

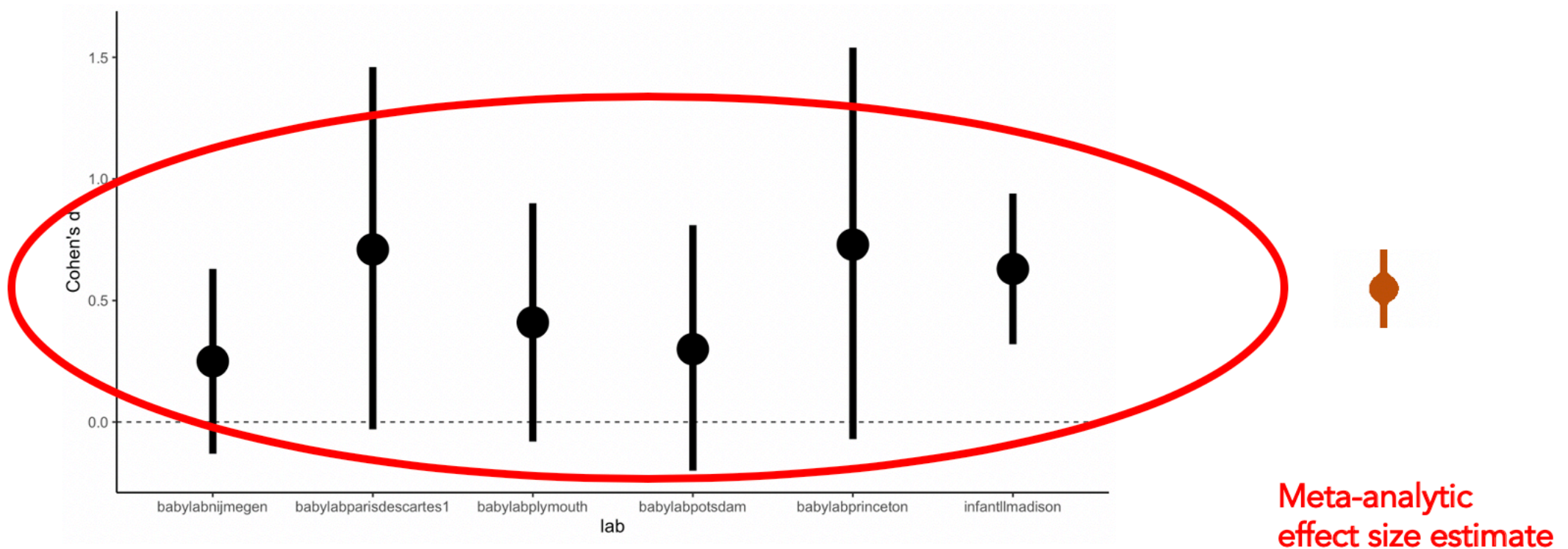
# Conducting your own MA

25 March 2020

*Modern Research Methods*

# Last time: Intro to meta-analysis

A quantitative approach to summarizing results across studies



# Today

- Wrap-up coding of Deak paper
- More on mutual exclusivity meta-analysis
- Introduction to final project/choosing a topic

# More practice coding effect sizes from papers

*J. Child Lang.* **28** (2001), 787–804. © 2001 Cambridge University Press  
DOI: 10.1017/S0305000901004858 Printed in the United Kingdom

## NOTE

**By any other name: when will preschoolers produce  
several labels for a referent?\***

GEDEON O. DEÁK AND LOULEE YEN  
*Vanderbilt University*

JEREMY PETTIT  
*David Lipscomb University*

(Received 23 February 2000. Revised 8 December 2000)

<https://cumulativescience.netlify.com/readings/deak2001.pdf>

# Your job #2

- In your breakout groups, calculate an effect size for the two age groups in Deak, et al. (2001) in Experiment 1.
- If you have time, you can also calculate the confidence intervals on the effect sizes.
- Disambiguation effect

