Dynamically aggregating evidence in community-augmented meta-analyses

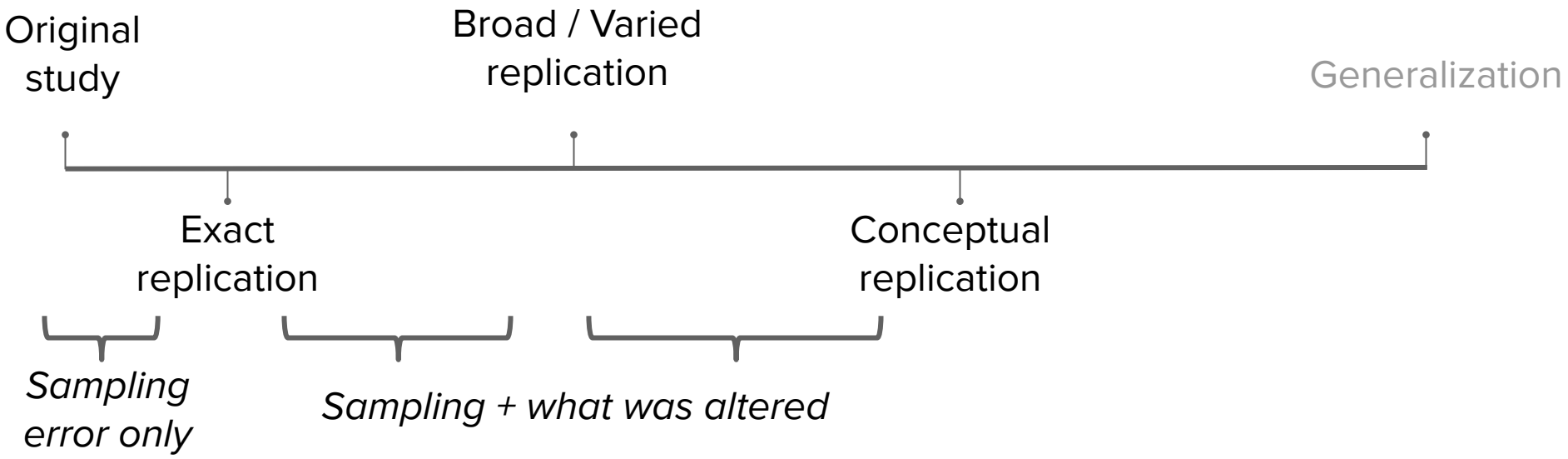
Christina Bergmann
christina.bergmann@mpi.nl
@chbergma

MAX PLANCK INSTITUTE
FOR PSYCHOLINGUISTICS

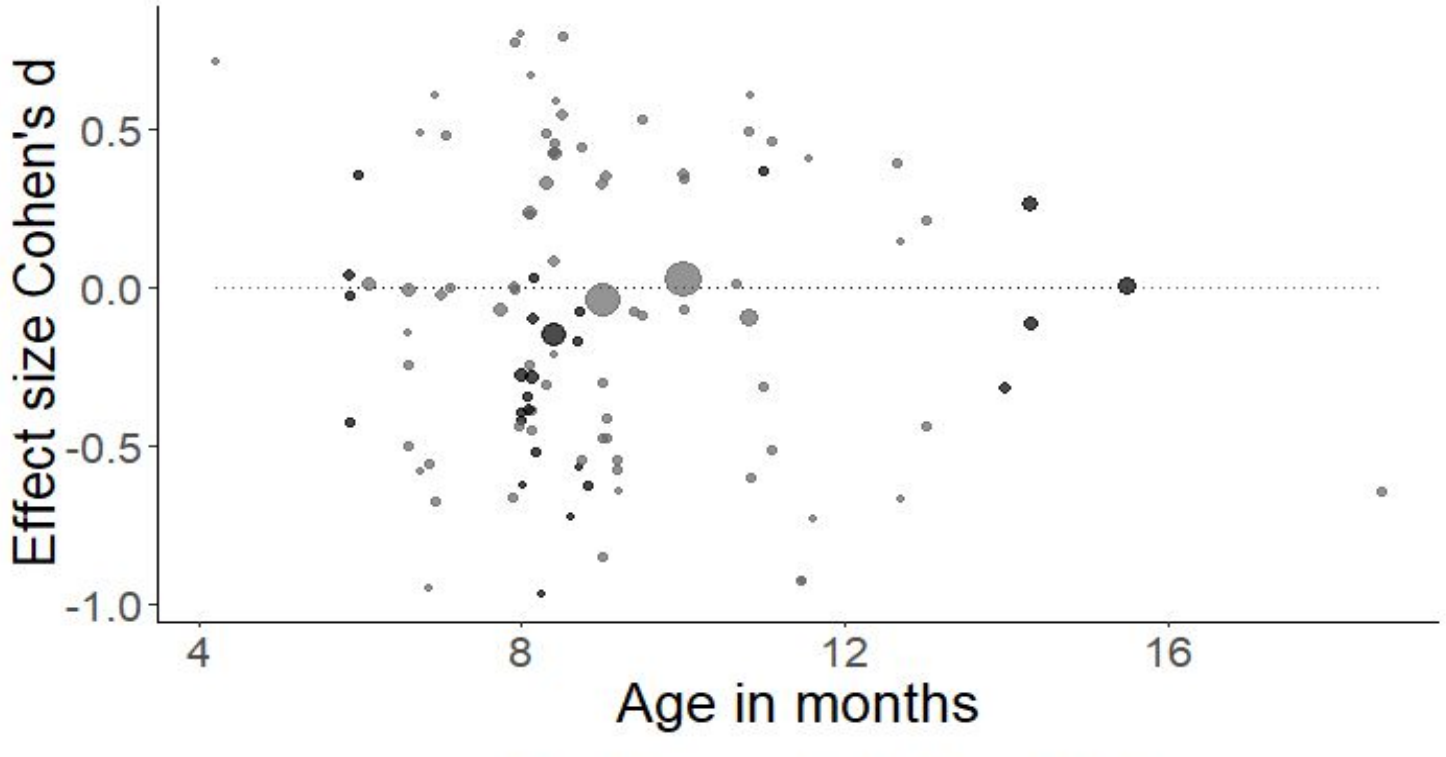
WWW.MPI.NL

A notion of replication in meta-analyses

Pairs of studies in a meta-analysis live on a replication continuum

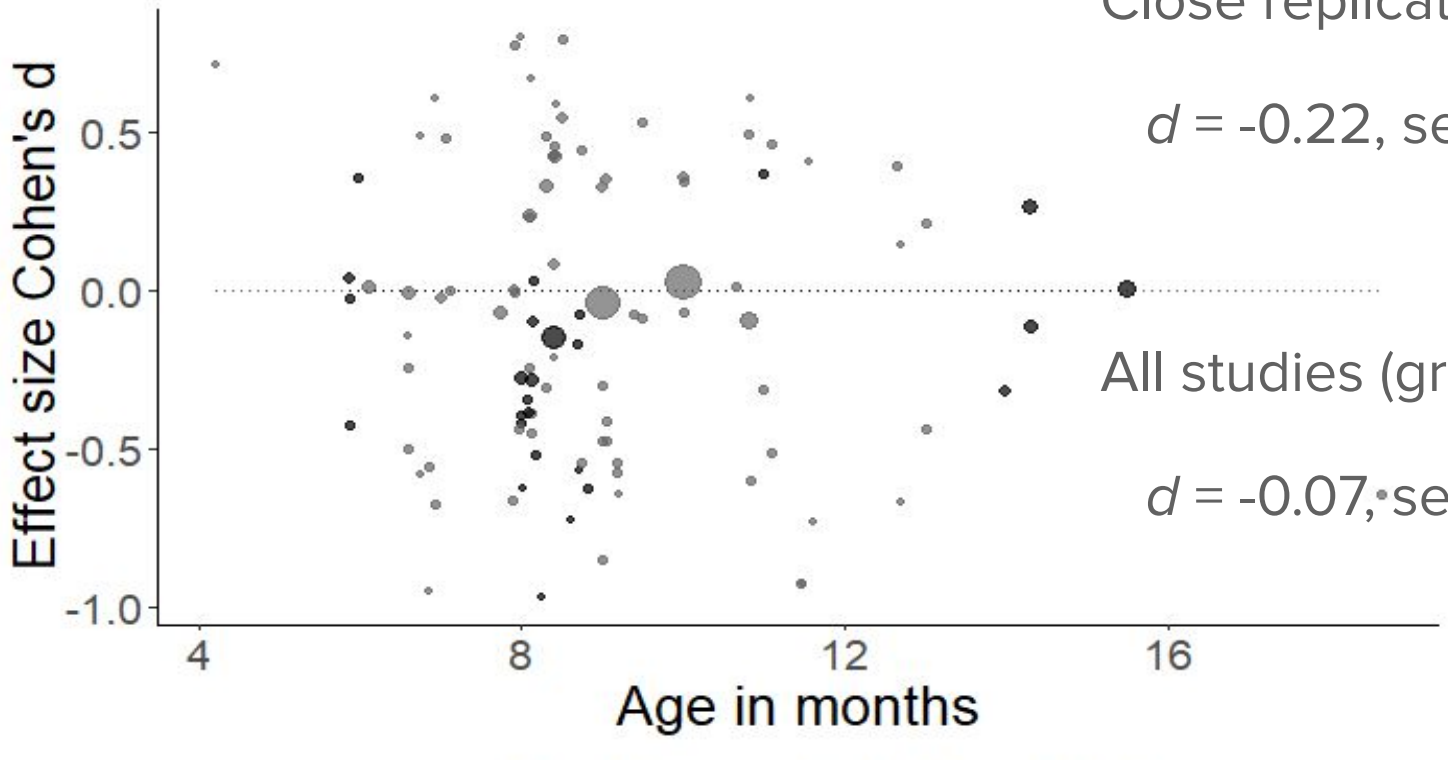


Just in case: Replication distance matters



Black & Bergmann (in prep/2017)

Just in case: Replication distance matters



Close replications (black)

d = -0.22, se = 0.06

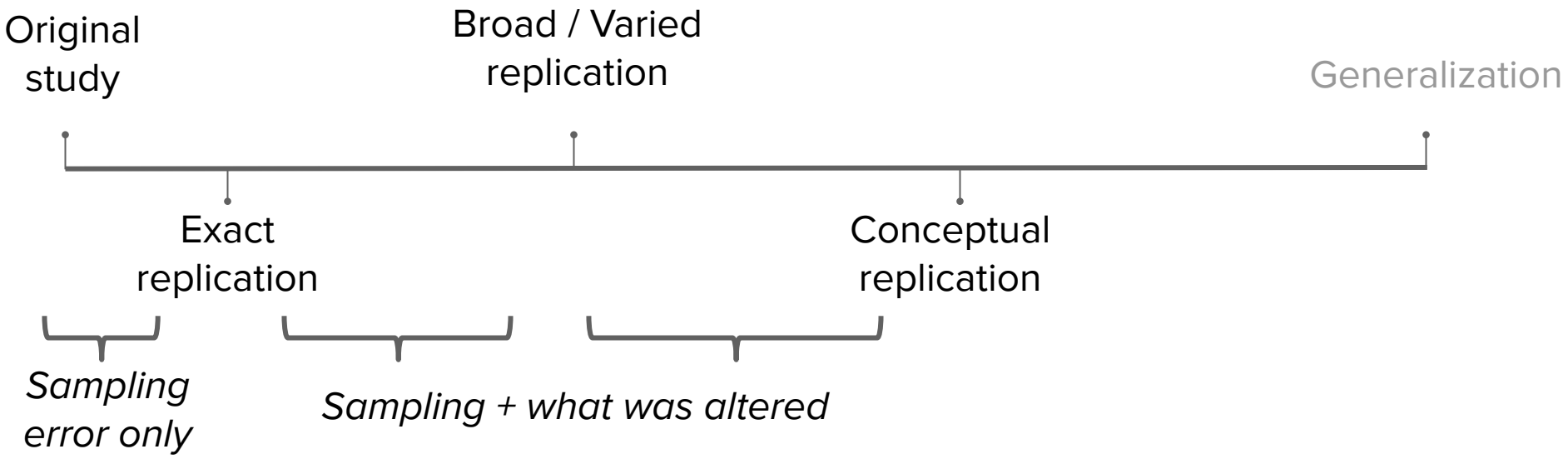
All studies (grey)

d = -0.07, se = 0.05

Black & Bergmann (in prep/2017)

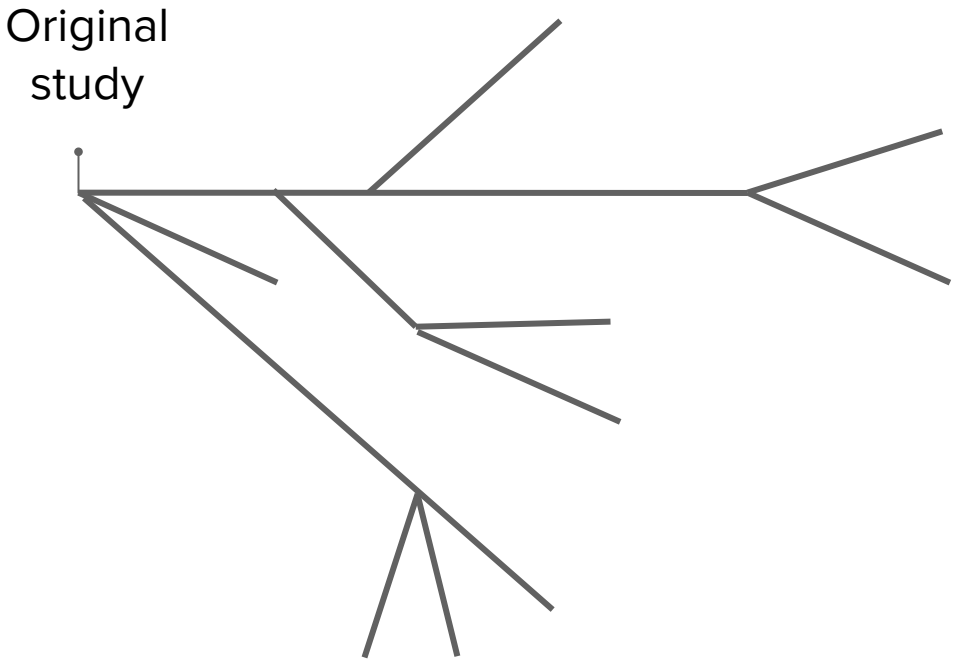
A notion of replication in meta-analyses

Pairs of studies in a meta-analysis live on a replication continuum



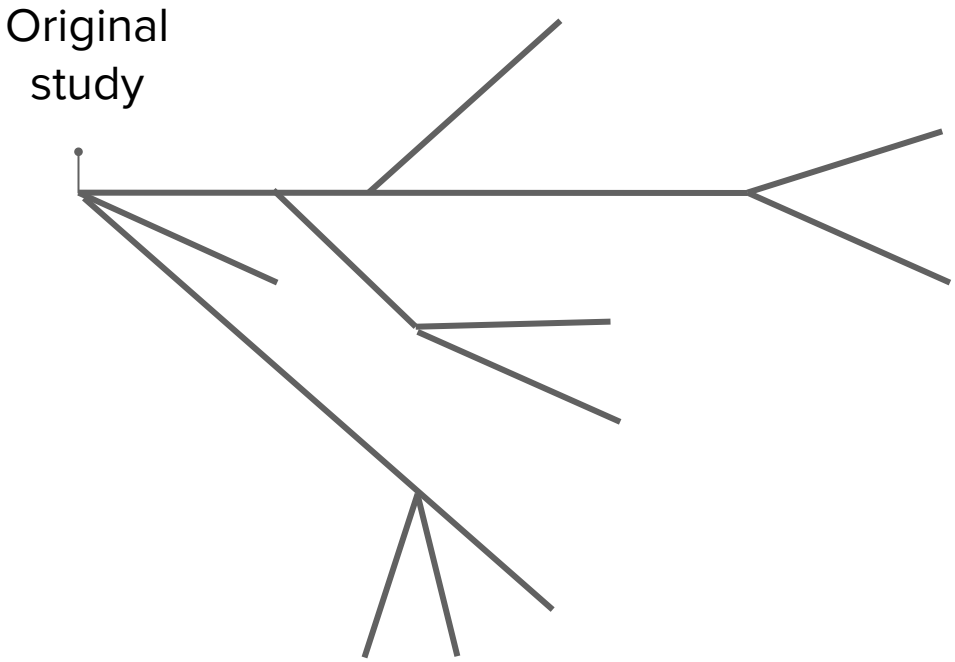
A notion of replication in meta-analyses

Groups of studies in a meta-analysis live in a replication space



A notion of replication in meta-analyses

Groups of studies in a meta-analysis live in a replication space



Accurate (statistical) modeling

➜ Determine effects of study differences / commonalities on

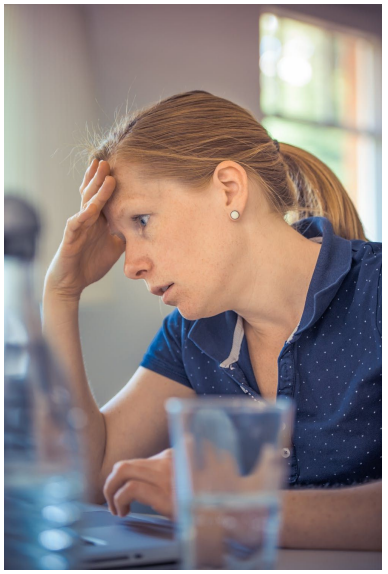
replication distance

Use CAMAs to understand replication distance?

