrapping literature: a classic funnel plot analysis, and a sensitivity analysis that assumes a more plausible model of the publication process.

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 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: $Z = 4.72$; $p < .0001$).

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 to determine how sensitive the meta-analytic effect size is to "missing" non-significant studies. Critically, because the degree of publication bias is not known (i.e. the 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 non-significant effect size estimates.

Conducting this sensitivity analysis on our data reveals that no amount of publication bias could attenuate the point estimate of the effect size to 0. Nevertheless, the

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

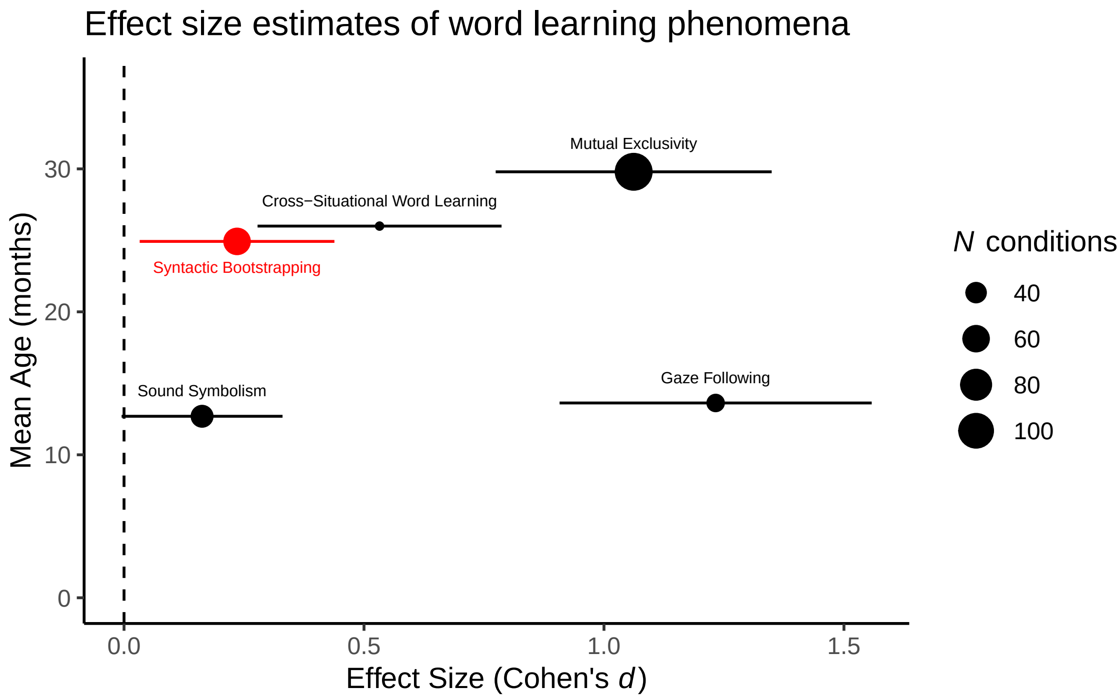

Figure 4. Meta-analytic effect sizes of five word learning phenomena, including syntactic bootstrapping (red). Point size corresponds to the number of individual conditions included in each meta-analysis. The x-axis shows the magnitude of the meta-analytic effect size estimate; the y-axis shows the mean age in months of children in each meta-analysis.

worst-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, 0.25]; see SI, Sec.2 for additional details).

In sum, across two types of analyses, we find some evidence for publication bias in the syntactic 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 may contribute 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 strategies