# Mutual exclusivity meta-analysis

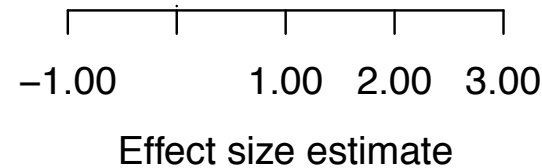
First author	Year	Age (m.)	N
1. bion	2013	18	22
2. bion	2013	24	25
3. bion	2013	30	20
4. byers	2009	17	16
5. grassman	2010	24	12
6. grassman	2010	48	12
7. markman	1988	45	10
8. spiegel	2011	30	72

"Forest plot"

Grand effect size

← Grand effect size estimate

*Pool effect sizes across studies, weighting by sample size*



# Actual Meta-Analysis



Contents lists available at [ScienceDirect](#)

Cognition

journal homepage: [www.elsevier.com/locate/cognit](http://www.elsevier.com/locate/cognit)



## Original Articles

### The role of developmental change and linguistic experience in the mutual exclusivity effect<sup>☆</sup>



Molly Lewis<sup>a,\*</sup>, Veronica Cristiano<sup>b</sup>, Brenden M. Lake<sup>c,d</sup>, Tammy Kwan<sup>c,d</sup>, Michael C. Frank<sup>e</sup>

<sup>a</sup> *Carnegie Mellon University, United States of America*

<sup>b</sup> *Gallaudet University, United States of America*

<sup>c</sup> *New York University, United States of America*

<sup>d</sup> *Cognitive ToyBox, Inc., United States of America*

<sup>e</sup> *Stanford University, United States of America*

# Meta-Analysis details

- Take 5 minutes to read the Methods section of this MA (pg. 3 – 4)
- Link on website
- Starting with “2. Meta-analysis”
- Then, we’ll discuss the results together

M. Lewis, et al.

children with larger vocabularies tend to have a larger ME bias (Blon et al., 2013; Deak, Yen, & Pettit, 2001; Graham, Poulin-Dubois, & Baker, 1998; Houston-Price et al., 2010; Law & Edwards, 2015; Lederberg & Spencer, 2008; Mervis & Bertrand, 1995).

Given the range of possible mechanisms producing experience-driven developmental change, a description of the developmental trajectory of the effect is needed in order to sufficiently constrain theories. There are a small set of studies that show developmental change in the mutual exclusivity effect by testing more than a couple age groups within the same experiment (Blon et al., 2013; Frank, Sugarman, Horowitz, Lewis, & Yurovsky, 2016; Grassmann et al., 2015; Halberda, 2003; Merriman & Bowman, 1989). For example, Halberda (2003) tested 14-16- and 17-mo in the ME paradigm, and found a pattern of developmental change: 14-mo children were biased to select the familiar object, 16-mo were at chance, and 17-mo were biased to select the novel object, demonstrating the ME effect.

However, while multi-age-group studies on ME provide clear evidence that there is a greater propensity to make the ME inference with development, they do not provide a continuous, quantitative description of the developmental trajectory of the effect that could help distinguish between theories of ME making qualitatively similar predictions. Instead, multi-age-group studies focus theorizing on accounting for why children at one or a few timepoints in development behave in a way that is consistent or not with the ME effect. In part, this focus on the “emergence” of the ME effect may be due to methodological challenges in conducting developmental experiments rather than to an underlying theoretical motivation: Since data collection from young children is expensive, it is costly for researchers to collect data from children across more than a couple age groups. In addition, experimental evidence from the ME paradigm is typically summarized as a binary description (children’s “success” or “failure” in the ME task) rather than as a more continuous estimate of the effect size, and this methodological choice may obscure evidence of more subtle changes in the cognitive system across development. In order to make stronger inferences about the cognitive mechanisms underlying the ME effect, a more fine-grained description of the developmental trajectory of the effect is therefore needed.

