wooooooow

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

[1] Department of Psychology, Stanford University

[2] Co-author affiliation

Author Note

okie

Correspondence concerning this article should be addressed to Anjie Cao, what is this. E-mail:

Abstract

so abstract!

Keywords: keyword1

Word count:

wooooooow

```
## # A tibble: 55 x 53
##      x1 coder unique_id link  grammatical_cla~ paper_eligibili~ short_cite
##   <dbl> <chr> <chr>     <chr> <chr>            <chr>            <chr>
##  1     5 alan~ arunacha~ http~ verb             include          Arunachal~
##  2     6 anjie arunacha~ http~ verb             include          Arunachal~
##  3     7 anjie arunacha~ http~ verb             include          Arunachal~
##  4     8 alan~ bunger20~ http~ verb             include          Bunger, A~
##  5     9 alan~ bunger20~ http~ verb             include          Bunger, A~
##  6    10 alan~ bunger20~ http~ verb             include          Bunger, A~
##  7    11 anjie bunger20~ http~ verb             include          Bunger, A~
##  8    12 anji~ fisher20~ http~ verb             include          Fisher, C~
##  9    13 anji~ fisher20~ http~ verb             include          Fisher, C~
## 10    14 anji~ fisher20~ http~ verb             include          Fisher, C~
## # ... with 45 more rows, and 46 more variables: expt_num <chr>,
## #   expt_condition <chr>, dependent_measure <chr>, test_type <chr>,
## #   same_infant <chr>, language <chr>, mean_age <dbl>,
## #   productive_vocab_mean <dbl>, productive_vocab_median <dbl>,
## #   population_type <chr>, sentence_structure <chr>, agent_argument_type <chr>,
## #   patient_argument_type <chr>, verb_type <chr>, stimuli_type <chr>,
## #   stimuli_modality <chr>, stimuli_actor <chr>, presentation_type <chr>,
## #   character_identification <chr>, practice_phase <chr>,
## #   test_mass_or_distributed <chr>, n_train_test_pair <dbl>,
## #   n_test_trial_per_pair <dbl>, n_repetitions_sentence <dbl>,
## #   n_repetitions_video <dbl>, example_target_sentence <chr>,
## #   test_question <chr>, inclusion_certainty <dbl>, note <chr>, n_1 <dbl>,
```

```
## #   x_1 <dbl>, x_2 <dbl>, x_2_raw <dbl>, sd_1 <dbl>, sd_2 <dbl>,
## #   sd_2_raw <dbl>, t <dbl>, d <dbl>, d_calc <dbl>, d_var_calc <dbl>,
## #   es_method <chr>, unique_infant <chr>, test_method <chr>,
## #   agent_argument_type_clean <chr>, patient_argument_type_clean <chr>,
## #   adult_participant <chr>
```

```
## [1] 29
```

```
## [1] 106
```

Woo! An intro.

Methods

Literature Search

We conducted our literature search following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [CITE].In Google Scholar, we conducted a keyword search with "Syntactic Bootstrapping"and a forward search on papers that cited the seminal paper, Naigles (1990) (total records identified: N = 3339; retrieved between May 2020 and July 2020). Additional papers were identified by consulting the references section of a recent literature review (N = 155) [CITE] and the experts in the field (N = 11). We included works that are published journal articles, conference proceedings, doctoral dissertations, and unpublished manuscripts. They will be collectively referred to as "papers" in the following sections. All papers include reports of experimental studies. Each paper may include multiple experimental conditions, and thus provide multiple effect sizes for the final analysis.

To be included in our final sample, the paper must satisfy the following criteria: First, the experimental paradigms must involve a two-alternative forced-choice situation, in which the participants were instructed to identify the scenes that match the linguistic stimuli. Second, the visual scenes are all displayed on a monitor, and the media can be either video recording or animation clips. Third, the linguistic stimuli used in the experiments must include at least one novel verb embedded in a syntactically informative frame For example, “Look, Kradding!” does not provide relevant syntactic information to interpret the novel verb “kradding”. In contrast, “Look, it’s kradding!” embeds the novel verb in an intransitive syntactic frame. If the linguistic stimuli used in the experiments only include the former example, then the experiments would be excluded. But as long as the participants are exposed to one exemplar of the latter example, the experiments would be included. Finally, we decided to focus on English-speaking, typically-developing children. Experimental conditions that tested on other populations were excluded from the sample.