How does the strength of the syntactic bootstrapping effect compare to that of other word learning strategies? To answer this question, we compared the meta-analytic syntactic bootstrapping effect size to effect sizes for other word learning strategies estimated from a meta-analysis of each literature. We considered an opportunity sample of word learning strategies, based on those strategies with available meta-analytic data. In particular, we selected all word learning strategies available in a database of language acquisition meta-analyses, called Metalab (Bergmann et al., 2018). We included a word learning strategy in our analysis if it could be considered to facilitate an inference about the mapping between a novel word and a meaning. This allowed for the comparison of the syntactic bootstrapping effect to four additional word learning strategies: (i) mutual exclusivity (Clark, 1987; Lewis et al., 2020; Markman & Wachtel, 1988), assuming a novel word refers to a novel object, (ii) cross-situational word learning (Yu & Smith, 2007), tracking word-object co-occurrences across situations, (iii) gaze following (Frank, Lewis, & MacDonald, 2016; Scaife & Bruner, 1975), following the eye gaze of a speaker to the intended referent, and (iv) sound symbolism (Fort et al., 2018), exploiting sound-meaning regularities in the lexicon. While these four strategies are not exhaustive of the strategies that have been proposed in the word learning literature, they are representative of the major theoretical perspectives, including constraints and biases (Markman, 1990), statistical learning (Romberg & Saffran, 2010), and communicative inferences (Tomasello, 2010). For each of these four comparison strategies, we calculated the meta-analytic effect size using the same model specification as for syntactic bootstrapping, restricting the sample to studies with a mean age of children younger than 48-month-olds.

Figure 4 shows the meta-analytic effect size for syntactic bootstrapping and each of the other four word learning strategies. The syntactic bootstrapping effect size (0.24 [0.03, 0.44]) was comparable in size to that of sound symbolism (d = 0.16 [-0.01, 0.33]; Mean age:

12.70 Months) and cross-situational learning (d = 0.53 [0.28, 0.79]; Mean age: 26 Months), and less than a quarter of the size of both mutual exclusivity (d = 1.06 [0.77, 1.35]; Mean age: 29.80 Months) and gaze following (d = 1.23 [0.91, 1.56]; Mean age: 13.60 Months). Importantly, the small effect size of syntactic bootstrapping relative to mutual exclusivity and gaze following cannot be due alone to differences in the ages of the samples in these different meta-analyses, since participants in the syntactic bootstrapping meta-analysis were older on average than those in the gaze following meta-analysis, and roughly the same age as those in the mutual exclusivity meta-analysis.

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. How does the strength of the syntactic bootstrapping effect change across development? We examined two measures of developmental change: age (months) and vocabulary size. These two measures were strongly correlated with each other ($r(30)$ = 0.85, p <.0001). There was no effect of age (β = -0.01 [-0.03, <.001], SE = 0.01, z = -1.50, p = 0.13; Fig. 5) or vocabulary size (β = -0.01 [-0.02, <.001], SE 0.01, z = -1.74, p = 0.08; Fig. 6a).

Sentence structure. We next asked how properties of the sentence structure influenced the strength of the syntactic bootstrapping effect. Predicate type (transitive vs. intransitive) was a significant moderator, (β = 0.24 [0.02, 0.46], SE = 0.11, z = 2.10, p = 0.04): the effect was larger for transitive sentences, relative to interanstive sentences. Further, the model intercept did not significanlty differ from zero (β = 0.1 [-0.14, 0.34], z = 0.80, p = 0.42), suggesting that the effect is only present in transitive conditions (M = 0.49, SD = 0.72) but not in intransitive conditions (M = 0.17, SD = 0.58). In contrast, there was no effect of agent argument type (pronoun vs. noun; β = 0.16 [-0.21, 0.54], SE =