## 1.1. The current study

We first describe the state of the evidence for developmental change in the ME effect via a meta-analysis of the extant empirical literature. By aggregating across studies that each test different ages, the meta-analytic approach allows us to take advantage of the large number of studies already conducted on the ME effect in order to characterize developmental change. We then present two new, relatively large-sample developmental experiments that investigate the causal role of linguistic experience in contributing to the ME effect. In Experiment 1, we examine the relationship between one correlate of language experience — vocabulary size — and the strength of the ME effect on a large sample of children. We find evidence that children with larger vocabularies tend to show a stronger ME effect, consistent with the notion that language experience influences the ME effect. In Experiment 2, we test the hypothesis that language experience plays a causal role in the ME effect, by directly manipulating children’s amount of experience with a word. We find greater experience with the familiar word-object mapping in the ME paradigm leads to a stronger ME effect. In the General Discussion, we conclude by discussing the role of developmental change and experience in the context of candidate theories of ME, in the context of our evidence.

## 2. Meta-analysis

To assess the strength of the ME effect as well moderating factors, we conducted a meta-analysis on the existing body of literature investigating the ME effect.

Cognition 198 (2020) 104191

## 1. Methods

### 1.1. Search strategy

We conducted a forward search based on citations of Markman and Wachtel (1988) in Google Scholar, and by using the keyword combination “mutual exclusivity” in Google Scholar (retrieved September 2013; November 2017).<sup>2</sup> Additional papers were identified through citations and by consulting experts in the field. We then narrowed our sample to the subset of studies that used one of two different paradigms: (a) an experimenter says a novel word in the context of a familiar object and a novel object and the child guesses the intended referent (the canonical paradigm; “Familiar-Novel”), or (b) experimenter first provides the child with an unambiguous mapping of a novel label to a novel object, and then introduces a second novel object and asks the child to identify the referent of a second novel label (“Novel-Novel”), or Familiar-Novel conditions, we included conditions that used more than one familiar object (e.g. Familiar-Familiar-Novel). From these conditions, we restricted our sample to only those that satisfied the following criteria: (a) participants were children (less than 12 years of age),<sup>3</sup> (b) referents were objects or pictures (not facts or object parts), (c) no incongruent cues (e.g. eye gaze at familiar object) and (d) children had visual access to the objects (versus exclusively touch). All papers used either forced-choice pointing or eye-tracking methodology. All papers were peer-reviewed with the exception of two dissertations (Williams, 2009; Frank, 1999). In total, we identified 48 papers that satisfied our selection criteria and had sufficient information to calculate an effect size. Papers included in the meta-analysis are marked with an asterisk in the bibliography.

### 1.2. Coding

For each paper, we coded separately each relevant condition with each age group entered as a separate condition. For each condition, we coded the paper metadata (citation) as well as several potential moderator variables: mean age of infants, estimates of mean vocabulary size if the sample population from the Words and Gestures form of the MacArthur-Bates Communicative Development Inventory when available (MCDI; Fenson et al., 1994, 2007), and participant population type.<sup>4</sup> We used production vocabulary as our estimate of vocabulary size since it was available for more studies in our sample. We coded participant population as one of three subpopulations that have been studied in the literature: (a) typically developing monolingual children, (b) multilingual children (including both bilingual and trilingual children), and (c) non-typically developing children. Non-typically developing conditions included children with selective language impairment, language delays, hearing impairment, autism spectrum disorder, and Down Syndrome.

In order to estimate effect size for each condition, we coded sample size, proportion novel-object selection, baseline (e.g., 0.5 in a 2-AFC paradigm), standard deviations for novel object selections, *t*-statistic, and Cohen’s *d*, where available. For several conditions, there was insufficient data reported in the main text to calculate an effect size (no means and standard deviations, *t*-statistics, or Cohen’s *d*), but we were able to estimate the means and standard deviations through measurement of plots ( $N = 13$ ), imputation from other data within the paper ( $N = 11$ ), or through contacting authors ( $N = 34$ ). Our final sample included 146 effect sizes ( $N_{typical-developing} = 119$ ;  $N_{multilingual} = 12$ ;

<sup>2</sup> Data and analysis code for this and subsequent studies are available in an online repository at: [https://github.com/mllewis/me\\_vocabs](https://github.com/mllewis/me_vocabs).