Our final sample included data from 1645 unique infants (Mean age: 936.91 Days), reported in 106 individual effect sizes from 29 individual papers.

Data Entry

We classified our final samples of effect sizes (N = 106) into two categories: the ones using “classical approach”(N = 47) and the ones using “non-classical approach”(N = 55). The experimental conditions were categorized as using “classical approach” if the the set-up and procedure were most similar to the seminal paper (Naigles, 1990). The ones categorized as using “non-classical approach” were experimental conditions to meet all of the inclusion criteria, but they deviated from the classical approach in some non-trivial aspects (The stimuli involved more complex syntactic structure or event structure: N = 94; the testing procedure includes a contrast phase: N = 10; the experiment was conducted through online platform: N = 2). To preserve the homogeneity of the experiments without reducing the sample sizes, we conducted our statistical analysis both on the entire sample and specifically on the experimental conditions labeled as “classical approach”.

Participants-related moderators included infants' mean age by days and the median productive vocabulary measured by MacArthur-Bates Communicative Development Inventories (CDI) Words and Sentences. Most papers reported mean age by months, so we converted it to mean age by days by multiplying the reported statistics with 30.44, the average number of days in a month. All conditions (N= 106) have the age information available. 40 conditions have the vocabulary size information available.

Stimuli related variables included features of both the linguistics stimuli and the visual stimuli. For linguistic stimuli during the training phase, we focused on the structure of the sentences infants heard. The sentences were categorized as transitive(N = 41), if and only if the novel verbs were embedded in a sentence with two or more noun arguments. The sentences were categorized as intransitive (N= 41) if the novel verbs were embedded in a sentence with one noun argument. We also coded the types of words used in the agent argument (One noun: N = 31; One pronoun: N = 10; Two nouns: N = 1; Noun phrase: N = 10; Varying across sentences: N = 26) and the patient argument(One noun: 34; One pronoun: 13; Noun pharse: N = 15; Varying across sentences: N = 4). For linguistic stimuli during the testing phase, we coded whether the infants were prompted to identify the action ("an example", N=) or the actors ("an example, N=). For visual stimuli, we coded both the types of media used (video or animation) and the types of protagonists in the events (person or non-person). [N=??] conditions used video recordings of human actors (N=??) or human actors in animal suits (N=??). The other [N=??] conditions used clips of animation with ??? as the protagonists of the events. We also coded how the onset of the linguistic stimuli aligned with the visual stimuli. The procedure was coded as"simultaneous" if the very first training sentence was presented along with the visual stimuli depicting relevant action (N=??). It was coded as "immediately after" if the first training sentence was presented along with an attention-getter or a blank screen, immediately followed by the relevant action (N=??). Finally, some experimental conditions first presented the linguistic stimuli paired with irrelevant visual scenes (for example, a person on the phone talking). The relevant visual stimuli were not shown until the training phase is over. For experimental conditions using this procedure, they were coded as “asynchronous”(N=??).

Coded Variables

For each experimental condition, we coded two types of variables: participants-related variables, the stimuli related variables, and the experimental procedure related variables The variables were either retrieved from the papers’ methods sections or obtained via contacting the authors. Allinformation was either retrieved from the paper’s methods sections and the figures illustrating the procedure, or obtained by contacting the authors.

Participants-related variables included infants’ mean age by days and the mean productive mean productive vocabulary measured by MacArthur-Bates Communicative Development Inventories (CDI) Words and Sentences. Most papers reported mean age by months, so we converted it to mean age by days by multiplying the reported statistics with 30.44, the average number of days in a month. All conditions (N=??) have the age variable availble. [???] conditions have the vocabulary information available.

Stimuli related variables included features of both the linguistics stimuli and the visual stimuli. For linguistic stimuli during the training phase, we focused on the structure of the sentences infants heard. The sentences were categorized as transitive(N =???), if and only if when the novel verb in the sentence was preceded by a subject and followed by an object. Other sentences were categorized as intransitive (N=???). We also coded the types of words used in the agent argument and the patient argument. [N=??] experimental conditions [EXPLAIN?? the arument]. For linguistic sstimuli during the testing phase, we coded whether the infants were prompted to identify the action (“an example”, N=) or the actors (“an example, N=). For visual stimuli, we coded both the types of media used (video or animation) and the types of protagonists in the events (person or non-person).