But first: What are CAMAs?

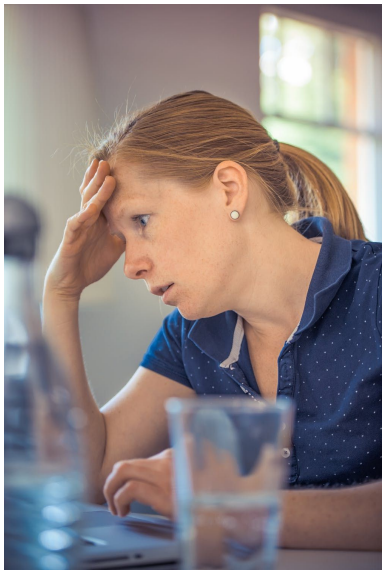
The classical meta-analysis

- Few researchers ➜ high workload



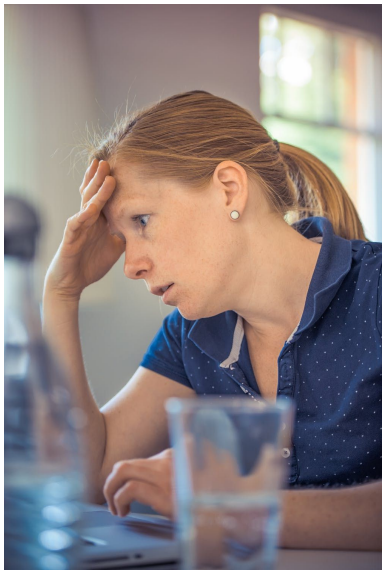
The classical meta-analysis

- Few researchers ➔ high workload
- Static, closed data
  - New study = new meta-analysis?
  - New moderator = new meta-analysis?



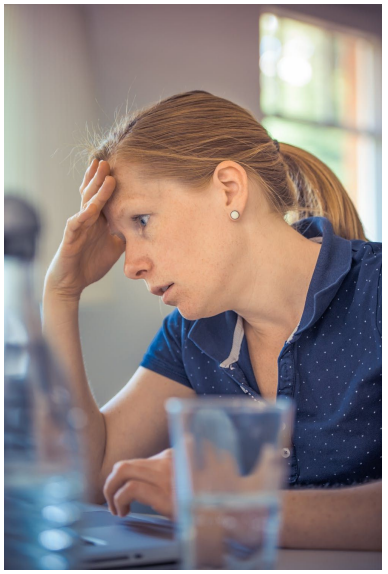
The classical meta-analysis

- Few researchers ➔ high workload
- Static, closed data
  - New study = new meta-analysis?
  - New moderator = new meta-analysis?
- Intransparent
  - Selection
  - Computation
  - Analysis decisions



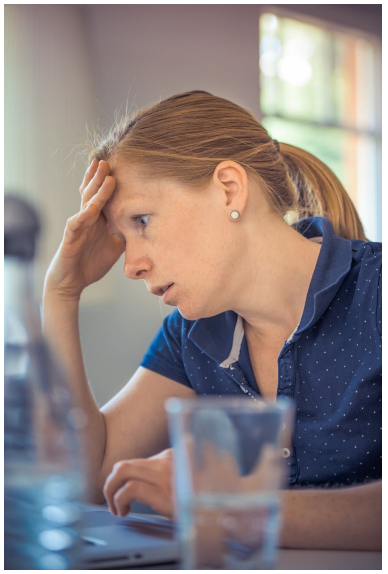
The classical meta-analysis

- Few researchers ➔ high workload
- Static, closed data
  - New study = new meta-analysis?
  - New moderator = new meta-analysis?
- Intransparent
  - Selection
  - Computation
  - Analysis decisions
- Idiosyncratic format



The classical meta-analysis

- Few researchers ➔ high workload
- Static, closed data
  - New study = new meta-analysis?
  - New moderator = new meta-analysis?
- Intransparent
  - Selection
  - Computation
  - Analysis decisions
- Idiosyncratic format
- Possible biases (see Tsuji et al., submitted)

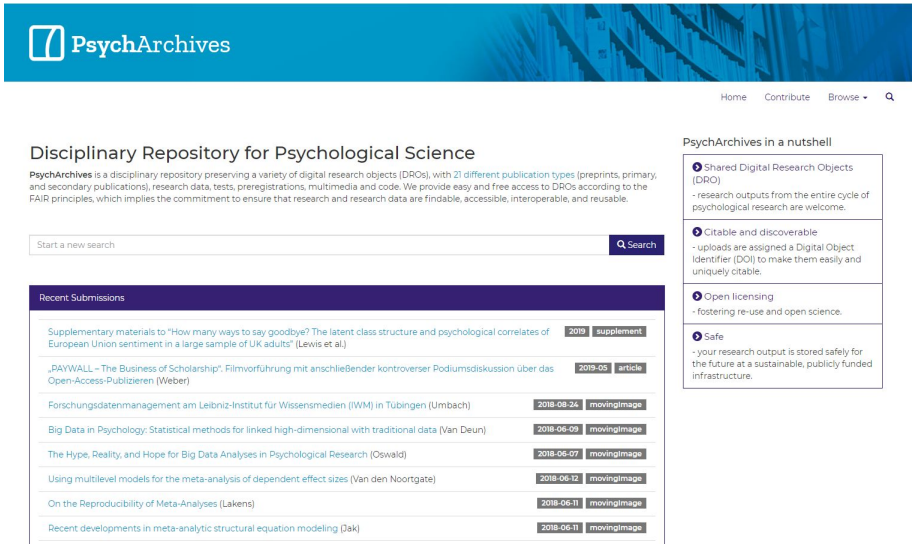


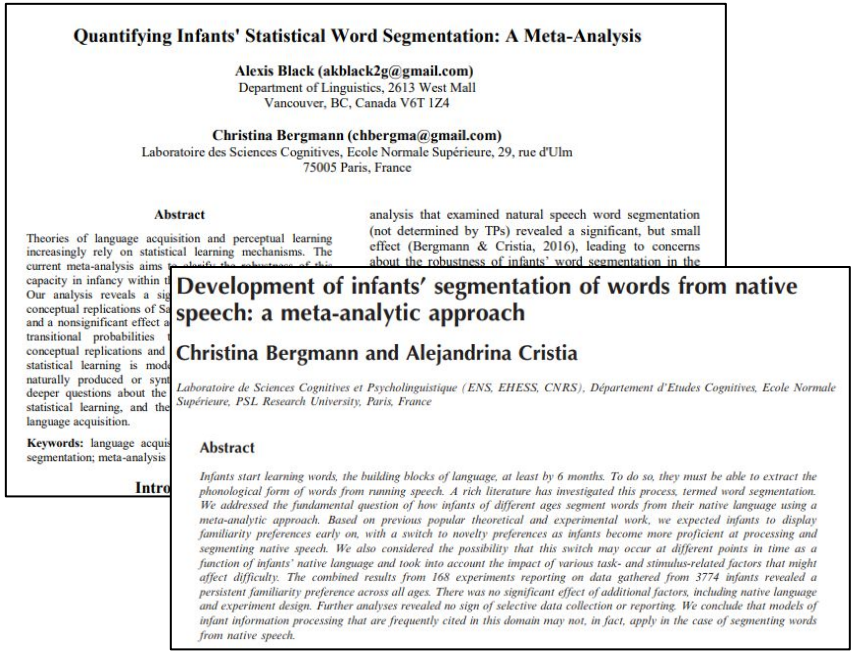
A proposal: Open and team up!

➜ Community-augmented meta-analyses (CAMAs)

Community-Augmented Meta-Analyses (CAMAs)

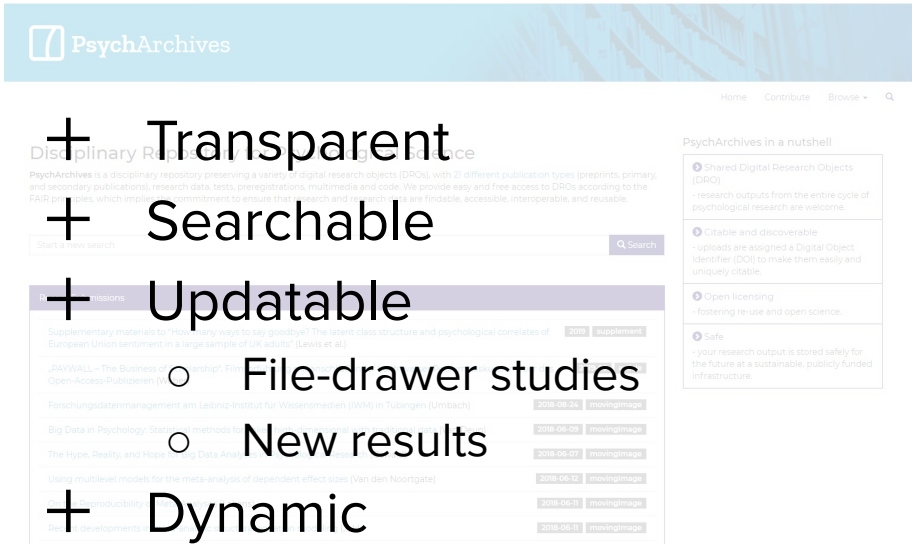
Open repository + Classical meta-analysis



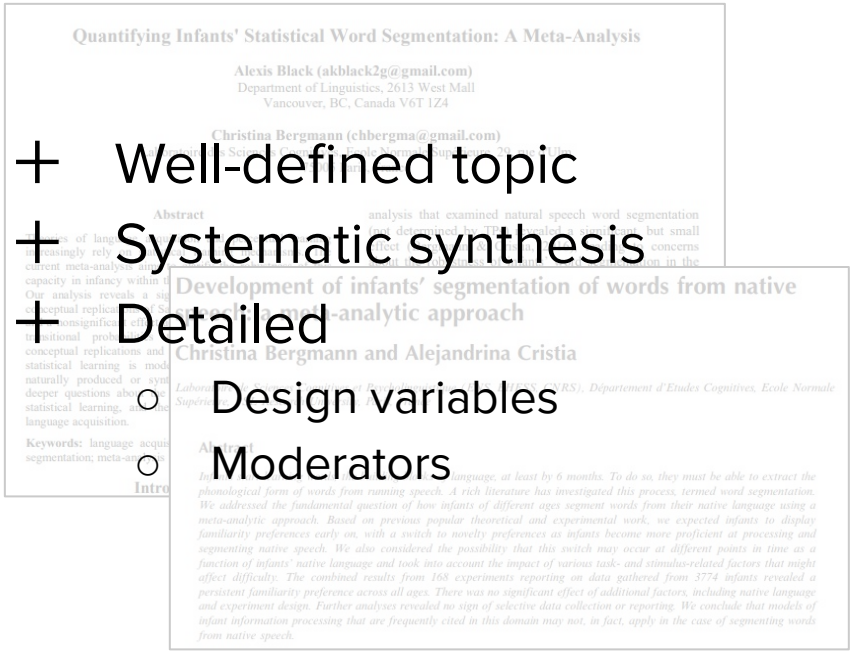


Community-Augmented Meta-Analyses (CAMAs)

Open repository + Classical meta-analysis



+ Transparent
+ Searchable
+ Updatable
  - File-drawer studies
  - New results
+ Dynamic
  - Adapt selection



+ Well-defined topic
+ Systematic synthesis
+ Detailed
  - Design variables
  - Moderators

CAMAs implemented

MetaLab

Early Language

How do children learn their native language?

19	436	1,649	26,328
Meta-analyses	Papers	Effect sizes	Participants

Cognitive Development

What is the nature of children's understanding?