<sup>3</sup> This cutoff was arbitrary but allowed us to include conditions from older children from non-typically-developing populations.

<sup>4</sup> We also coded a number of other moderating variables not included here: sex (eye-tracking or pointing), number of alternatives in the forced choice method, and task modality (paper vs. object). See <https://meta-lab.stanford.edu/> for more analyses.

M. Lewis, et al.

$N_{non-typically-developing} = 15$ ).

### 2.1.3. Statistical approach

We calculated effect sizes (Cohen’s *d*) from reported means and standard deviations where available, otherwise we relied on reported *t*-statistics (*t* or *d*). Effect sizes were computed by a script, `compute_es.R`, available in the GitHub repository. All analyses were conducted with the `metafor` package in R (Viechtbauer, 2010) using multi-level random effect model with grouping by paper and participant group (for conditions with the same or overlapping participant samples).<sup>5</sup> In models with moderators, moderator variables were included as additive fixed effects. Age was entered as logarithmic in months (where one month equals 30.44 days) to facilitate interpretation. All estimate ranges are 95% confidence intervals.

### 2.2. Results

In a model with all conditions in our sample, we estimated the overall ME effect size to be 1.27 (0.99, 1.55), and reliably greater than zero ( $p < .001$ ; Fig. 1).<sup>6</sup>

We next conducted a separate meta-analysis for four theoretically-relevant conditions: Familiar-Novel trials with typically developing participants, Novel-Novel trials with typically developing participants, conditions with multilingual participants, and conditions with non-typically developing participants.

#### 2.2.1. Typically developing population: Familiar-Novel trials

We first examined effect sizes of ME for typically-developing children in the canonical familiar-novel paradigm. This is the central data point that theories of ME must explain.

The overall effect size for these conditions was 1.37 (1, 1.75), and reliably greater than zero ( $p < .001$ ; Fig. 1). The effect sizes contained considerable heterogeneity, however ( $Q = 823.09$ ,  $p < .001$ ).

We next tried to predict this heterogeneity with two moderators corresponding to developmental change: age and vocabulary size. In a model with age as a moderator, age was a reliable predictor of effect size ( $\beta = 2.08$ ,  $Z = 6.15$ ,  $p < .001$ ; see Table 1), suggesting that the ME effect becomes larger as children get older (Fig. 2). For the conditions for which we had estimates of vocabulary size ( $N = 18$ ), age of participants was highly correlated with vocabulary size in our sample ( $r = 0.50$ ,  $p < .01$ ). In a model with only vocabulary as a moderator, vocabulary was also a reliable predictor of effect size ( $\beta = 0.003$ ,  $Z = 2.66$ ,  $p = 0.01$ ). Next we asked whether vocabulary size predicted independent variance in the magnitude of the ME effect. To test this, we fit a model with both age and vocabulary size as moderators. In this model, neither vocabulary size ( $\beta = 0.002$ ,  $Z = 1.23$ ,  $p = 0.22$ ) nor age ( $\beta = 1.06$ ,  $Z = 1.01$ ,  $p = 0.31$ ) was a reliable predictor of ME effect size, likely due in part to the high intercorrelation between the two predictors.

These analyses confirm that the ME effect is robust, and associated with a very large effect size ( $d = 1.37$  [1, 1.75]) relative to other experimental psychology phenomena (Bosco, Aguinis, Singh, Field, & Pierce, 2015; Open Science Collaboration, 2015). They also suggest that

<sup>5</sup> The exact model specification was as follows: `metafor::rma.mv(yi = effect.size, V = effect.size.var, random = ~ 1 | paper|participant, group)`.