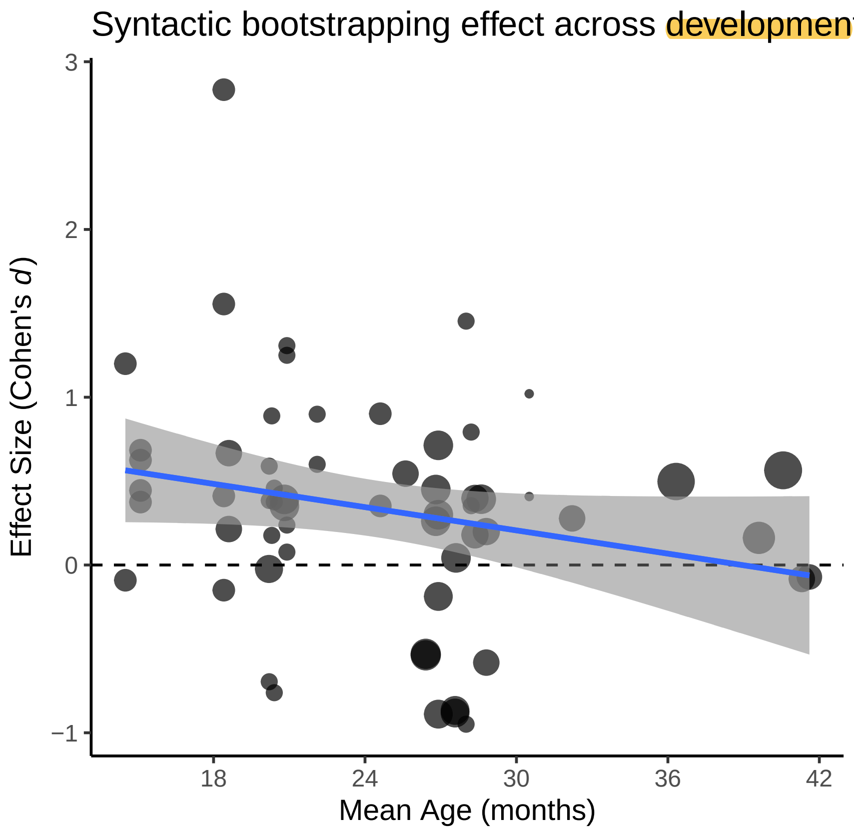

Figure 5. Syntactic bootstrapping effect size (Cohen's d) as a function of age in months. Each point corresponds to one effect size (condition), and point size corresponds to the number of children in that condition. The blue line shows a linear model fit and the corresponding standard error. The dashed line indicates an effect size of zero. The slope of the model fit does not significantly differ from zero, suggesting no appreciable developmental change in the size of the syntactic bootstrapping effect.

0.19, $z = 0.85$, $p = 0.4$).

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). Figure 6 shows estimates for each of the single-predictor models along with the additive linear model. The additive model revealed estimates that were highly comparable to the single-predictor model.

In summary, we find that predicate type is a significant predictor of the effect size:

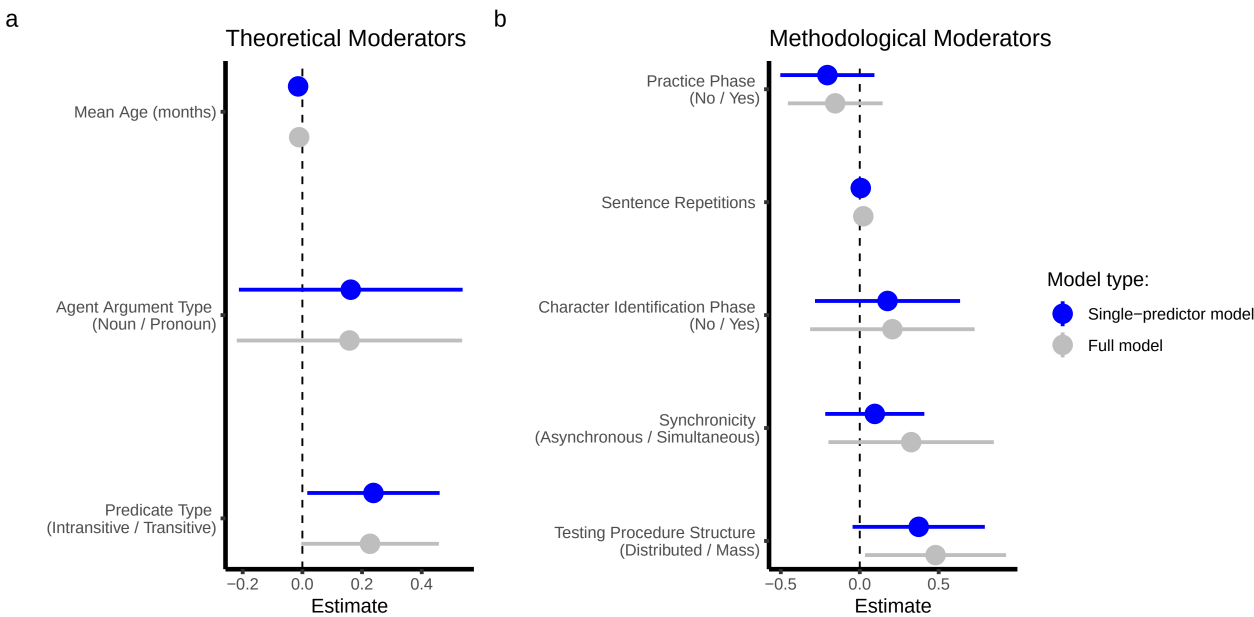

Figure 6. Meta-analytic models parameter estimates for (a) theoretical and (b) methodological moderators. Blue points show model estimates from single-predictor model; grey points show model estimates from additive linear model with all moderators included. Ranges correspond to 95% confidence intervals. Levels for categorical variables are given in parentheses, with the first level indicating the base level in the model.

conditions with transitive sentences were associated with larger effect sizes than those tested with intransitive sentences. No other theoretical variable significantly moderated the syntactic bootstrapping effect.

Methodological Moderators

One limiting factor in interpreting the moderating role of theoretical variables is that there was appreciable 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, developmental change in the strength of the effect may not be detectable (Bergmann et al., 2018).

To evaluate this possibility, we asked whether five different methodological variables (practice phase, sentence repetitions, character identification 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. 6; see SI, Sec.3 for exact estimates). In an additive linear model with all five methodological predictors, there was a significant effect of testing procedure structure (β = 0.48 [0.03, 0.93], $SE = 0.23$, $z = 2.11$, $p = 0.04$), with mass testing designs tending to have larger effect sizes than distributed designs. This finding suggests that children tested in a procedure with only one train-test pair performed better than those tested in a procedure with multiple train-test pairs. Finally, we asked how these methodological moderators related to our theoretical moderators of interest. Notably, controlling for methodological variables did not qualitatively change the role of any of the theoretical moderators (see SI, Sec.5). 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

Three decades of research on syntactic bootstrapping have examined the role that syntax plays in facilitating early verb acquisition. Here we built upon these previous studies by presenting a quantitative synthesis of the literature using meta-analytic methods. We find a small effect of syntactic bootstrapping, comparable in size to that of sound-symbolism and cross-situational learning. We then used the meta-analytic method to examine how the strength of the syntactic bootstrapping effect varies as a function of developmental change and syntactic structure. We found no evidence that the strength of the syntactic bootstrapping effect changes across development. In contrast, we found some evidence that the strength of the effect is influenced by predicate type: the syntactic bootstrapping effect is present for transitive sentences, but not intransitive sentences. We also examined a range of methodological features and found no evidence that these features had an appreciable moderating influence on the size of the syntactic bootstrapping effect.

Small effect size

The syntactic bootstrapping effect ($d = 0.24$ [0.03, 0.44]) is small by conventional Cohen's d standards (Cohen, 1988), even in the presence of some publication bias. It is also small relative to other early word learning biases and strategies typically studied with nouns, like mutual exclusivity and gaze-following. On the one hand, this small effect size is consonant with the fact that verbs tend to be learned later in development than nouns: If verb learning strategies are weaker than those for nouns, we would expect verbs to be learned later. On the other hand, children do eventually develop a large vocabulary of verbs, despite no developmental change in the strength of the syntactic bootstrapping effect. This raises an important puzzle: If syntactic bootstrapping does not play a prominent role in helping children learn verbs, then how do children learn them?