3	41	155	1,941
Meta-analyses	Papers	Effect sizes	Participants

The metafor Package

A Meta-Analysis Package for R

CAMAs implemented

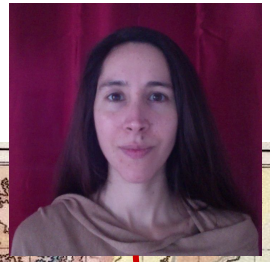
MetaLab

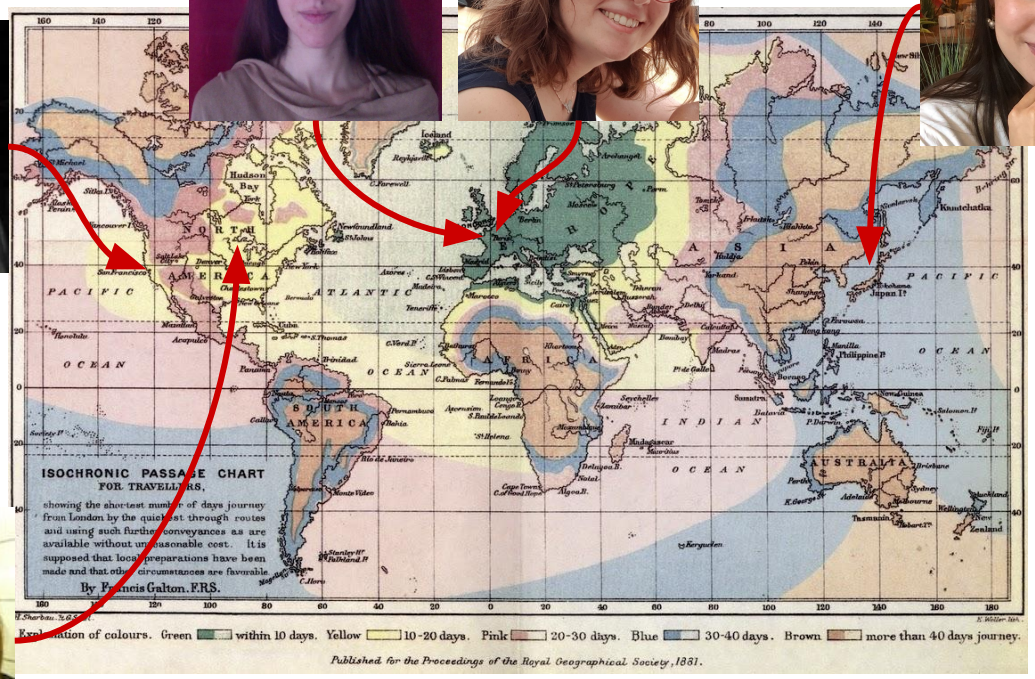
Alejandrina Cristia
ENS, Paris

Cécile Issard
ENS, Paris

Sho Tsuji
Tokyo

Michael C. Frank
Stanford



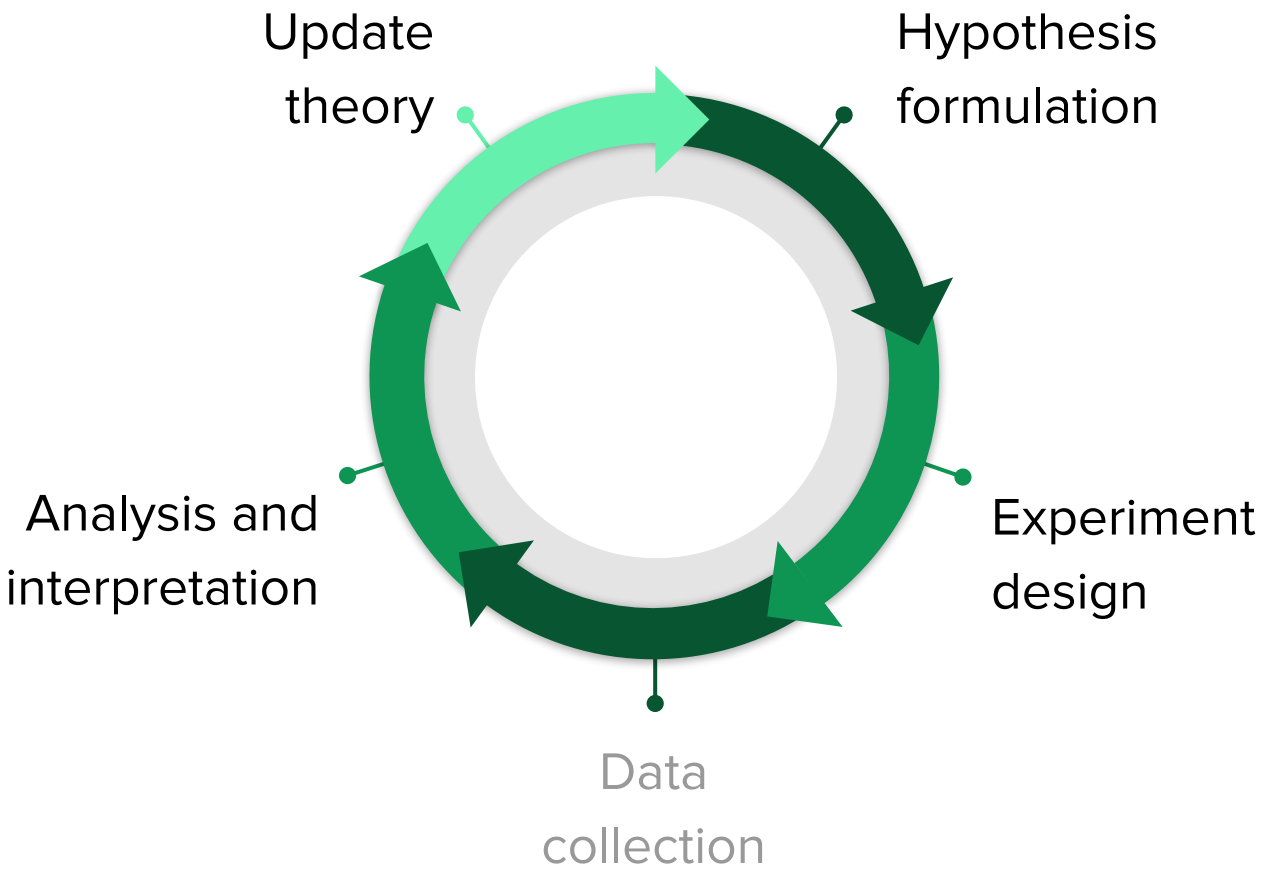


Molly L. Lewis
Chicago/Wisc.

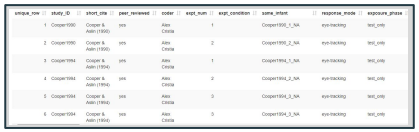
+ Team of Curators

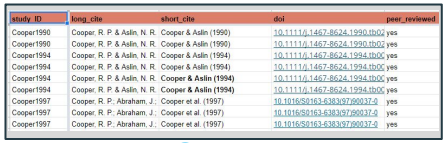
Map: Francis Galton(Life time: 17 January 1911) [Public domain]

MetaLab: Making a researcher's life easier



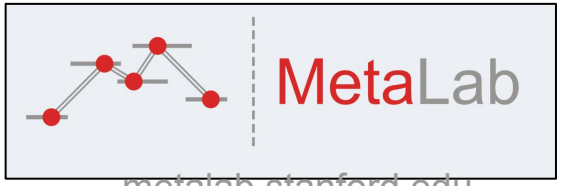
Improving meta-analyses

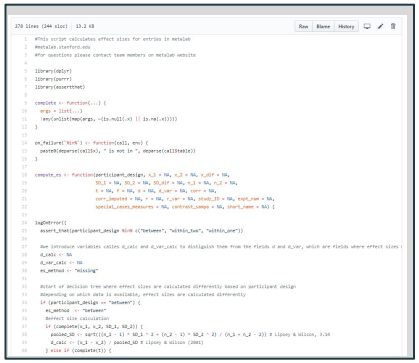




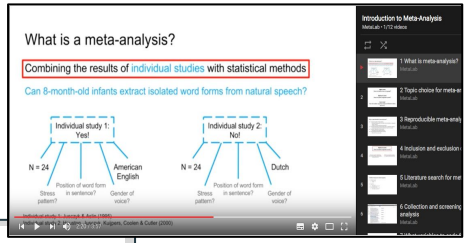
Open Datasets

Visualizations





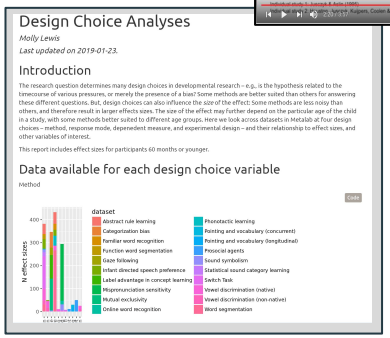
Open Code



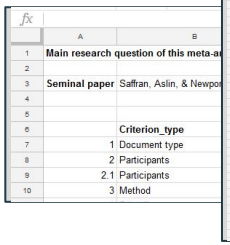
Tutorials

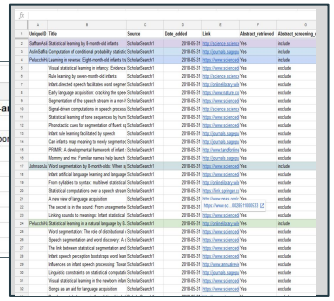
The metafor Package

A Meta-Analysis Package for R



Dynamic Reports

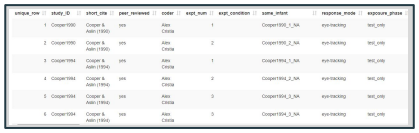


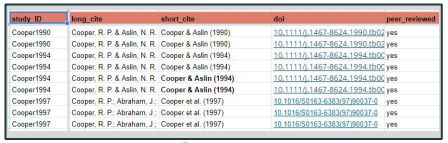


Standardized Templates

Improving meta-analyses

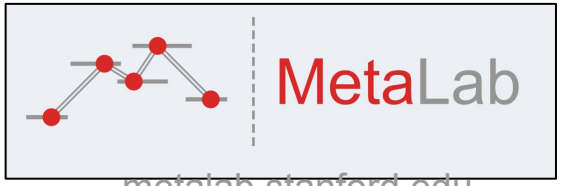
- Few researchers
- Static, closed data
  - New study = new MA?
  - New moderator = new MA?
- Intransparent
  - Selection
  - Computation
  - Analysis decisions
- Idiosyncratic format
- Possible biases



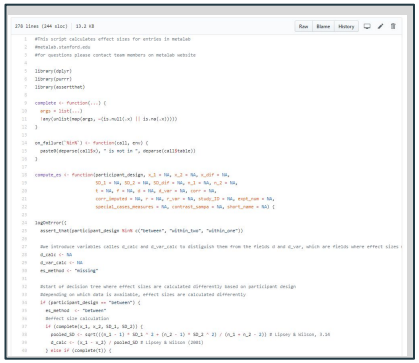


Open Datasets

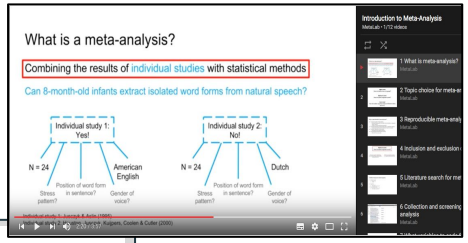
Visualizations



metalab.stanford.edu



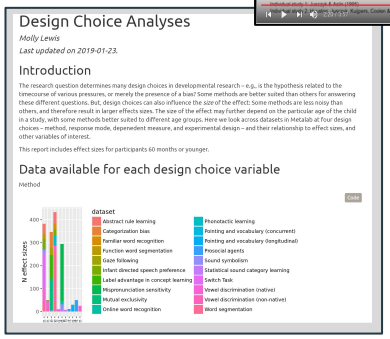
Open Code



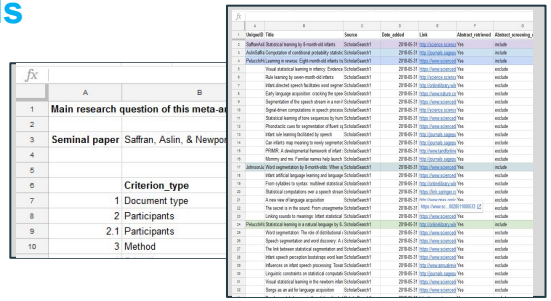
Tutorials

The metafor Package

A Meta-Analysis Package for R



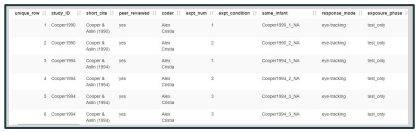
Dynamic Reports

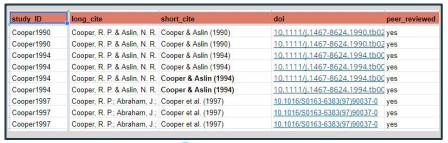


Standardized Templates

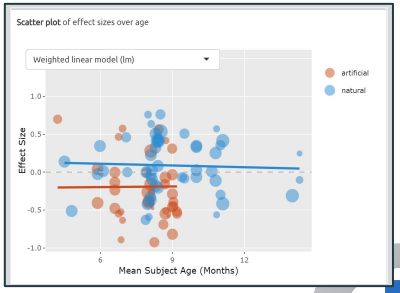
Improving meta-analyses

- ☑ Open
- ☑ Dynamic
  - ❑ New study = expand MA
  - ❑ New moderator = expand MA
- ☑ Transparent
  - ❑ Selection
  - ❑ Computation
  - ❑ Analysis decisions
- ☑ Standardized format
- ☑ Change/Check selection

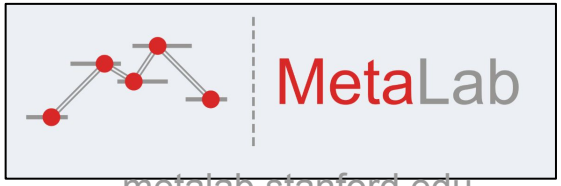




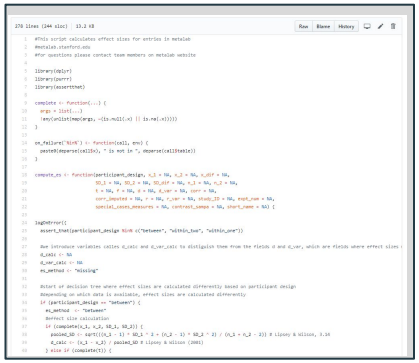
Open Datasets



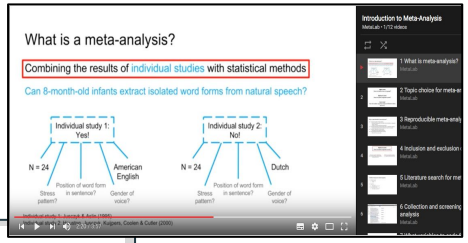
Visualizations



metalab.stanford.edu



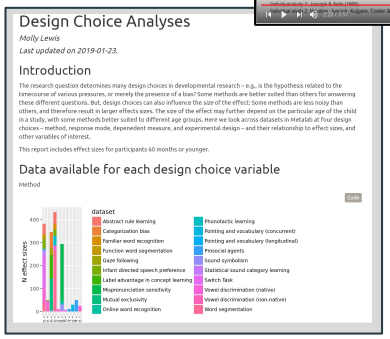
Open Code



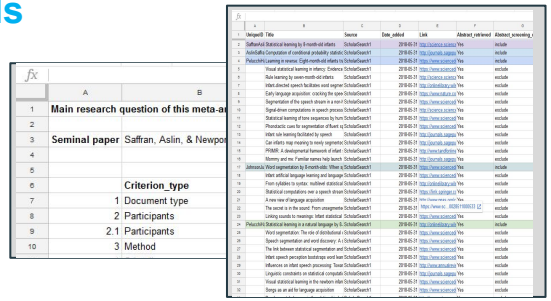
Tutorials

The metafor Package

A Meta-Analysis Package for R



Dynamic Reports



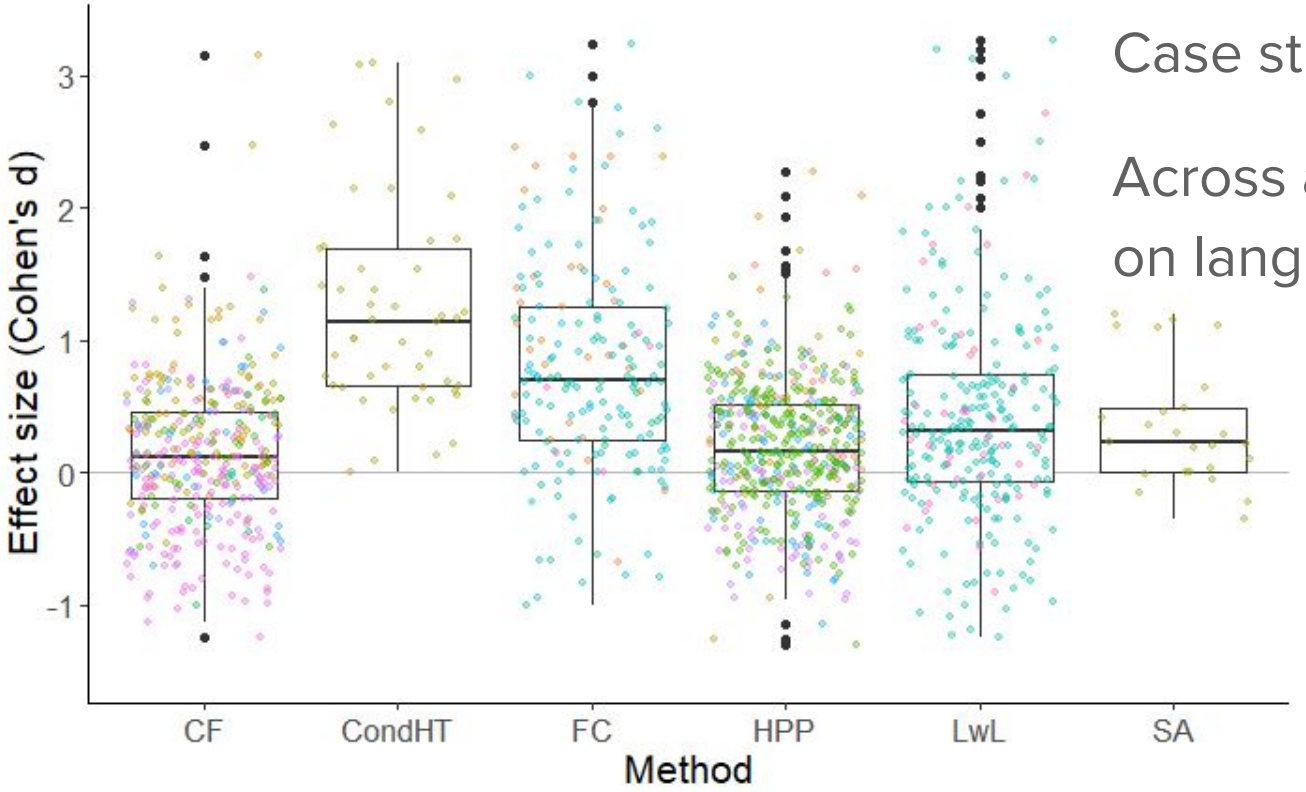
Standardized Templates

CAMAs: Tracking study differences

study_ID	response_mode	exposure_phase	method	dependent_measure	native_lang
Cooper1990	eye-tracking	test_only	CF	looking_time	American English
Cooper1990	eye-tracking	test_only	CF	looking_time	American English
Cooper1994	eye-tracking	test_only	CF	looking_time	American English
Cooper1994	eye-tracking	test_only	CF	looking_time	American English
Cooper1994	eye-tracking	test_only	CF	looking_time	American English
Cooper1994	eye-tracking	test_only	CF	looking_time	American English
Cooper1997	eye-tracking	test_only	CF	looking_time	American English
Cooper1997	eye-tracking	test_only	CF	looking_time	American English
Cooper1997	eye-tracking	test_only	CF	looking_time	American English
Droucker2013	eye-tracking	familiarization	CF	looking_time	Canadian English
Droucker2013	eye-tracking	familiarization	CF	looking_time	Canadian English
Droucker2013	eye-tracking	familiarization	CF	looking_time	Canadian English
Droucker2013	eye-tracking	familiarization	CF	looking_time	Canadian English
Droucker2013	eye-tracking	familiarization	CF	looking_time	Canadian English
Fernald1985	behavior	conditioning	CHT	target_selection	American English
Fernald1987	behavior	conditioning	CHT	target_selection	American English
Fernald1987	behavior	conditioning	CHT	target_selection	American English
Fernald1987	behavior	conditioning	CHT	target_selection	American English
Glenn1983	behavior	conditioning	FC	target_selection	British English
Glenn1983	behavior	conditioning	FC	target_selection	British English
Hayashi2001	behavior	familiarization	HPP	looking_time	Japanese
Hayashi2001	behavior	familiarization	HPP	looking_time	Japanese
Hayashi2001	behavior	familiarization	HPP	looking_time	Japanese
Inoe2011	behavior	familiarization	HPP	looking_time	Japanese
Inoe2011	behavior	familiarization	HPP	looking_time	Japanese