<sup>6</sup> These conditions were more than three standard deviations beyond the overall mean effect size (two typically developing Familiar-Novel conditions and one non-typically-developing condition). These outliers contributed to heterogeneity in our sample (Breusch-Pagan test  $\chi^2(1) = 11.86$ ,  $p < .001$ ). With these outliers excluded (Lipsey & Wilson, 2001), the heteroskedasticity was reduced ( $\chi^2(1) = 0.13$ ,  $p = .72$ ) and the overall effect size (1.22 [0.96, 1.48]) was slightly smaller, but qualitatively the same. Given that we have no theoretical reason to exclude these conditions, we have included all conditions in our analyses presented here.

Cognition 198 (2020) 104191

the magnitude of the effect strengthens over development. Vocabulary size, though correlated with age, does not predict additional effect size variance over and above age. This finding is difficult to interpret however, given the fact that estimates of vocabulary size are likely far more accurate than those of age, and we likely have less power to detect an effect of vocabulary size relative to age, since estimates of vocabulary size are available for only a minority of conditions (18%).

#### 2.2.2. Typically developing population: Novel-Novel trials

One way that vocabulary knowledge could lead to increased performance on the Familiar-Novel ME task is through increased certainty about the label associated with the familiar word: If a child is more certain that a ball is called “ball,” then the child should be more certain the novel label applies to the novel object. Novel-Novel trials control for potential variability in certainty about the familiar object by having participants a new label for a novel object prior to the critical ME trial, where this previously-learned label becomes the “familiar” object in the ME task. If knowledge of the familiar object is not the only contributor to age-related changes in the ME effect, then there should be an increase in the magnitude of the ME effect in Novel-Novel trials, as well as Familiar-Novel trials. In addition, if the strength of knowledge of the “familiar” object influences the strength of the ME effect, then the overall effect size should be smaller for Novel-Novel trials, compared to Familiar-Novel trials.

For conditions with the Novel-Novel trial design, the overall effect size was 1.29 (0.69, 1.89) and reliably greater than zero ( $p < .001$ ). We next asked whether age predicted some of the variance in these trials by fitting a model with age as a moderator. Age was a reliable predictor of effect size ( $\beta = 0.93$ ,  $Z = 3$ ,  $p < .001$ ), suggesting that the strength of the ME effect increases with age. There were no Novel-Novel conditions in our dataset where the mean vocabulary size of the sample was reported, and thus we were not able to examine the moderating role of vocabulary size on this trial type.

Finally, we fit a model with both age and trial type (Familiar-Novel or Novel-Novel) as moderators of the ME effect. Both moderators predicted independent variance in ME effect size (age:  $\beta = 1.89$ ,  $Z = 6.94$ ,  $p < .0001$ ; trial type:  $\beta = -0.88$ ,  $Z = -5.06$ ,  $p < .0001$ ), with Familiar-Novel conditions and conditions with older participants tending to have larger effect sizes.

These analyses suggest that both development (either via maturation or experience-related changes) as well as the strength of the familiar word representation are related to the strength of the ME effect. A successful theory of ME will need to account for both of these empirical facts.

#### 2.2.3. Multilingual population

We next turn to a different population of participants: Children who are simultaneously learning multiple languages. This population is of theoretical interest because it allows us to isolate the influence of linguistic knowledge from the influence of domain-general capabilities. If the ME effect relies on mechanisms that are domain-general and independent of linguistic knowledge, then we should expect the magnitude of the effect to be the same for multilingual children compared to monolingual children.

Children learning multiple languages did not reliably show the ME effect in a model not controlling for age ( $d = 0.57$  [–0.13, 1.28]). We next fit a model with both monolingual (typically-developing) and multilingual participants, predicting effect size with language status (monolingual vs. multilingual), while controlling for age. Both language status ( $\beta = 0.61$ ,  $Z = 1.91$ ,  $p = 0.06$ ) and age ( $\beta = 1.61$ ,  $Z = 6.57$ ,  $p < .0001$ ) were reliable predictors of effect size: Being monolingual and older were each predictive of a larger effect size.