One possibility is that children learn verbs using many of the same strategies typically thought of as noun-learning strategies. Indeed, while word learning is typically studied with nouns, there is some evidence that a wide range of word learning strategies may be helpful in verb learning, in addition to syntactic bootstrapping. Mutual exclusivity (Golinkoff, Jacquet, Hirsh-Pasek, & Nandakumar, 1996; Merriman, Evey-Burkey, Marazita, & Jarvis, 1996; Merriman, Marazita, & Jarvis, 1993), cross situational learning (Monaghan, Mattock, Davies, & Smith, 2015; Scott & Fisher, 2012), sound symbolism (Imai, Kita, Nagumo, & Okada, 2008; Kantartzis, Imai, & Kita, 2011) and the use of social cues (Roseberry, Hirsh-Pasek, Parish-Morris, & Golinkoff, 2009; Tomasello & Barton, 1994) have all been found to facilitate the mapping between verbs and actions. Importantly, these learning strategies, including syntactic bootstrapping, are not mutually exclusive with one another: children may be using a combination of strategies to varying degrees in different contexts and at different ages. Understanding how these different strategies are flexibly used in verb learning is an important area for future research.

No developmental change

Notably, we do not find the syntactic bootstrapping effect gets larger with age. This finding is consistent with the proposal that syntactic bootstrapping is an unlearned bias and that the accumulation of experience with age does not have a significant impact on the strength of the effect (Fisher et al., 2020; Gleitman, 1990). An alternative explanation of the lack of developmental change is the relational nature of verbs and grammatical constructions (Gentner, 2006; Goldwater, 2017). It is possible that the developmental change would only emerge as children experience the "relational shift", that is, a period of time that they gradually shift their attention from objects to the relations between objects (Gentner, 1988; Gentner & Rattermann, 1991). Moreover, there is some evidence to suggest that older children are more likely to rely on syntactic information to infer verb meaning: Nappa, Wessel, McEldoon, Gleitman, and Trueswell (2009) found that five-years-old children were more likely than three-years-old children to rely on syntactic cues relative to socio-pragmatic cues when both sources of information were present. It is therefore possible that although no developmental change was observed within the age range of the current meta-analysis, syntactic information may play a more prominent role in verb learning later on in development.

A final possibility is that researchers make methodological adaptations for children of different ages, such that older children are tested in more challenging paradigms than younger children. Under this possibility, we do not observe developmental increase in the size of the effect because of children's increasing ability to use syntactic information to infer meaning is concealed by more challenging task demands. We examined this hypothesis in the current meta-analysis by testing whether a wide range of methodological factors interacted with age, and found no evidence that they did. We also informally examined how the complexity of the visual stimuli used in the studies in our meta-analysis varied as a function of age (See SI, Sec. 7). With the exception of one study designed for

15-month-olds (Jin, 2015), there was no observable variability in visual stimuli complexity as a function of the age of children tested. While it is of course possible that there is a relevant, unmeasured methodological factor, the present set of analyses are suggestive that the lack of developmental change in the strength of effect is not due to methodological factors.

Transitivity effect

Our meta-analysis revealed that the syntactic bootstrapping effect is only observed with certain kinds of syntactic structures: children tend to select the correct novel action in conditions using transitive sentences (e.g. “The girl is gorping the boy”), but not in conditions using intransitive sentences (e.g., “The girl is gorping”). This pattern is consistent with the observations that several researchers have made (Arunachalam & Waxman, 2010; Yuan et al., 2012). Three factors may have given rise to this transitivity advantage. First, under certain contexts, the transitive sentences are less ambiguous than intransitive sentences. 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). Third, the scenes corresponding to transitive sentences may have been more perceptually salient than those in intransitive conditions. Children were found to have a baseline preference for two actors events over one actor event (e.g., Yuan et al., 2012), and for synchronous movement over causative movement (e.g., Naigles & Kako, 1993). These preferences may make detecting an effect in the transitive condition easier than in the intransitive condition.

Importantly, the lack of effect in intransitive sentences calls into question on the generalizability of syntactic bootstrapping effect: What other syntactic structures can syntactic bootstrapping be extended to? Can it be applied to the learning of adjectives and adverbs, which are two syntactic categories similarly abstract like verbs? Alternatively, is syntactic bootstrapping only plausible in transitive sentences? If that is the case, then what features of transitive sentences make them uniquely conducive to verb learning?