1. Standardized
2. Expandable

CAMAs: Quantify the impact of study differences

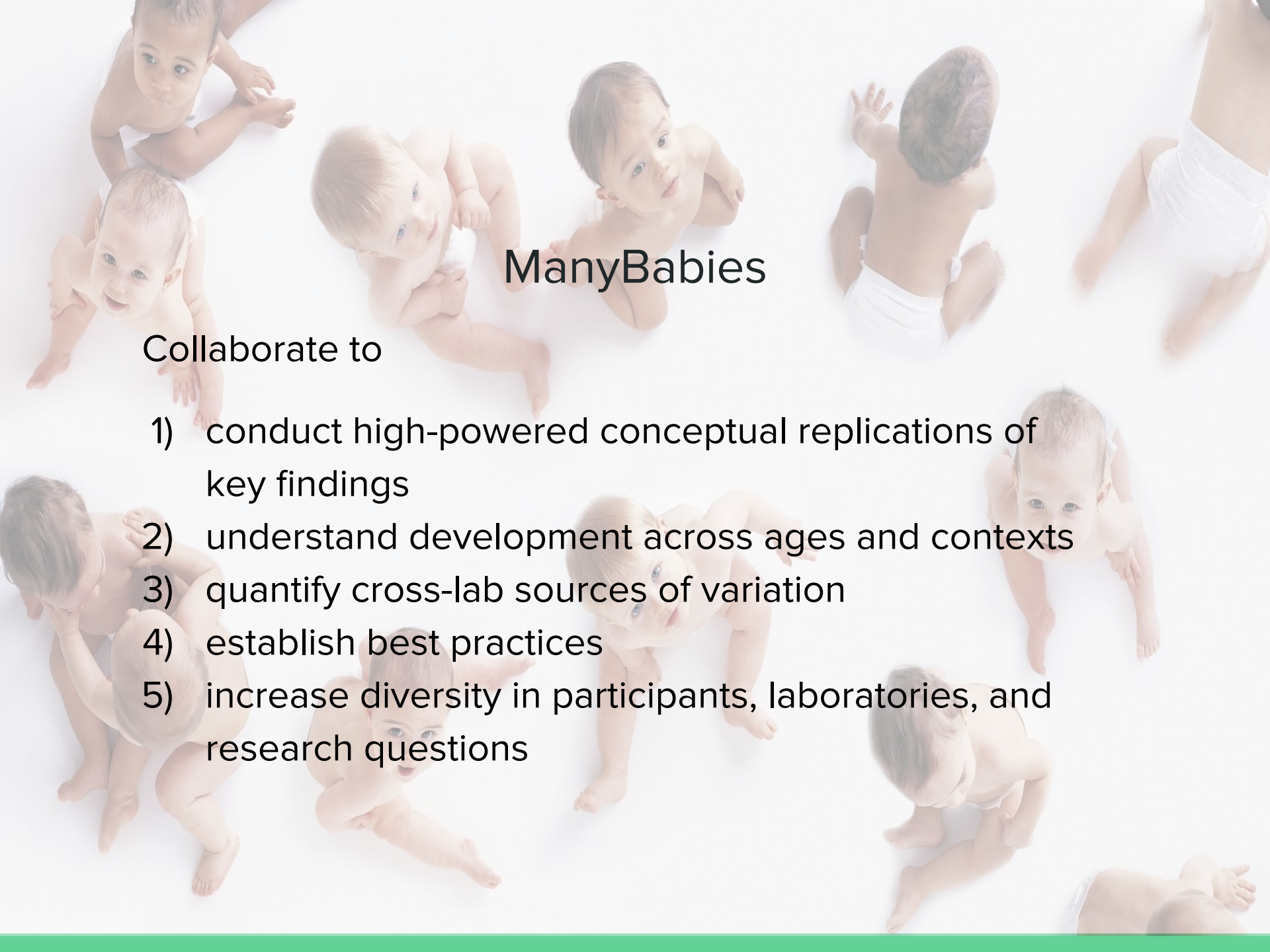


Case study: Method

Across **all** meta-analyses on language development

Updated from Bergmann et al. (2018)

CAMAs in the times of large-scale replication projects



ManyBabies

Collaborate to

1) conduct high-powered conceptual replications of key findings
2) understand development across ages and contexts
3) quantify cross-lab sources of variation
4) establish best practices
5) increase diversity in participants, laboratories, and research questions

ManyBabies 1 - Goals

- Test a key phenomenon

ManyBabies 1 - Goals

- Test a key phenomenon
- Quantify differences across labs

ManyBabies 1 - Goals

- Test a key phenomenon
- Quantify differences across labs
  - Standardize where possible

  & Investigate effects of systematic and incidental differences

  ➜ Most prominent: Method effects

Meta-analysis for planning ManyBabies 1

- ➔ Exploit literature overview
- ➔ Selecting efficient design and stimuli

Infant-Directed Speech Preference

Infant-Directed Speech (IDS): higher, slower, more variable in pitch, preferred by infants over adult-directed speech

Infant-Directed Speech Preference

Infant-Directed Speech (IDS): higher, slower, more variable in pitch, preferred by infants over adult-directed speech

Theoretical importance

- ➢ Universal? Language Specific?
- ➢ Signal tailored for learning language?

Infant-Directed Speech Preference

Infant-Directed Speech (IDS): higher, slower, more variable in pitch, preferred by infants over adult-directed speech

Theoretical importance

- ➢ Universal? Language Specific?
- ➢ Signal tailored for learning language?

Practical relevance

- ➢ Key recommendation for parents
- ➢ (Almost) All infant studies use IDS

Infant-Directed Speech Preference: Meta-Analysis

Studies: 71 (from 23 papers)

Age: 0-18 months

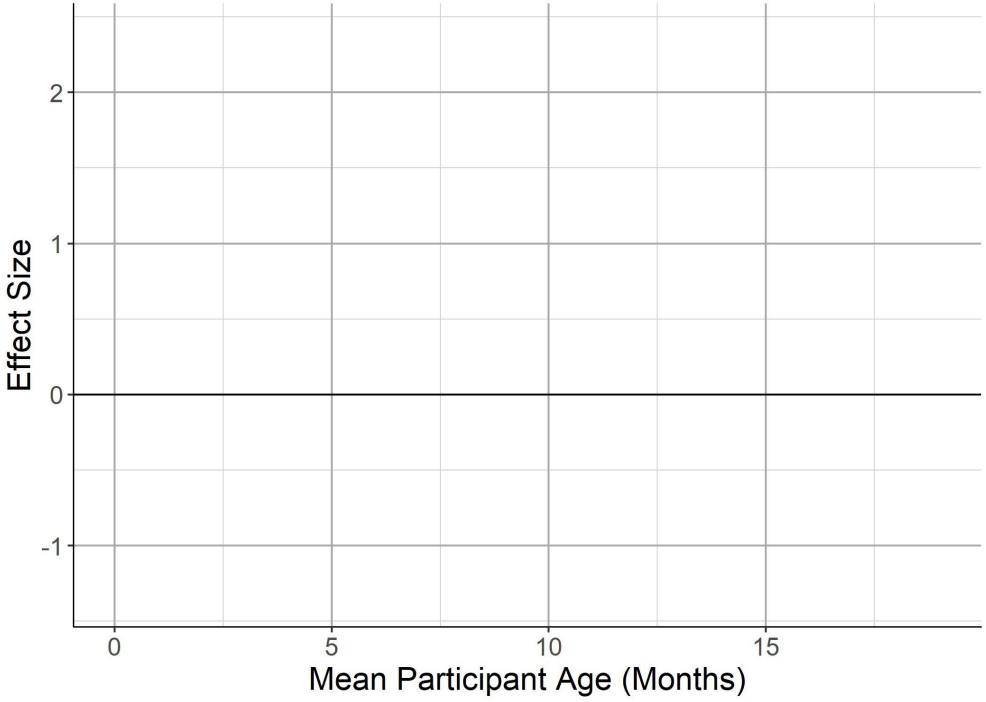
Effect size: d = 0.62, se = 0.1

Sample size: 20 (9-60)

Average power: 69%

Note: Updated from the paper, original meta-analysis: Dunst, Gorman, & Hamby (2012)

Infant-Directed Speech Preference: Meta-Analysis



Studies: 71 (from 23 papers)

Age: 0-18 months

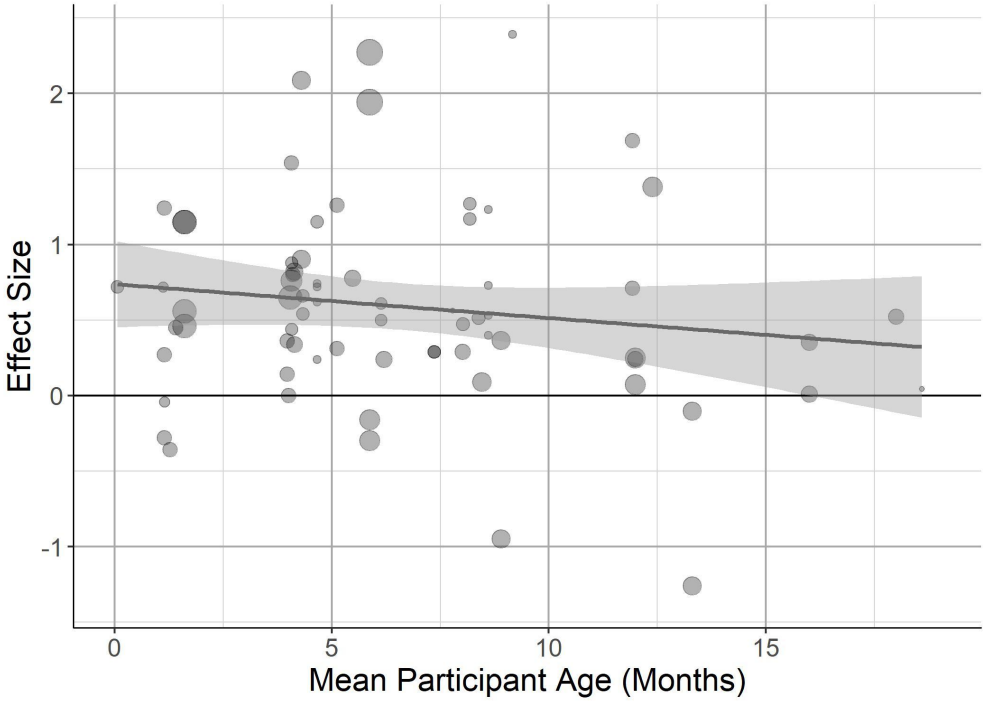
Effect size: *d* = 0.62, se = 0.1

Sample size: 20 (9-60)

Average power: 69%

Note: Updated from the paper, original meta-analysis: Dunst, Gorman, & Hamby (2012)

Infant-Directed Speech Preference: Meta-Analysis



Studies: 71 (from 23 papers)

Age: 0-18 months

Effect size: *d* = 0.62, se = 0.1

Sample size: 20 (9-60)

Average power: 69%

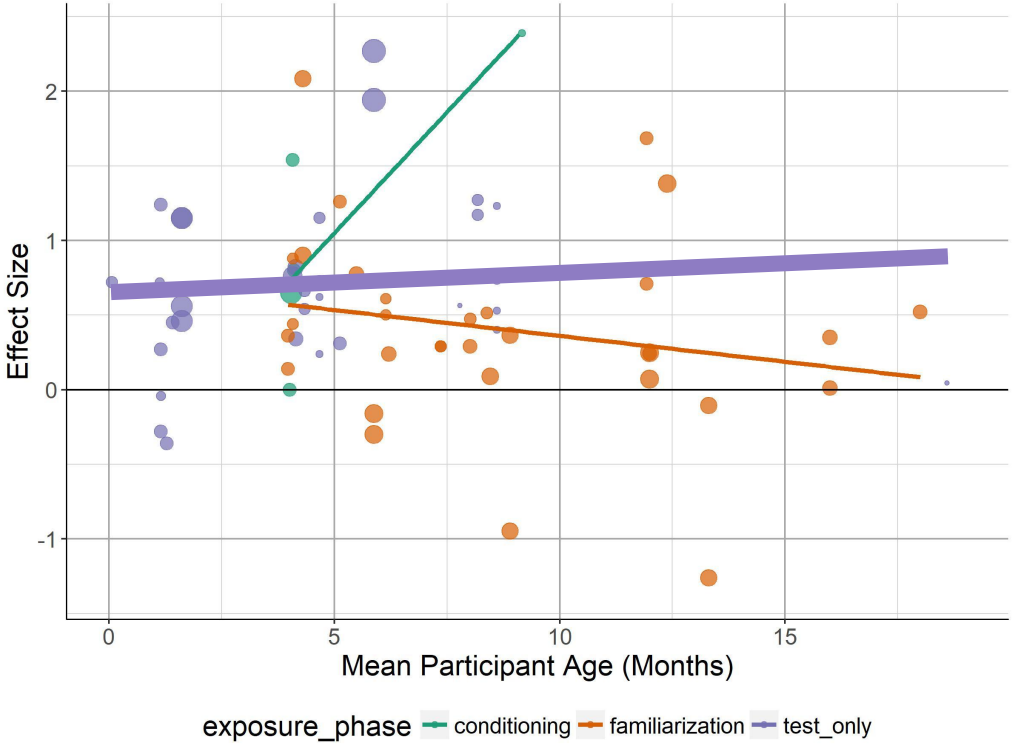
Note: Updated from the paper, original meta-analysis: Dunst, Gorman, & Hamby (2012)