These data provide some evidence that language-specific knowledge influences the magnitude of the ME bias, consistent with the experimental work with multilinguals.



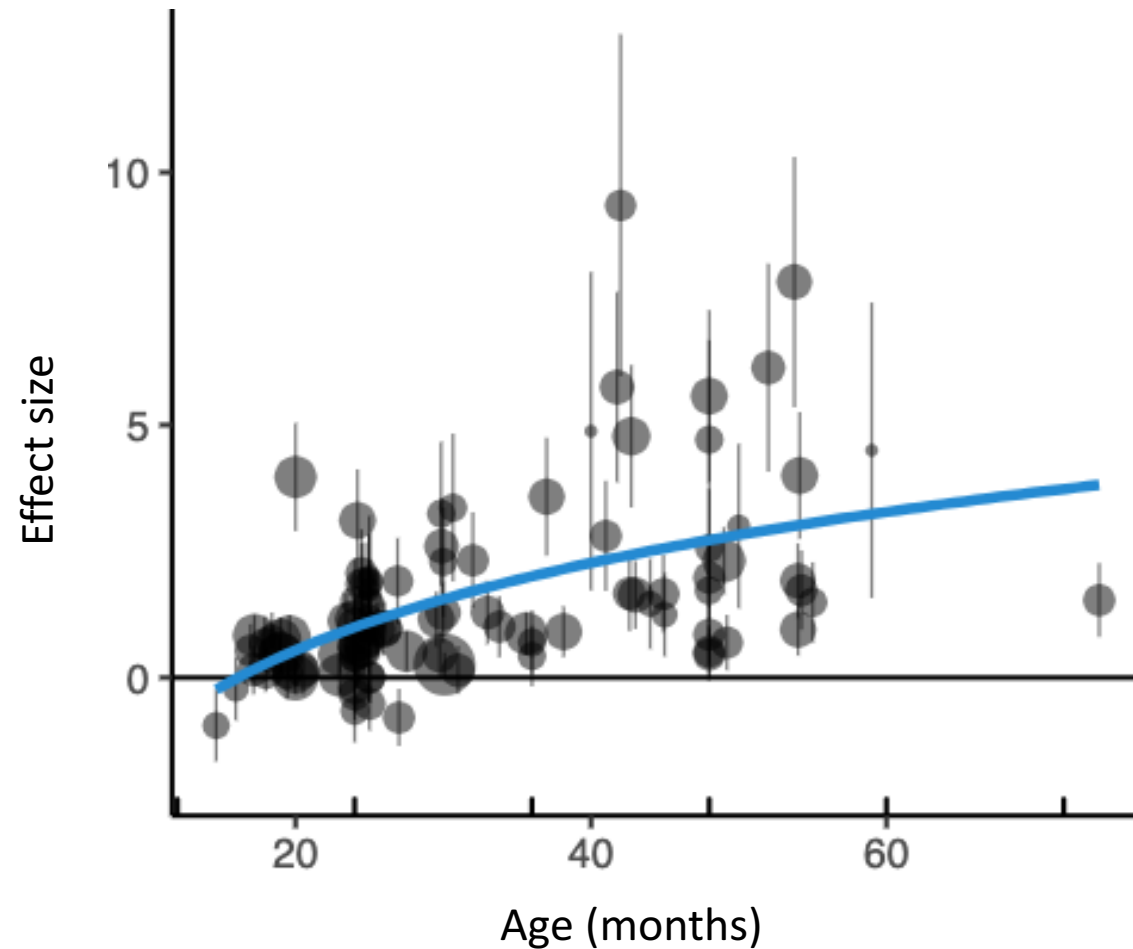
# Meta-analysis details

- Search strategy?
- Coding?
  - How many effect sizes total?
  - What else did we code for? (i.e. “moderators”)
- Statistical approach?