Limitations and recommendations for future works

Like any method, meta-analysis has a number of limitations. Most notably, the meta-analytic method is influenced by publication bias of the literature it draws upon. In the present study, we find some evidence for publication bias in the syntactic bootstrapping literature in two different analyses, suggesting that there may be a number of “missing”, null or negative studies in our meta-analysis potentially leading to an overestimation of the overall effect size. Although our sensitivity analysis suggests that the effect size can not be reduced to zero even after assuming the “worst-case scenario”, the presence of publication bias limits the evidential value of the current literature and suggests that the magnitude of the effect should be interpreted with caution. Further, the relatively small effect size also means that large sample sizes are needed to detect moderating factors, meaning that some of the null effects found in our meta-analysis could be the result of insufficient power.

Future research could address the issue of publication bias in several ways. One way is for researchers to conduct studies with substant[illegible] larger sample sizes. The mean sample size of studies in the meta-analysis was 14. Based on our meta-analytic effect size estimate, we estimate that studies in our sample had approximately 0.10 power to de[illegible] a true effect – much smaller than the typical target of 0.8. In order to reach 0.8 power, we estimate that researchers would need about 142 participants per condition (substantially more if moderators were tested; SI sec. 9). Sample size is important because conducting underpowered studies not only increases the false negative rate (Type II error), it also inflates the false positive rate (Type I error; Button et al., 2013; Oakes, 2017). And, critically, with more false positives, there are more opportunities for publication pressures to influence the literature. Increasing sample size in studies designed to test hypotheses about syntactic bootstrapping would thus reduce the rate of false positives in the literature and decrease publication bias, leading to a more robust estimate of the aggregate effect size. Second, publication bias could be improved by pre-registering study hypotheses, designs, and analytical methods (Nosek, Ebersole, DeHaven, & Mellor, 2018). Pre-registration is a relatively low-cost way for researchers to distinguish both for themselves and for other researchers which decisions in the research process were made before and after seeing the data. By making this distinction explicit, researchers are less likely to inflate the false positive rate through analytical flexibility. Substantial insight into the robustness and strength of the syntactic bootstrapping phenomenon could be therefore be gained through a pre-registered, high-powered direct replication of the seminal syntactic bootstrapping studies across one or more labs (e.g., ManyBabies Consortium, 2020).

A second limitation of the current meta-analysis is the narrow scope of participants in our sample: English-speaking, typically developing children. This homogeneity limits the extent to which our study can shed light on the generalizability of syntactic bootstrapping to other populations. For example, English relies strongly on word order to signal syntactic relations, raising the possibility that children learning languages with more explicit morphosyntactic markers could rely more strongly on syntactic information to infer verb meanings. There are, notably, an increasing number of studies that examine the syntactic bootstrapping effect in children learning a diverse set of languages (e.g. Mandarin: Lee & Naigles, 2008; Korean: Jin, 2015; Turkish: Göksun, Küntay, & Naigles, 2008) and in children with a range of developmental disorders (e.g. Naigles, Kelty, Jaffery, & Fein, 2011; Grela, 2002; O'Hara & Johnston, 1997). As this literature grows, additional studies can be contributed to the current meta-analysis to systematically examine how differences in language and developmental trajectory interact with the size of the syntactic bootstrapping effect.

Conclusion

In sum, syntactic bootstrapping is a prominent proposal for how young children learn an important part of their vocabulary—verbs. Our meta-analysis suggests that while syntactic bootstrapping may be one route to early verb learning, it is a small effect relative to other word learning strategies and may not be robust to a wide range of syntactic structures. Our work highlights the need for pre-registered, high-powered replications of the syntactic bootstrapping effect, and future research examining how syntactic bootstrapping interacts with other word learning strategies.

References

* Arunachalam, S., & Waxman, S. R. (2010). Meaning from syntax: Evidence from 2-year-olds. *Cognition*, *114*(3), 442–446.

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.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J., … Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, (21), 85–123.

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.

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.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376.

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.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge.

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.

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., Jin, K. S., & Scott, R. M. (2020). The developmental origins of syntactic bootstrapping. *Topics in Cognitive Science*, *12*(1), 48–77.

Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., & Tsuji, S. (2018). Symbouki: A meta-analysis on the emergence of sound symbolism in early language acquisition. *Developmental Science*, *21*(5), e12659.

Frank, M., Braginsky, M., Marchman, V., & Yurovsky, D. (in press). Variability and consistency in early language learning: The wordbank project.

Frank, M. C., Lewis, M. L., & MacDonald, K. (2016). A performance model for early word learning. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.

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. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, 47–59.

Gentner, D. (2006). Why verbs are hard to learn. *Action Meets Word: How Children Learn Verbs*, 544–564.

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. *Perspectives on Language and Thought: Interrelations in Development*, *225*.

* Gertner, Y., & Fisher, C. (2012). Predicted errors in children's early sentence comprehension. *Cognition*, *124*(1), 85–94.

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.

Goldwater, M. B. (2017). Grammatical constructions as relational categories. *Topics in Cognitive Science*, *9*(3), 776–799.

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.

Golinkoff, R. M., Jacquet, R. C., Hirsh-Pasek, K., & Nandakumar, R. (1996). Lexical principles may underlie the learning of verbs. *Child Development*, *67*(6), 3101–3119.

Göksun, T., Küntay, A. C., & Naigles, L. R. (2008). Turkish children use morphosyntactic bootstrapping in interpreting verb meaning. *Journal of Child Language*, *35*(2), 291.

Grela, B. G. (2002). Lexical verb diversity in children with down syndrome. *Clinical*

Linguistics &Amp; Phonetics, *16*(4), 251–263.

* 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., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*(1), 54–65.

* Jin, K. S. (2015). *The role of syntactic and discourse information in verb learning* (PhD thesis). University of Illinois at Urbana-Champaign.

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science*, *35*(3), 575–586.

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.

Lee, J. N., & Naigles, L. R. (2008). Mandarin learners use syntactic bootstrapping in verb acquisition. *Cognition*, *106*(2), 1028–1037.

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*.

ManyBabies Consortium. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science*, *3*(1), 24–52.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*(1), 57–77.

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

Mathur, M., & VanderWeele, T. (2020). Sensitivity analysis for publication bias in meta-analyses.

Merriman, W. E., Evey-Burkey, J. A., Marazita, J. M., & Jarvis, L. H. (1996). Young two-year-olds' tendency to map novel verbs onto novel actions. *Journal of Experimental Child Psychology*, *63*(3), 466–498.

Merriman, W. E., Marazita, J., & Jarvis, L. H. (1993). Four-year-olds disambiguation of action and object word reference. *Journal of Experimental Child Psychology*, *56*(3), 412–430.

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.

Monaghan, P., Mattock, K., Davies, R. A., & Smith, A. C. (2015). Gavagai is as gavagai does: Learning nouns and verbs from cross-situational statistics. *Cognitive Science*, *39*(5), 1099–1112.

* 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., 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.

Nappa, R., Wessel, A., McEldoon, K. L., Gleitman, L. R., & Trueswell, J. C. (2009). Use of speaker's gaze and syntax in verb learning. *Language Learning and Development*, *5*(4), 203–234.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606.

Oakes, L. M. (2017). Sample size, statistical power, and false conclusions in infant looking-time research. *Infancy*, *22*(4), 436–469.

O'Hara, M., & Johnston, J. (1997). Syntactic bootstrapping in children with specific language impairment. *International Journal of Language &Amp; Communication Disorders*, *32*(2), 189–205.

Rabagliati, H., Ferguson, B., & Lew-Williams, C. (2019). The profile of abstract rule learning in infancy: Meta-analytic and experimental evidence. *Developmental Science*, *22*(1), e12704.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906–914.

Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., & Golinkoff, R. M. (2009). Live action: Can young children learn verbs from video? *Child Development*, *80*(5), 1360–1375.

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*(5489), 265–266.

Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*(2), 163–180.

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

Tomasello, M. (2010). *Origins of human communication*. MIT press.

Tomasello, M., & Barton, M. E. (1994). Learning words in nonostensive contexts. *Developmental Psychology*, *30*(5), 639.

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.