Selecting design and stimuli

➔ Identify most "effective" way to test IDS preference

Note: Not a close replication

IDS preference - Design

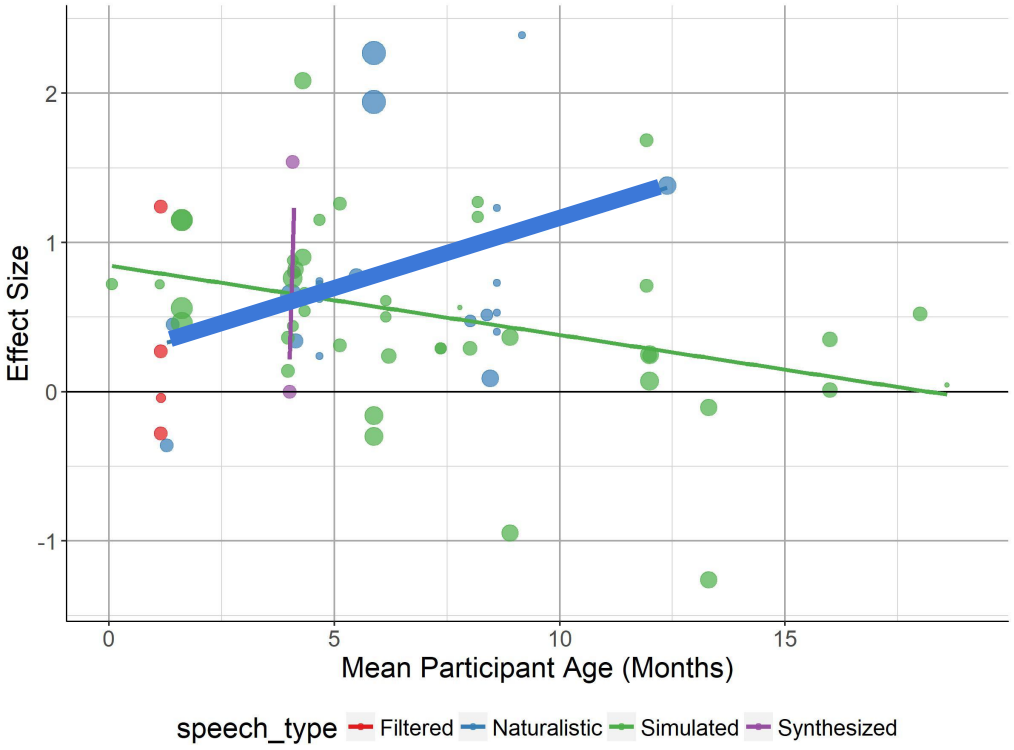


conditioning (n = 5) >

test only (n = 33) >

familiarization (n = 32)

IDS preference - Stimuli



Stimulus effect (interaction with age):

naturalistic (n = 20) > simulated (n = 43)

Other options: filtered, synthesized

Re-designing the optimal study: ManyBabies 1

Design = test only (no exposure phase)

Stimuli = naturalistic

Re-designing the optimal study: ManyBabies 1

Design = test only (no exposure phase)

Stimuli = naturalistic

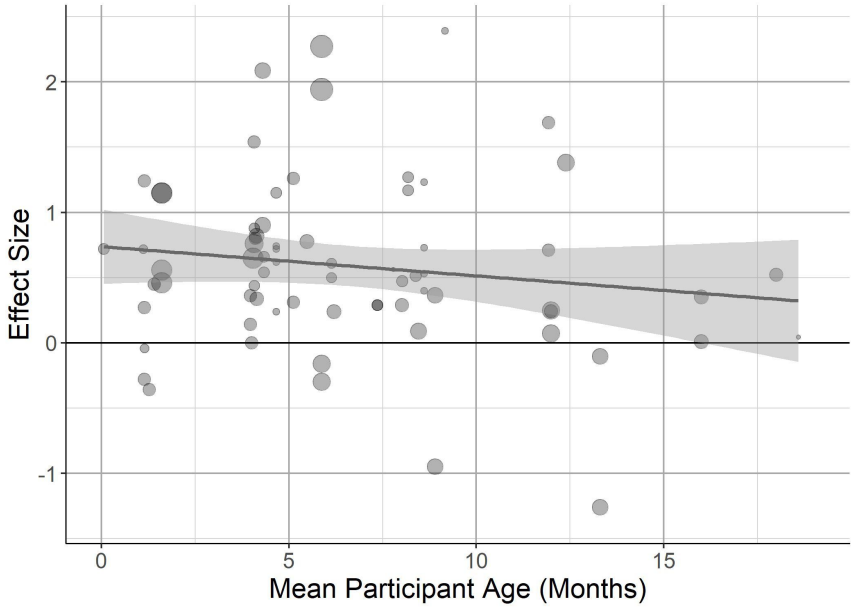
➜ Subset meta-analysis

Re-designing the optimal study: ManyBabies 1

Design = test only (no exposure phase)

Stimuli = naturalistic

➜ Subset meta-analysis

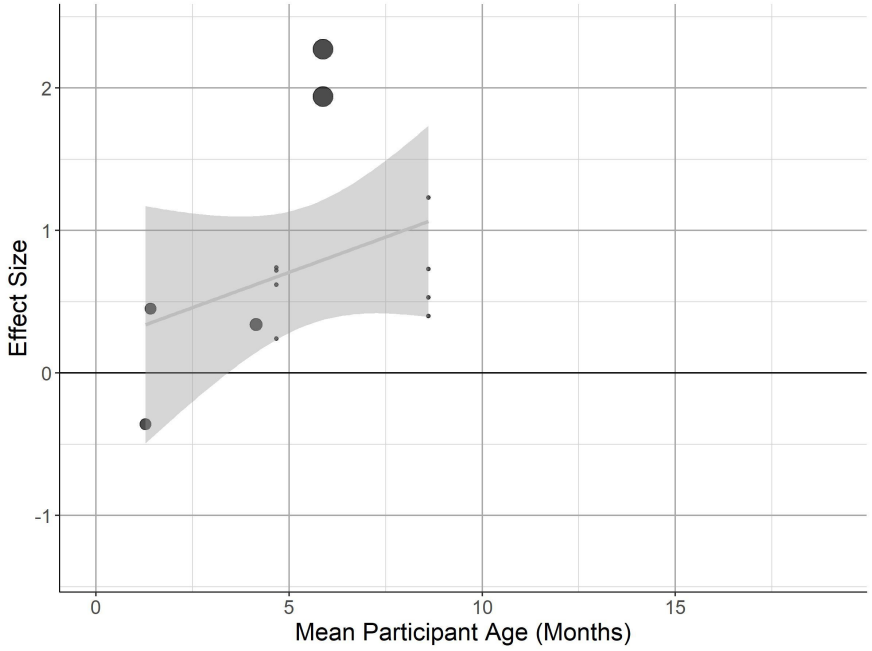


Re-designing the optimal study: ManyBabies 1

Design = test only (no exposure phase)

Stimuli = naturalistic

➜ Subset meta-analysis



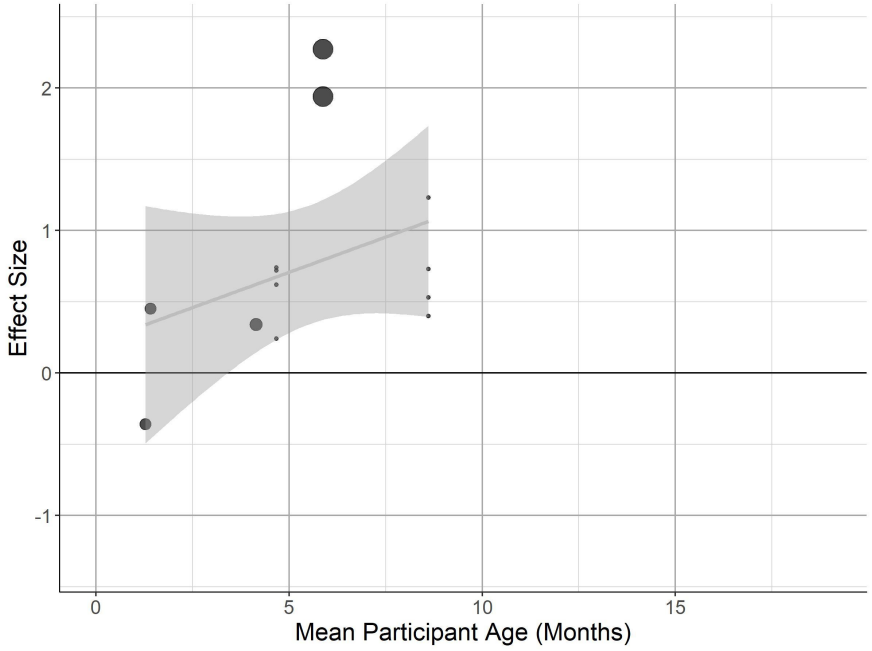
Re-designing the optimal study: ManyBabies 1

Design = test only (no exposure phase)

Stimuli = naturalistic

➜ Subset meta-analysis

New effect size: 0.66 (vs 0.62)



Re-designing the optimal study: ManyBabies 1

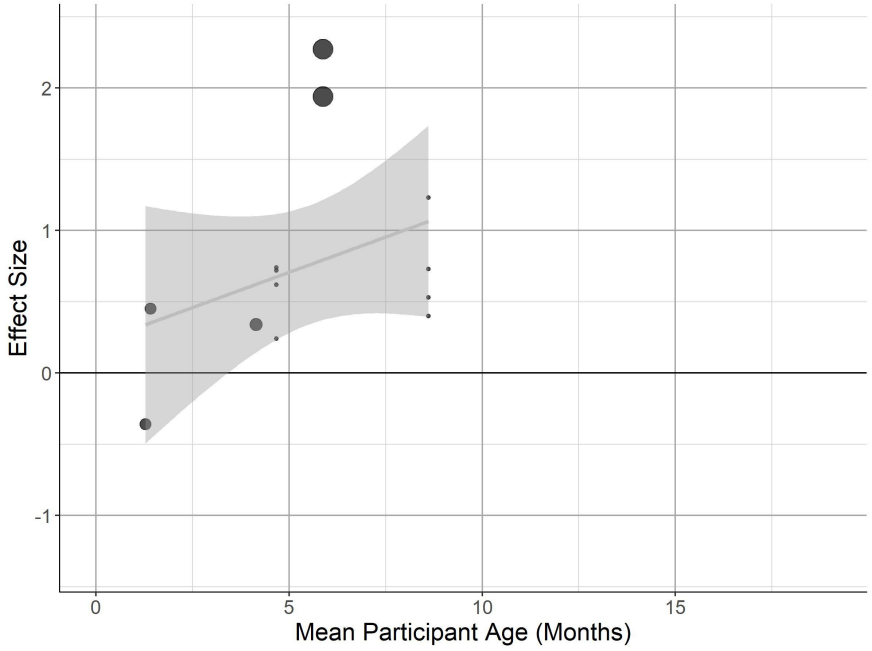
Design = test only (no exposure phase)

Stimuli = naturalistic

➜ Subset meta-analysis

New effect size: 0.66 (vs 0.62)

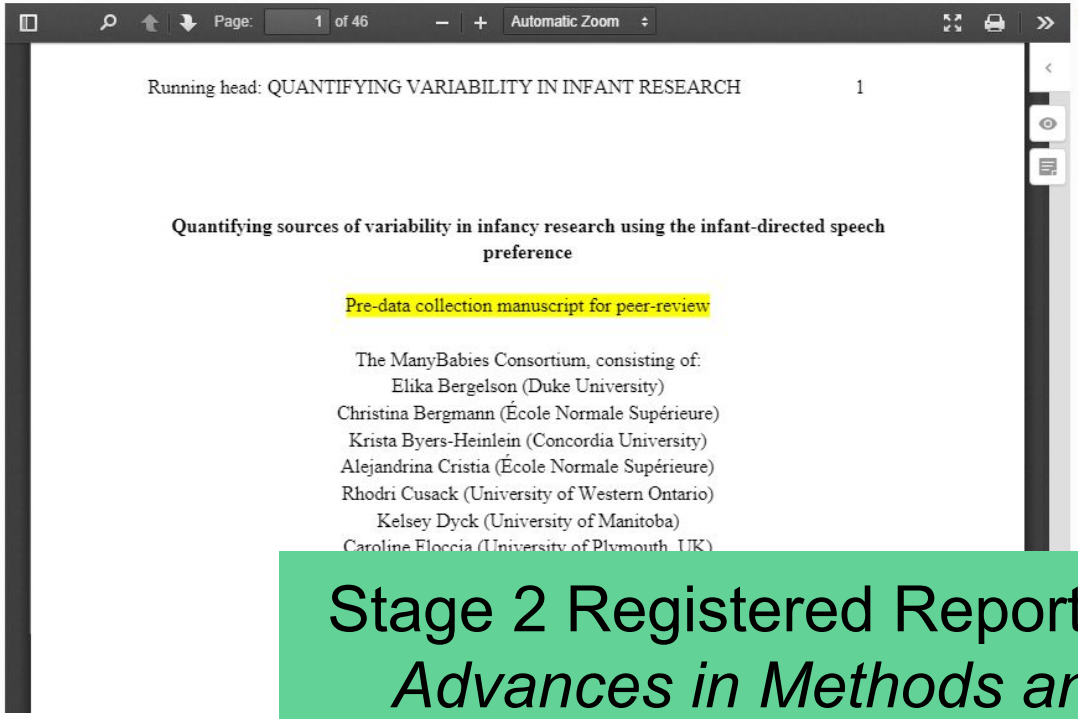
Note: Interaction with age positive



Quantifying sources of variability in infancy research using the infant-directed speech preference

AUTHORS
Elika Bergelson, Christina Bergmann, Krista Byers-Heinlein, Alejandrina Cristia, Rhodri Cusack, Kelsey Dyck, caroline floccia, Michael Frank, Judit Gervain, Nayeli Gonzalez, Kiley Hamlin, Erin Hannon, Danielle Kellier, Melissa Kline, Casey Lew-Williams, Thierry Nazzi, Robin Panneton, Hugh Rabagliati, Jennifer Rennels, Amanda Seidl, Melanie Soderstrom, Daniel Yurovsky

CREATED ON April 14, 2017 **LAST EDITED** July 02, 2018 **SUPPLEMENTAL MATERIALS** osf.io/ctvd5/



Page: 1 of 46 Automatic Zoom

Running head: QUANTIFYING VARIABILITY IN INFANT RESEARCH 1

Quantifying sources of variability in infancy research using the infant-directed speech preference

Pre-data collection manuscript for peer-review

The ManyBabies Consortium, consisting of:
Elika Bergelson (Duke University)
Christina Bergmann (École Normale Supérieure)
Krista Byers-Heinlein (Concordia University)
Alejandrina Cristia (École Normale Supérieure)
Rhodri Cusack (University of Western Ontario)
Kelsey Dyck (University of Manitoba)
Caroline Floccia (University of Plymouth, UK)

Download preprint Downloads: 857

Abstract

The field of psychology has become increasingly concerned with issues related to methodology and replicability. Infancy researchers face specific challenges related to replicability: high-powered studies are difficult to conduct, testing conditions vary across labs, and different labs have access to different infant populations, amongst other ...