# Moderators

- = anything you think might influence the effect size
- Age
- Design
- Stimuli type
- # of languages spoken

# Exploring a moderator of effect size: Age



Each point is a study

# Design

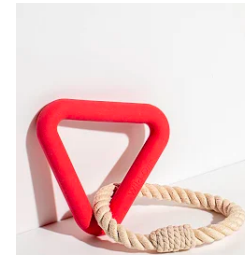
## Familiar-Novel Design

Can you find the “fep”?



## Novel - Novel Design

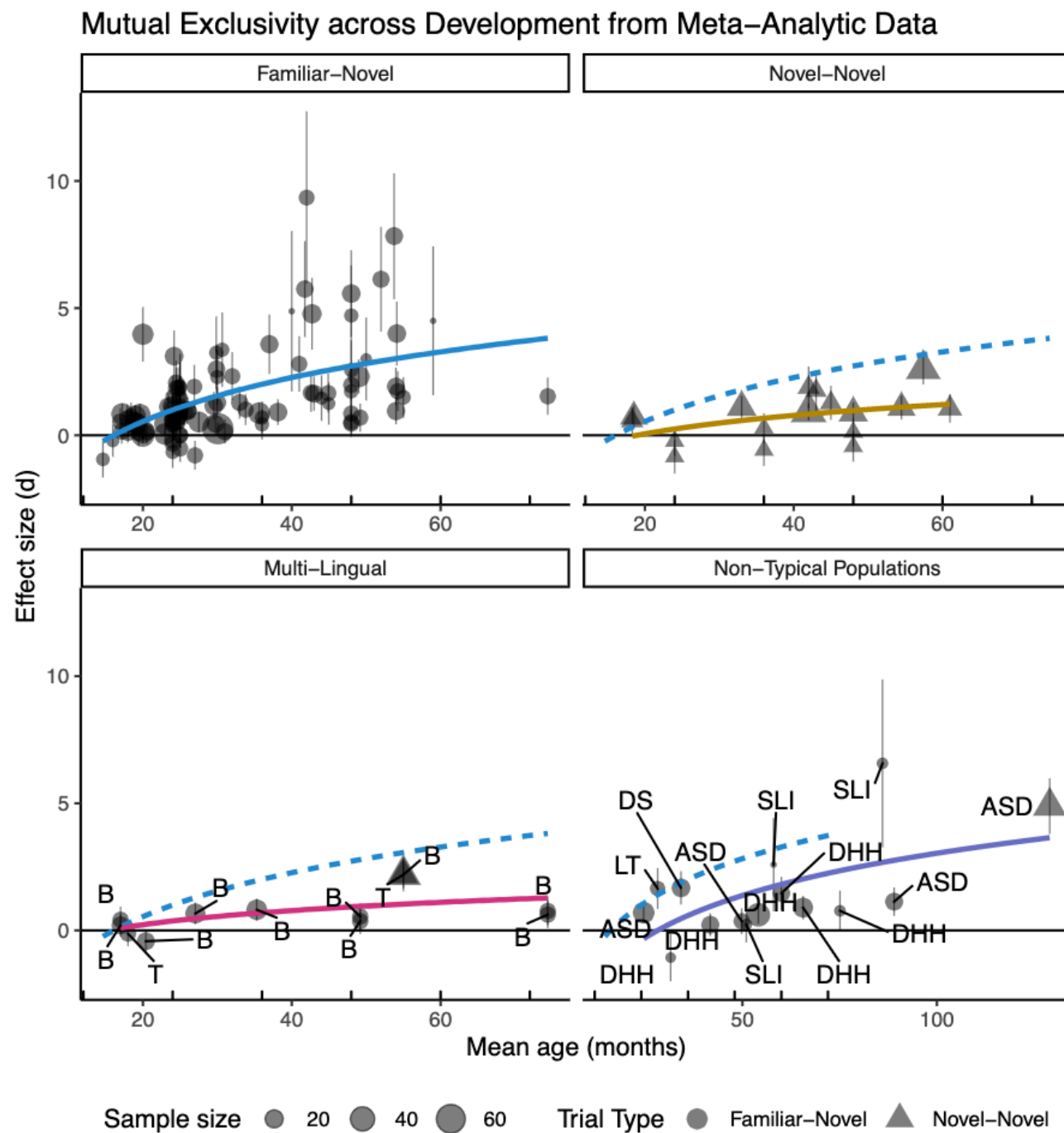
This is a “Dax”