[N=??] conditions used video recordings of human actors (N=??) or human actors in animal suits (N=??). The other [N=??] conditions used clips of animation with ??? as the protagonists of the events. We also coded how the onset of the linguistic stimuli aligned with the visual stimuli. The procedure was coded as"simultaenous" if the very first training sentence was presented along with the visual stimuli depicting relevant action (N=??). It was coded as "immediate after" if the first training sentence was presented along with an attention-getter or a blank screen, immediately followed by the relevant action (N=??). Finally, some experimental conditions first presented the linguistic stimuli paired with irrelevant visual scenes (for example, a person on the phone talking). The relevant visual stimuli were not shown until the training phase is over. For experimental conditions using this procedure, they were coded as "asynchronous"(N=??).

Finally, experimental procedure related variables included the type of response elicited from the participants: whether the infants were explicitly prompted to point or their eye gaze duration were measured as they heard the linguistic stimuli. (Pointing: N = [??]; Looking: N=[???]). Three characteristics of the experimental procedures were coded as categorical variables: the inclusion of practice phase (Yes: N= ; No: N=), the inclusion of character identification phase (Yes: N=: No:), and the distribution of the training and the testing trials. A procedure was categorized as "mass" if and only if the infants were trained exclusively on one novel verb and tested on the very same verb (N=??). It was "distributed" if the infants were given multiple train and test pairs on multiple novel verbs (N=??) To better characterize the experience infants had prior to testing, we also coded how many train-test pair the infants were given (for mass procedure it was always 1), how many trials during the test phase infants were given, how many times the visual stimuli showing the relevant actions were presented, and how many timeseach novel verb was spoken in a syntactically-informative way. If the prompt question provided relevant syntactic context (EXAMPLE), then it was also counted.

```
##  [1] "x1"                          "coder"
##  [3] "unique_id"                   "link"
##  [5] "grammatical_class"           "paper_eligibility"
##  [7] "short_cite"                  "expt_num"
##  [9] "expt_condition"              "dependent_measure"
## [11] "test_type"                   "same_infant"
## [13] "language"                    "mean_age"
## [15] "productive_vocab_mean"       "productive_vocab_median"
## [17] "population_type"             "sentence_structure"
## [19] "agent_argument_type"         "patient_argument_type"
## [21] "verb_type"                   "stimuli_type"
## [23] "stimuli_modality"            "stimuli_actor"
## [25] "presentation_type"           "character_identification"
## [27] "practice_phase"              "test_mass_or_distributed"
## [29] "n_train_test_pair"           "n_test_trial_per_pair"
## [31] "n_repetitions_sentence"      "n_repetitions_video"
## [33] "example_target_sentence"     "test_question"
## [35] "inclusion_certainty"         "note"
## [37] "n_1"                         "x_1"
## [39] "x_2"                         "x_2_raw"
## [41] "sd_1"                        "sd_2"
## [43] "sd_2_raw"                    "t"
## [45] "d"                           "d_calc"
## [47] "d_var_calc"                  "es_method"
## [49] "unique_infant"               "test_method"
## [51] "agent_argument_type_clean"   "patient_argument_type_clean"
## [53] "adult_participant"
```

Data Preprocessing

We calculated the Cohen's d effect size for each experimental conditions. Because we only included studies that used a two alterantive forced-choice test method, we compared the mean proportion of correct response against chance level (0.5) for the group[DO I need to explain the deviation?]. A small portion of the looking time studies (N=??) reported their own baseline looking time measurements, i.e., the participants' looking time before and after hearing the relevant linguistic prompts. For these studies, we calculated two versions of the effect sizes. One with the chance level and the other with the reported baseline. [maybe say something about they don't correlate very well? and said we ran two versions of the statistics analysis] The mean and standard deviation were obtained from one of the three ways: a)directly retreived from the results section or the tables presenting data(N=??); b) recovered from barplots by measuring the height of the bars and the error bars (N=??); c) contacting with the original authors (N=??). For looking time studies, when the paper only reported the raw looking time, we calculated the proportion of correct response by dividing the mean looking time toward the matching scenes by the sum of mean looking time toward the matching scenes and the mean looking time toward the non-matching scenes (i.e., excluding the look away itme). The standard deviations were also scaled by being divided by the sum.

All the effect sizes and the coded variables were then analyzed with the metafor package in R [CITE] # Results

A Section Name

General Discussion

References

Appendix