See more

Preprint DOI

10.31234/osf.io/s98ab

License

Stage 2 Registered Report under review at *Advances in Methods and Practices in Psychological Science*

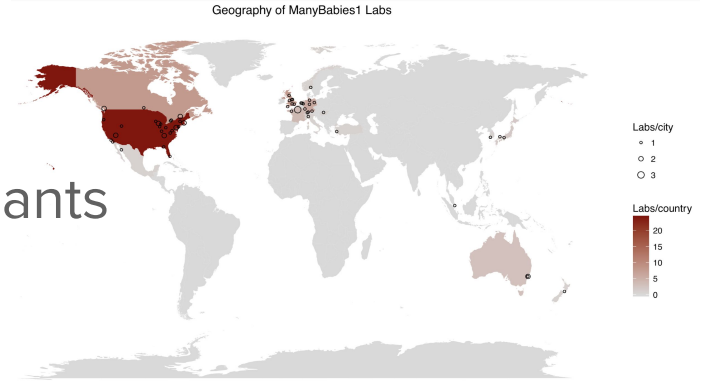
Stage 1 preprint: https://psyarxiv.com/s98ab/

Final Sample ManyBabies1

Number of Participating labs: 69 contributed data (67 in final sample)

Number of Countries: 17

Final Sample with Exclusions: **2329** participants



North American English: N = 1066

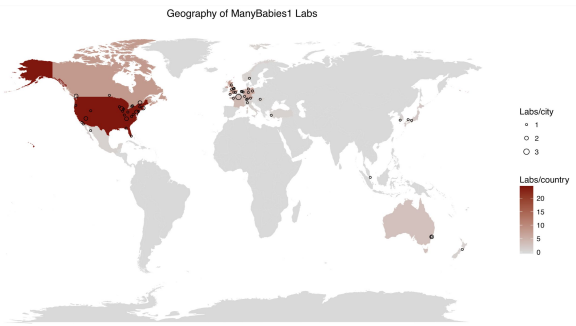
Non-North American English: N = 1263

Meta-Analyses versus Replication

- ★ Meta-Analysis:
  - Incidental variation and uncontrolled co-variation
    - Stimuli
    - Method
    - Population (American / British / Canadian English; Cantonese; Japanese)

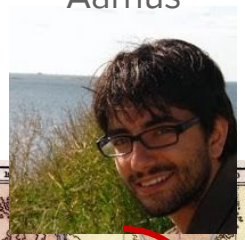
Meta-Analyses versus Replication

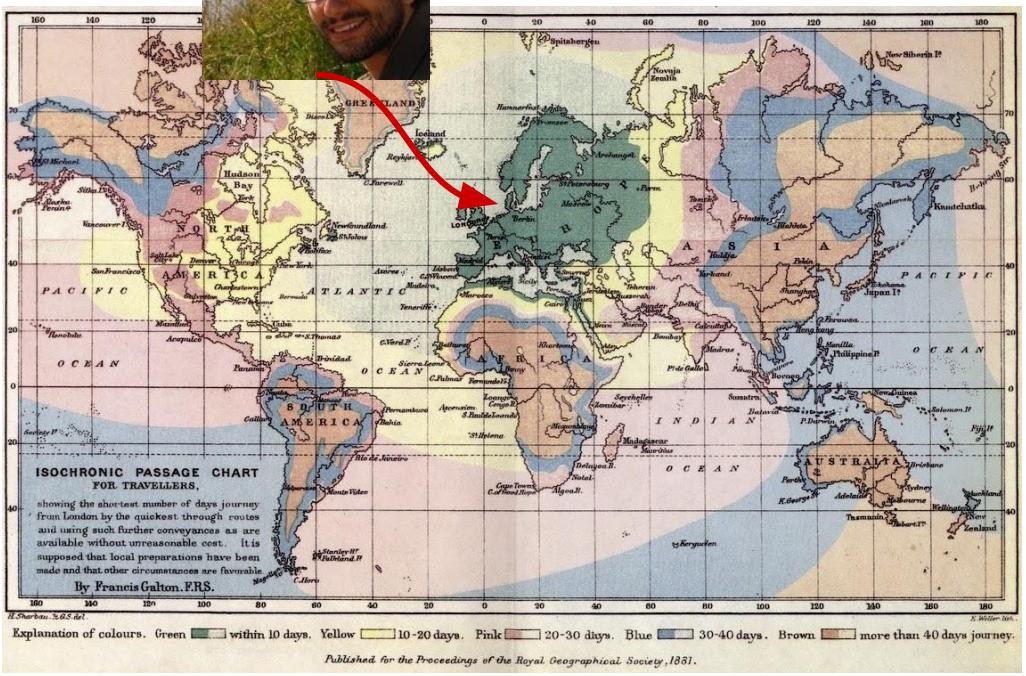
- ★ Meta-Analysis:
  - Incidental variation and uncontrolled co-variation
    - Stimuli
    - Method
    - Population (American / British / Canadian English; Cantonese; Japanese)
- ★ ManyBabies 1:
  - Standardization (as much as possible)
    - 1 stimulus set
    - 3 Methods
  - More diverse population (16 languages)
  - New variable (nativeness)



Meta-Analyses versus Replication

Riccardo Fusaroli
Aarhus





Map: Francis Galton(Life time: 17 January 1911) [Public domain]

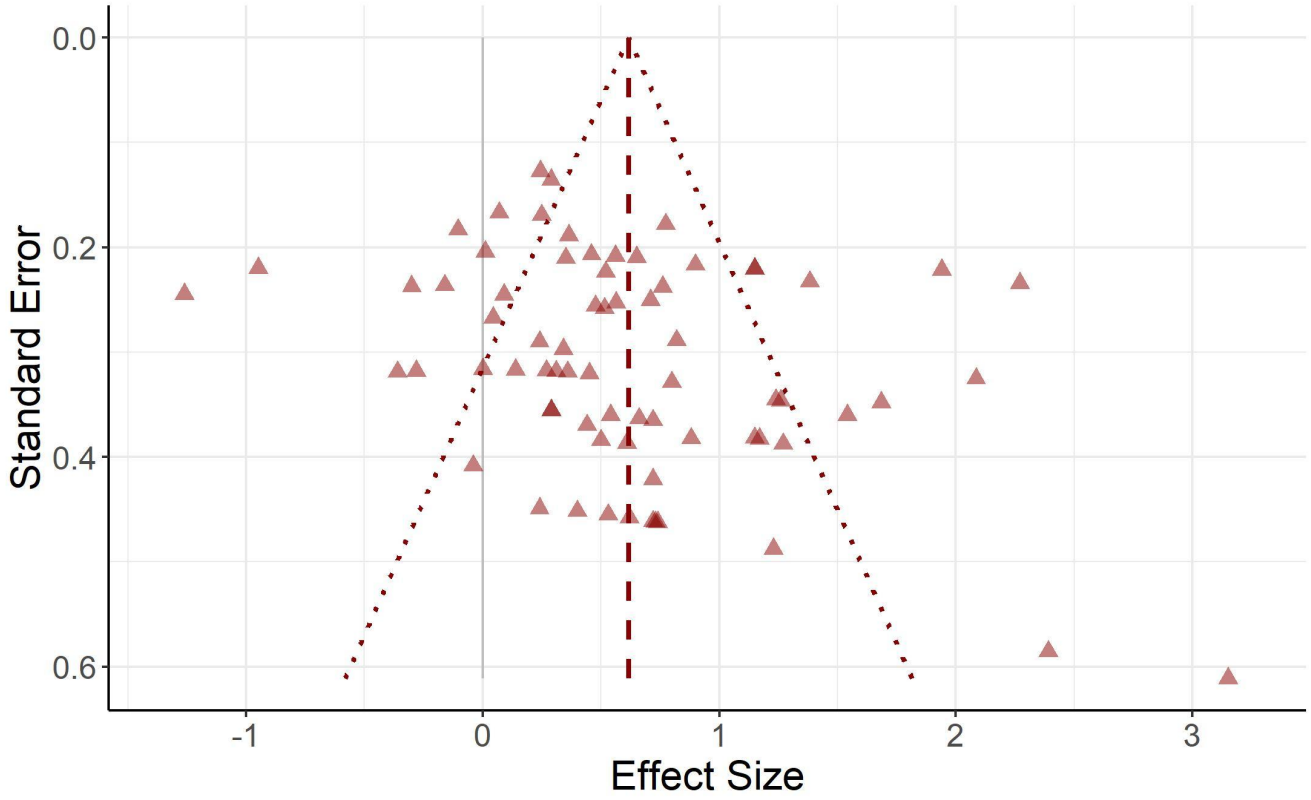
Goals

1. Compare the main effect size to the meta-analysis
2. Consistent moderator effects - method?

Visual comparison

Meta-analysis

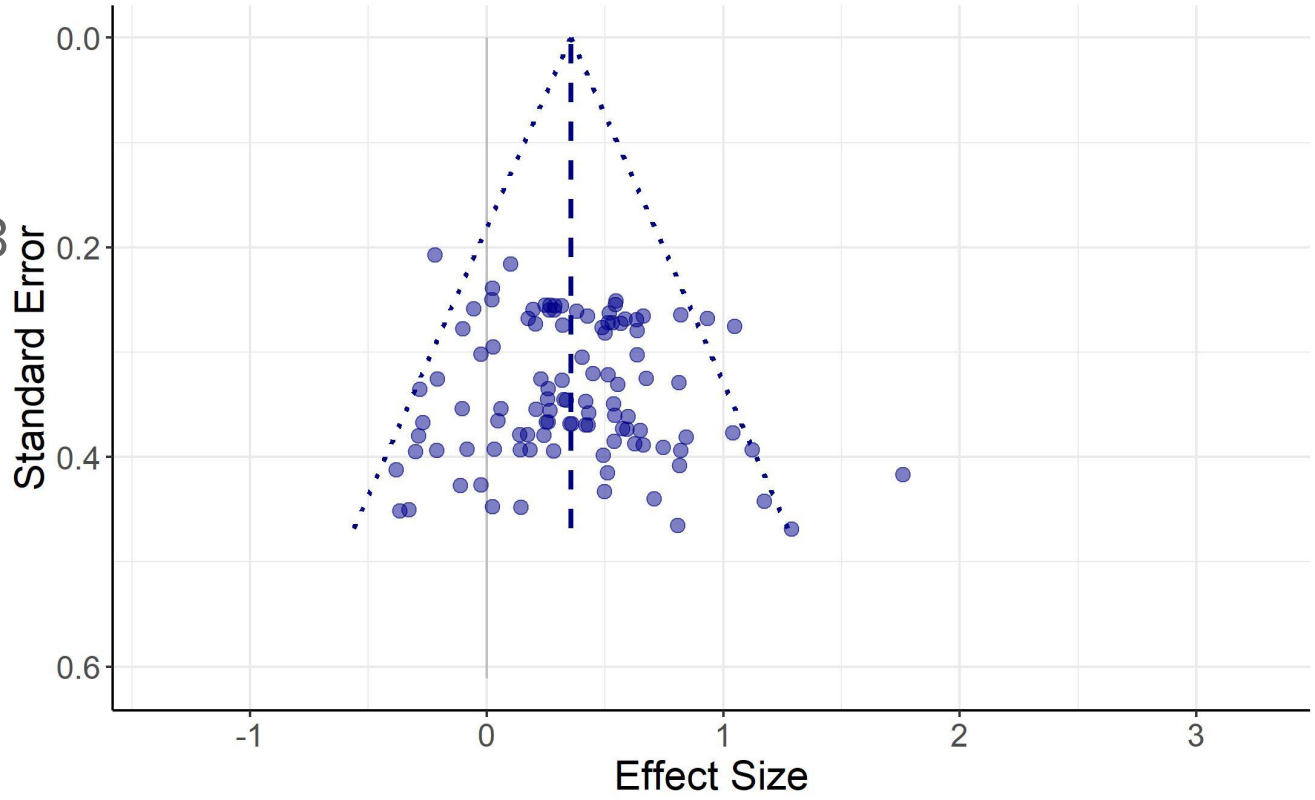
d = 0.62, se = 0.1



Visual comparison

ManyBabies 1

d = 0.36, se = 0.03



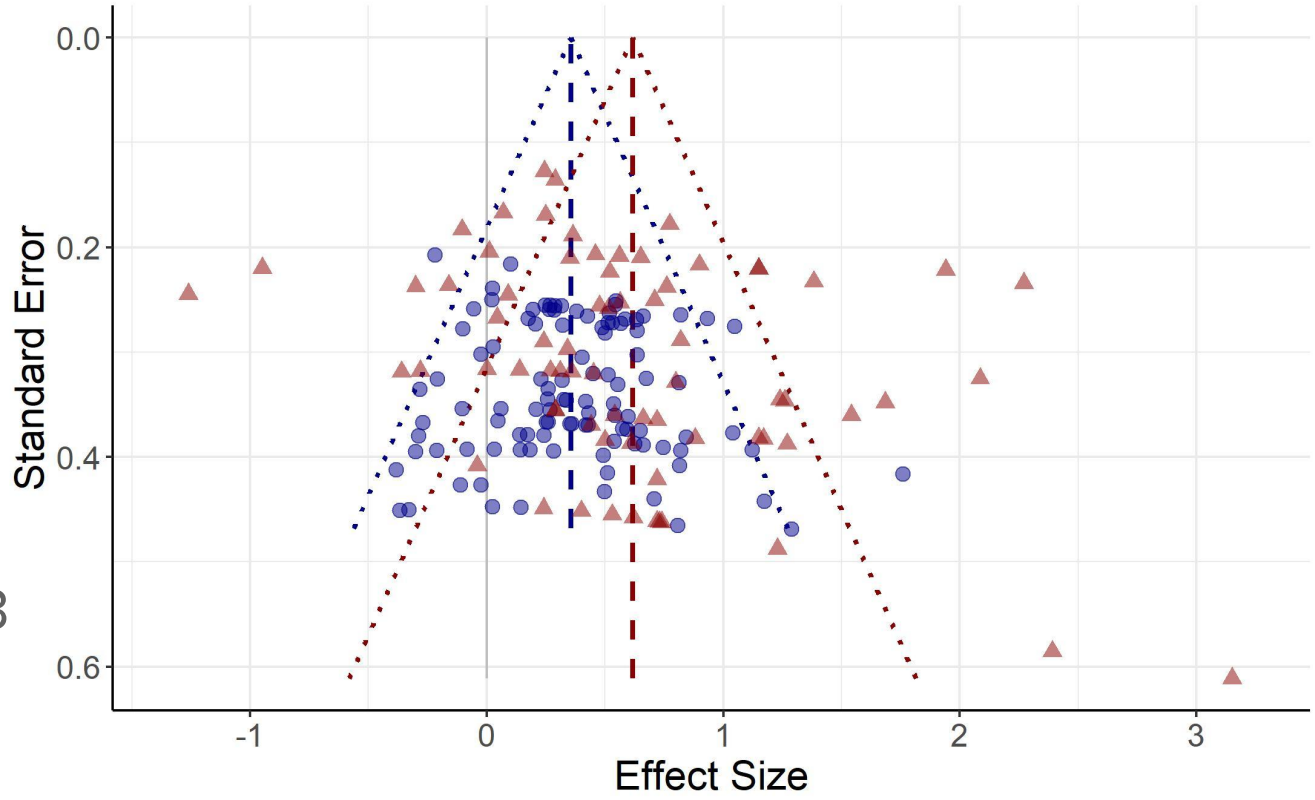
Visual comparison

Meta-analysis

d = 0.62, se = 0.1

ManyBabies 1

d = 0.36, se = 0.03



Predicting ManyBabies overall effect size

Bayesian approach:

Skeptic prior versus informed prior

⇓ ⇓

m = 0, sd = 0.3 m = 0.62, sd 0.1

Predicting ManyBabies overall effect size

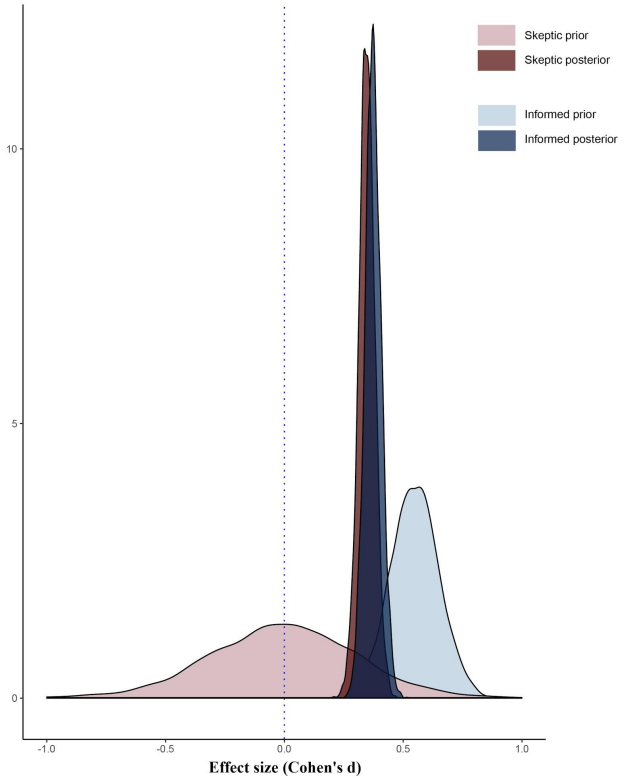
Bayesian approach:

Skeptic prior versus informed prior

⇓ ⇓

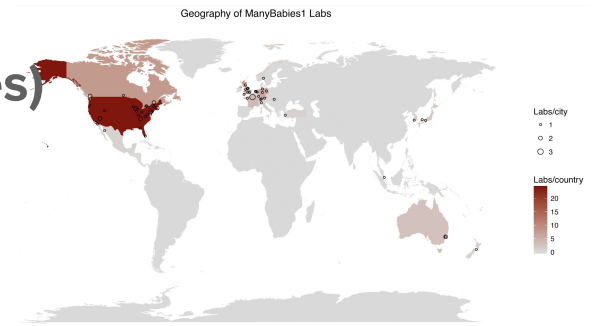
m = 0, sd = 0.3 m = 0.62, sd 0.1

➜ No gain from informed prior



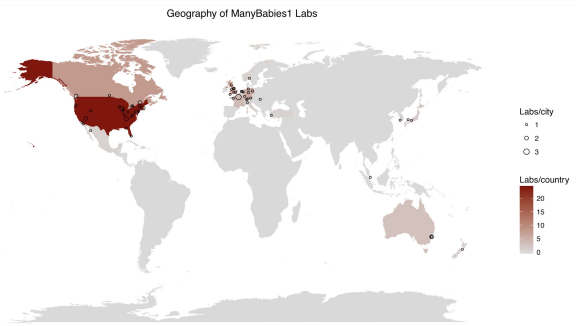
Meta-Analyses versus Replication: Main effect size

- ★ Meta-Analysis:
  - **Incidental variation and uncontrolled co-variation**
    - **Stimuli**
    - **Method**
    - **Population** (American / British / Canadian English; Cantonese; Japanese)
- ★ ManyBabies 1:
  - **Standardization (as much as possible)**
    - **1 stimulus set**
    - 3 Methods
  - **More diverse population (16 languages)**
  - **New variable (nativeness)**



Meta-Analyses versus Replication: Moderators

- ★ Meta-Analysis:
  - Incidental variation and uncontrolled co-variation
    - Stimuli
    - **Method (5, 2 dominant)**
    - Population (American / British / Canadian English; Cantonese; Japanese)
- ★ ManyBabies 1:
  - Standardization (as much as possible)
    - 1 stimulus set
    - **3 Methods, evenly distributed**
  - More diverse population (16 languages)
  - New variable (nativeness)

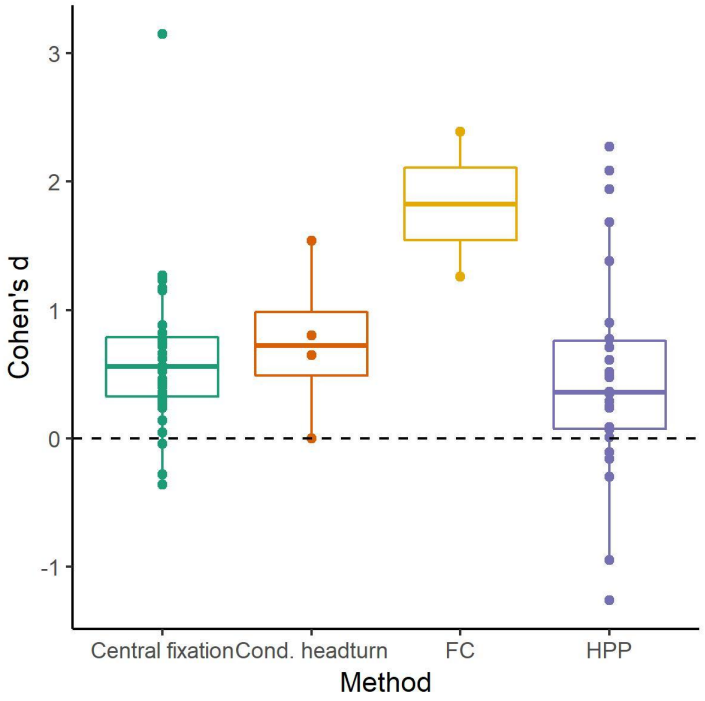


Predicting moderators: Method effect

Skeptic prior versus informed prior

⇓

m = 0, sd = 0.3

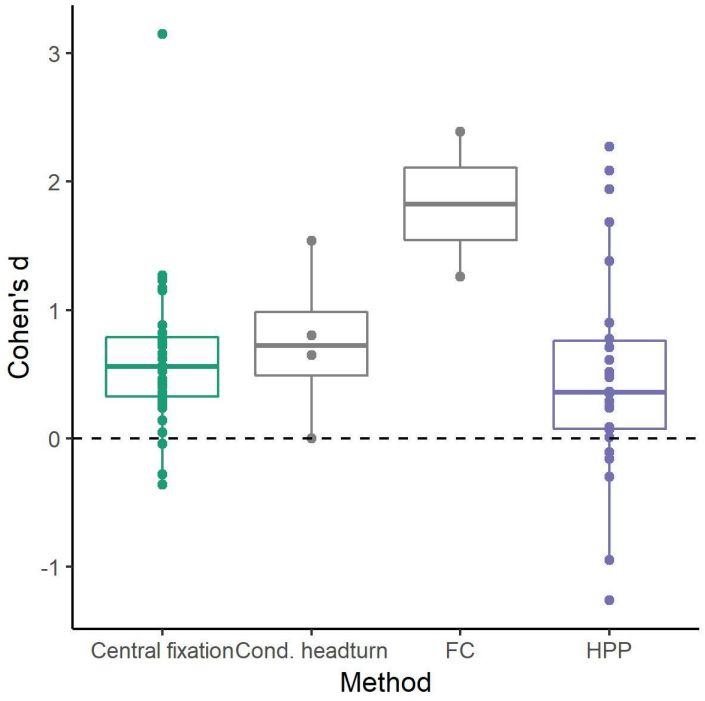


Predicting moderators: Method effect

Skeptic prior versus informed prior

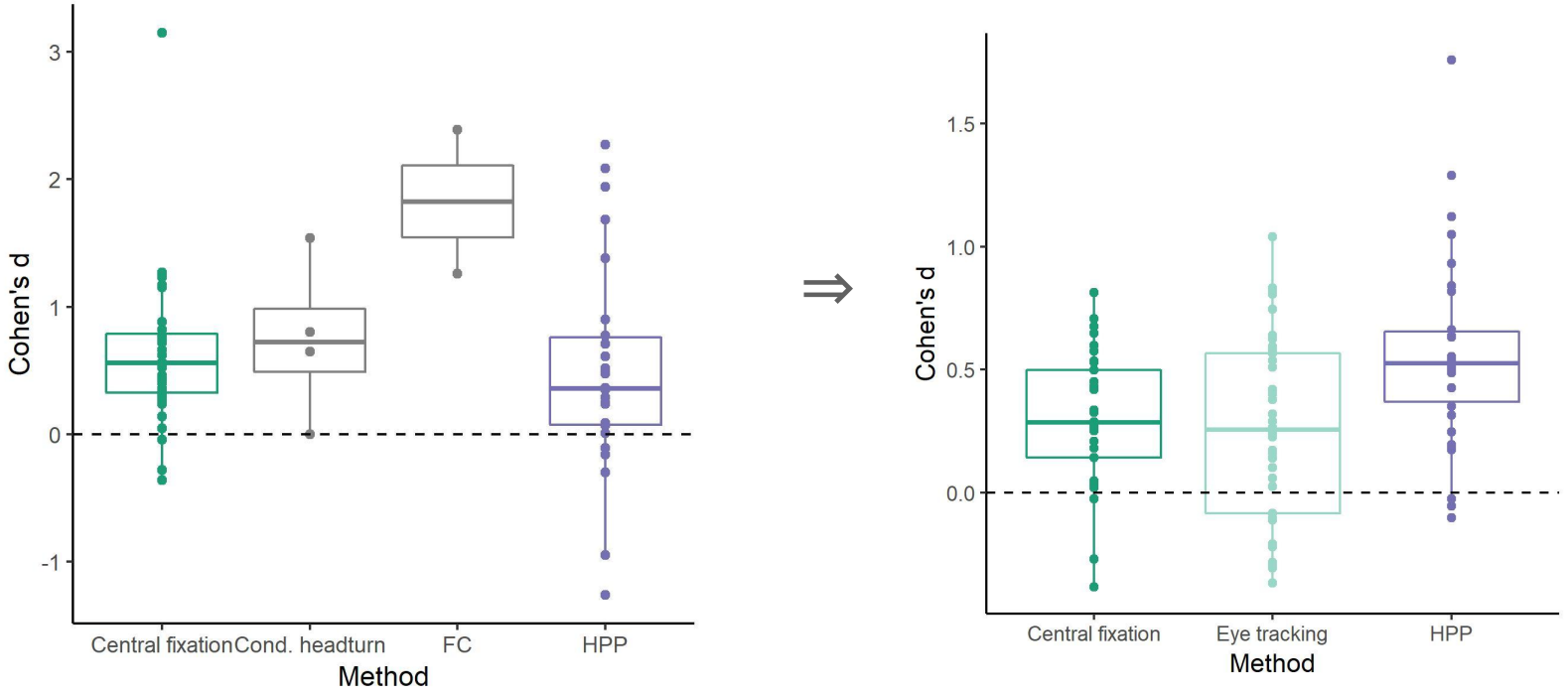
⇓

m = 0, sd = 0.3



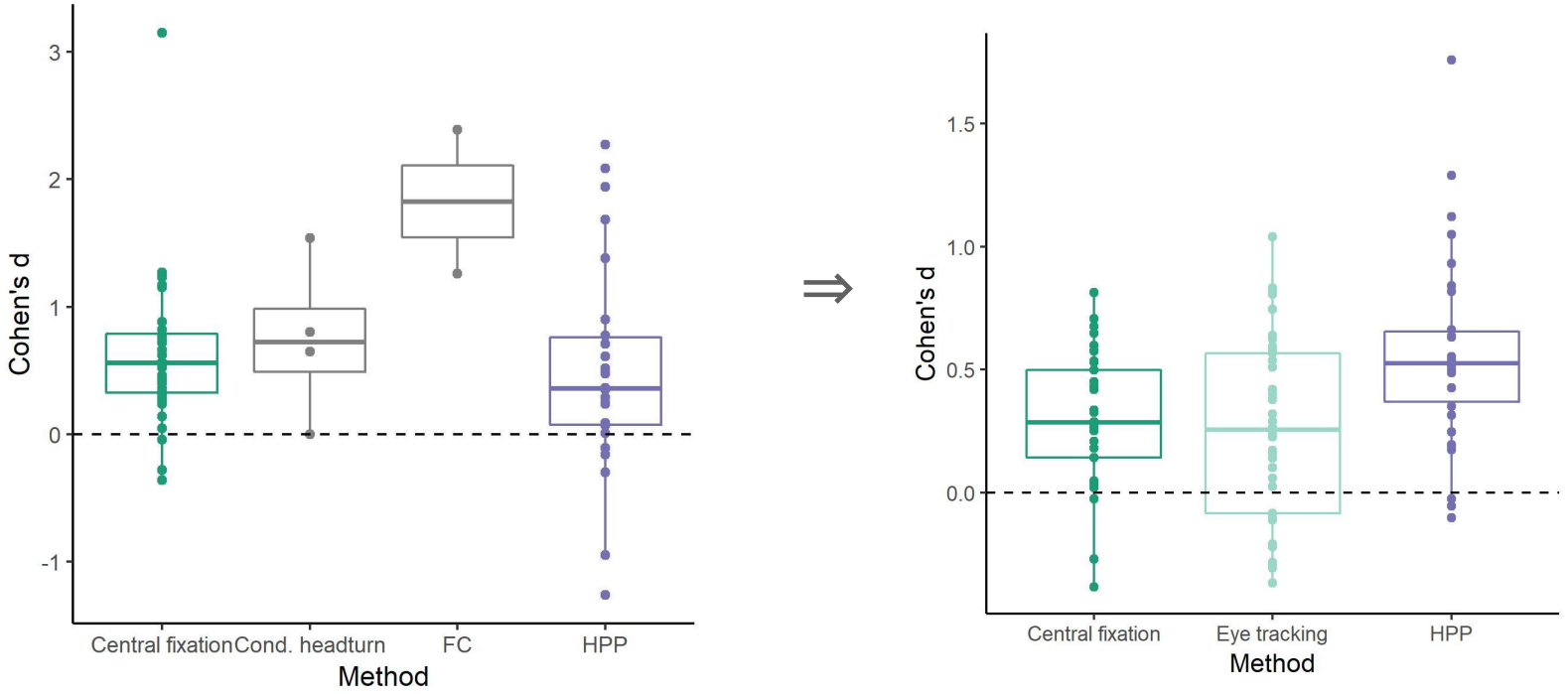
Predicting moderators: Method effect

Skeptic prior versus informed prior



Predicting moderators: Method effect

Skeptic prior versus **informed prior**



Summary: CAMAs and ManyBabies

Experiment planning:

- ➔ Design
- ➔ Stimuli

Summary: CAMAs and ManyBabies

Experiment planning:

➔ Design
➔ Stimuli

Results:

➔ Meta-analysis informative for moderators

Summary: CAMAs and ManyBabies

Experiment planning:

- ➔ Design
- ➔ Stimuli

Results:

- ➔ Meta-analysis informative for moderators

Next steps:

- ➔ Update meta-analysis (in progress)
- ➔ Further dig into possible joint analyses
  - ◆ Open for participation!

Theory adjudication though meta-analysis

Theories of language acquisition

Key characteristics

- Age as driving factor
- Universal stages

Theories of language acquisition

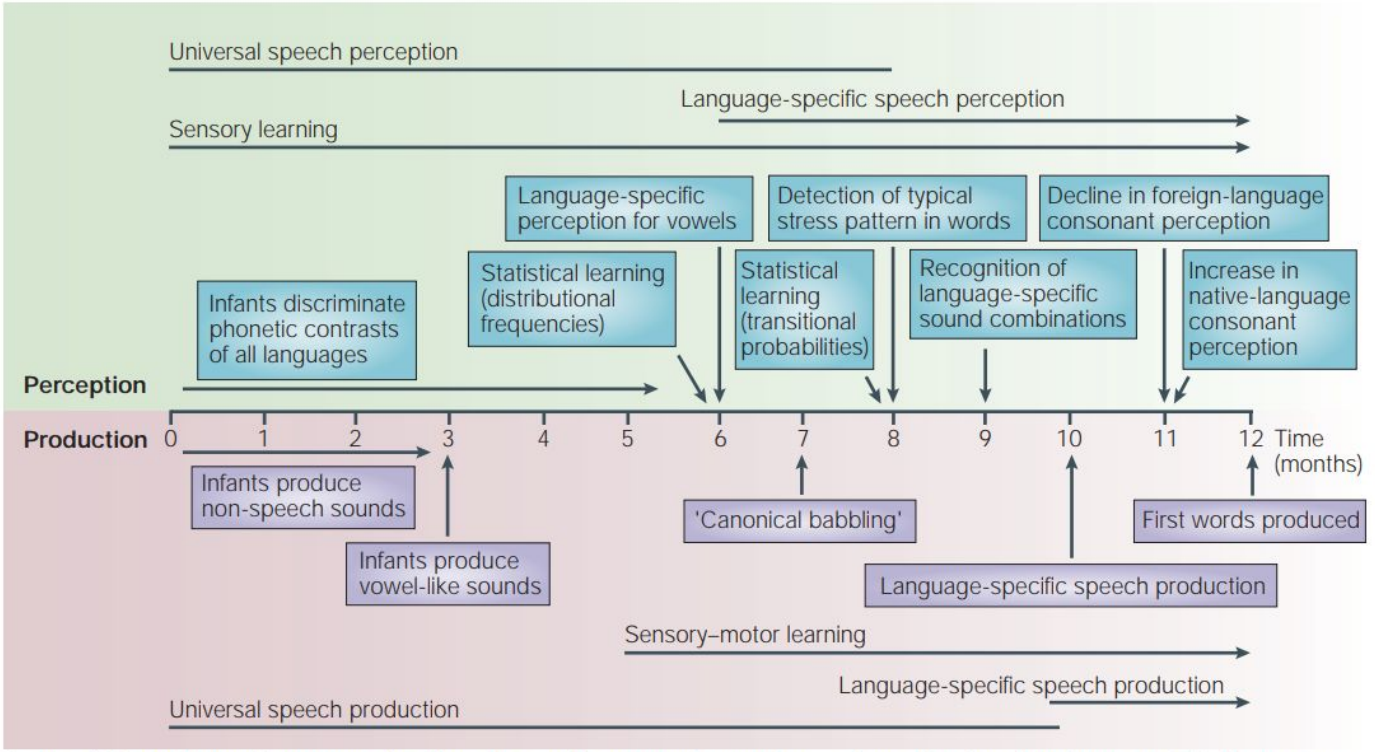
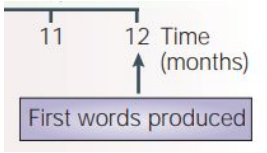


Figure 1 | **The universal language timeline of speech-perception and speech-production development.** This figure shows the changes that occur in speech perception and production in typically developing human infants during their first year of life.