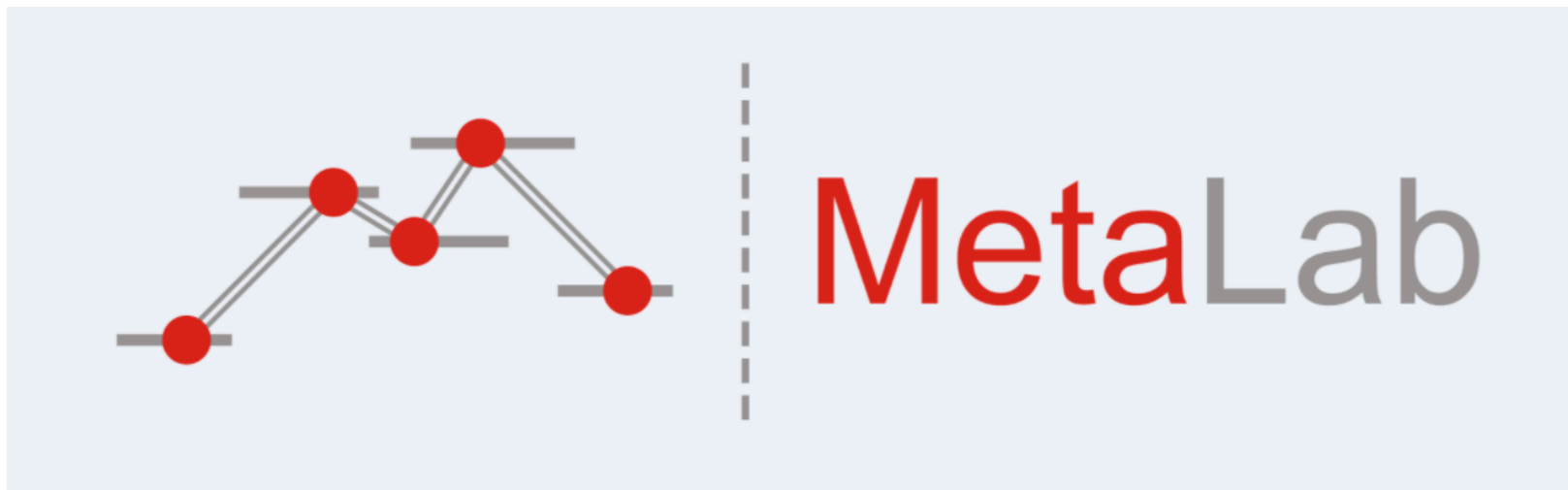


Can you find the “fep”?



# Exploring other moderators










- Aggregate of meta-analyses of different phenomena in cognitive development (focus on language acquisition)
- Interactive visualizations
- <http://metalab.stanford.edu/>

# Next Time: Forming groups for final project

Before lab on Friday, look over papers that are possible topics for your final project.

In google drive folder.

AGGREGATING MANY EXPERIMENTS		READING	SLIDES	ASSIGNMENT	NOTES
Week 10 [M, 3/23]	Intro to meta-analysis (MA)				
[W]	Choosing a MA topic				[FINAL PROJECT] [Lewis et al. 2020] [Seminal papers for final project]
[F]	Lab: Choosing a MA topic				Skills: papaja