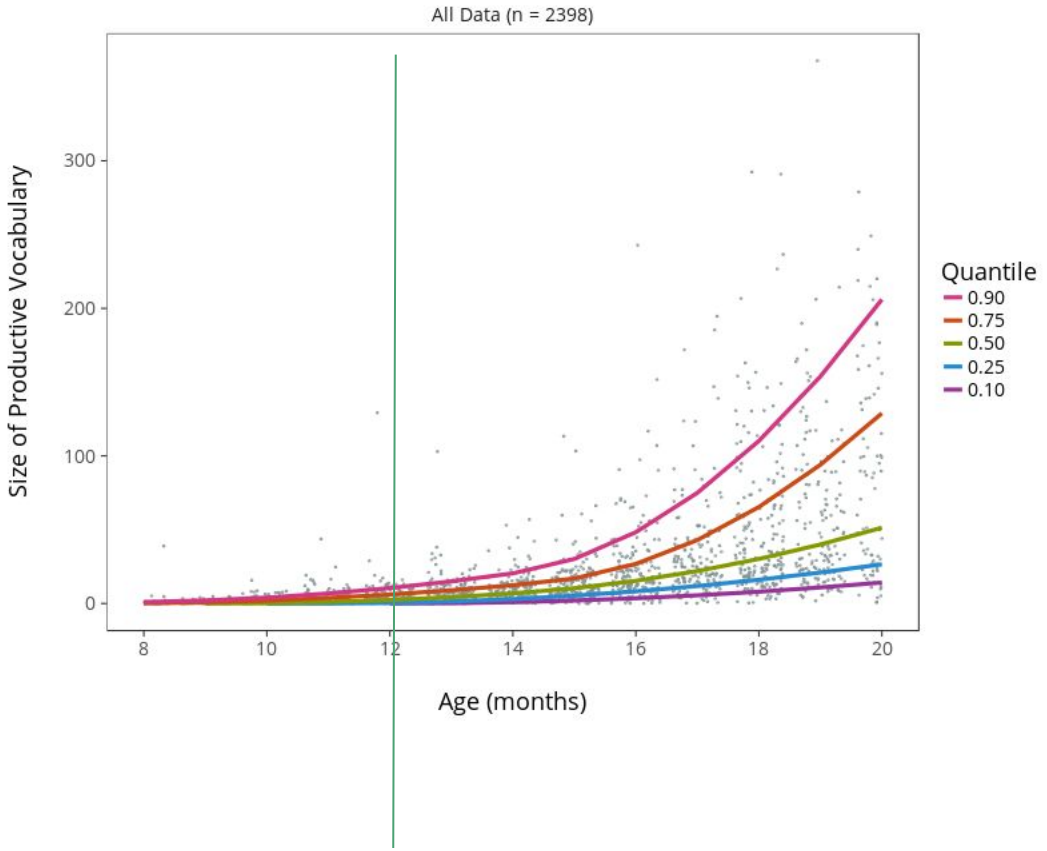
Kuhl (2004). https://www.nature.com/articles/nrn1533

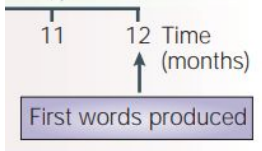
Theory vs Reality



Kuhl (2004). https://www.nature.com/articles/nrn1533

Theory vs Reality

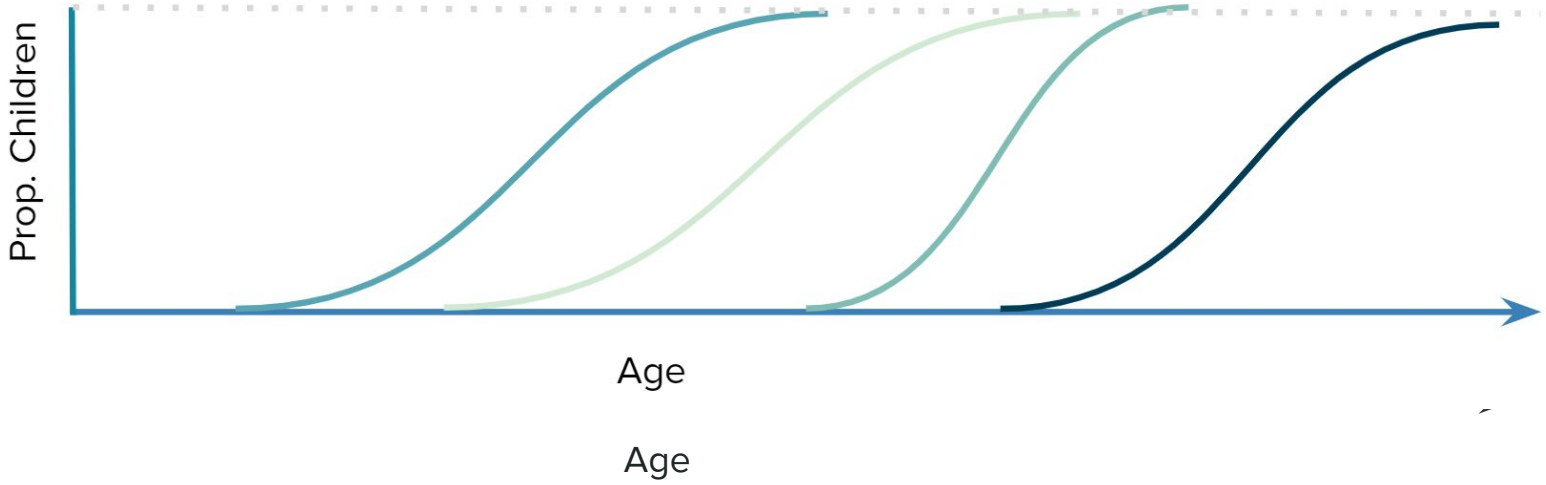




wordbank.stanford.edu, Danish data

Kuhl (2004). https://www.nature.com/articles/nrn1533

A different way of conceptualizing



Age

From point-estimates towards model comparison

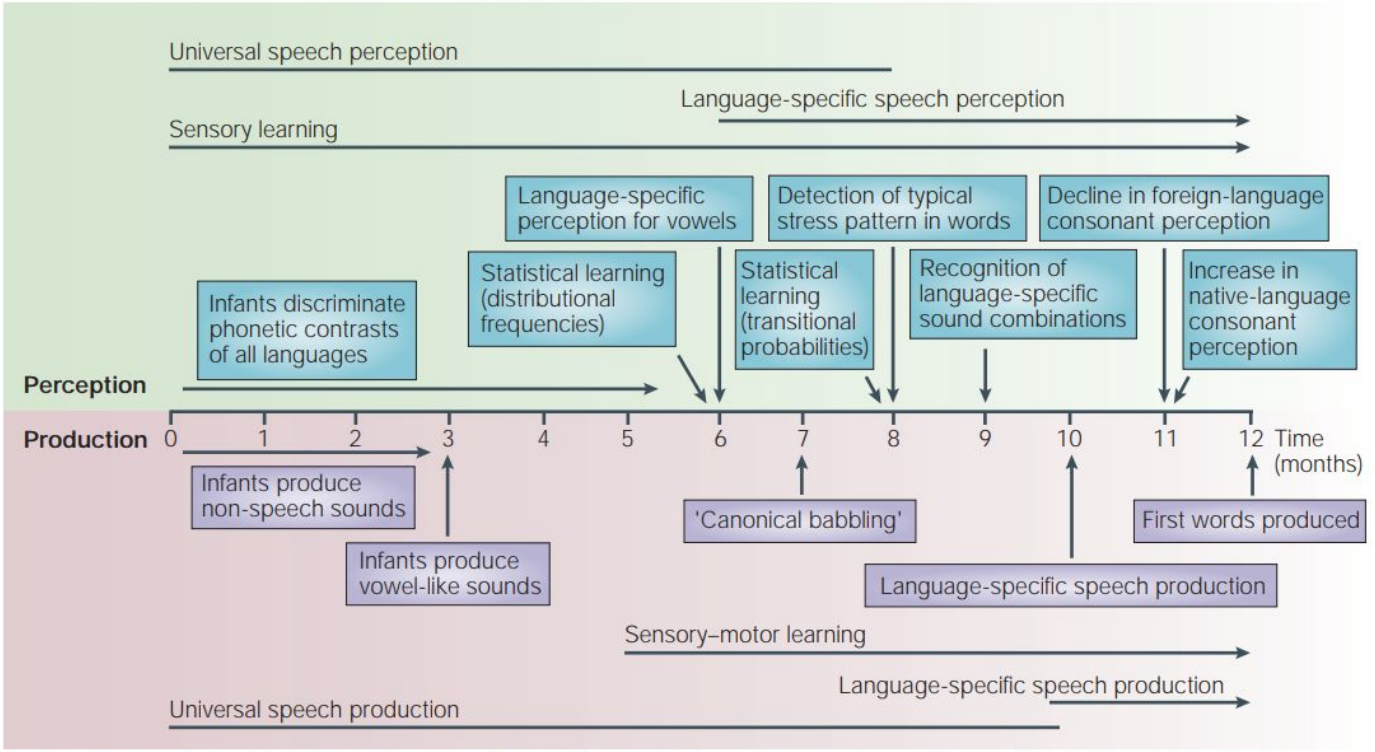
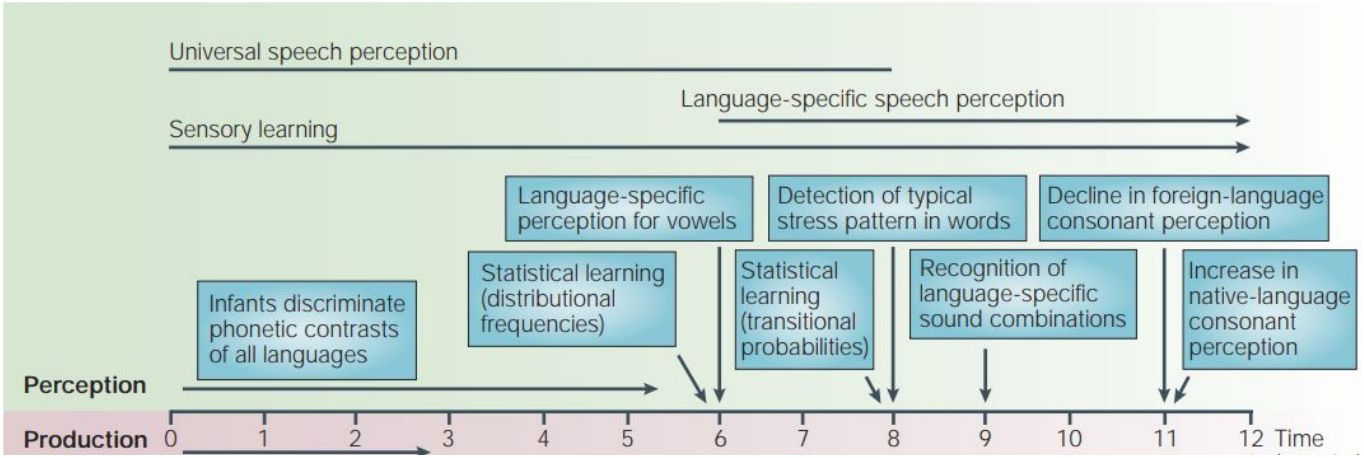


Figure 1 | **The universal language timeline of speech-perception and speech-production development.** This figure shows the changes that occur in speech perception and production in typically developing human infants during their first year of life.

Kuhl (2004). https://www.nature.com/articles/nrn1533

From point-estimates towards model comparison



Kuhl (2004). https://www.nature.com/articles/nrn1533

From point-estimates towards model comparison

Required:

- Clear (verbal) model predictions
- Various skills on a common scale

Example: Sounds and words

2 different levels of processing

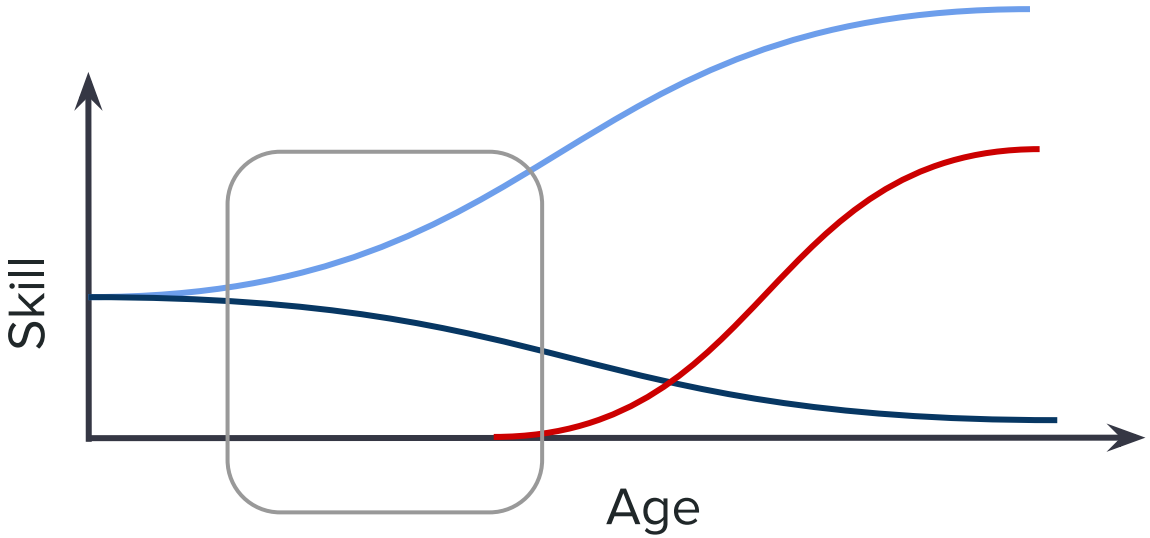
But: Knowing one helps with the other

Example: Sounds and words

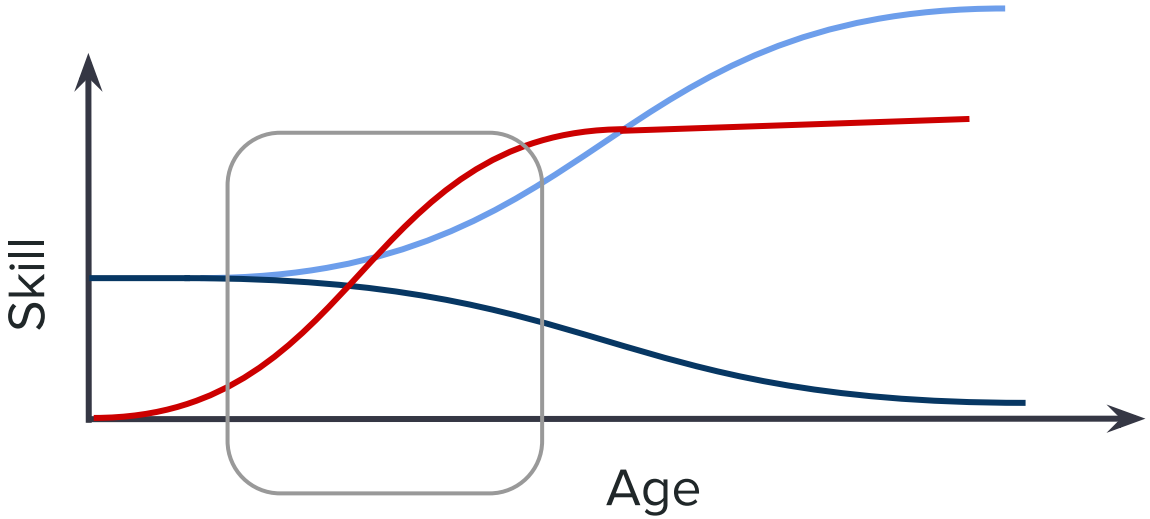
2 different levels of processing

But: Knowing one helps with the other - **Which comes first?**

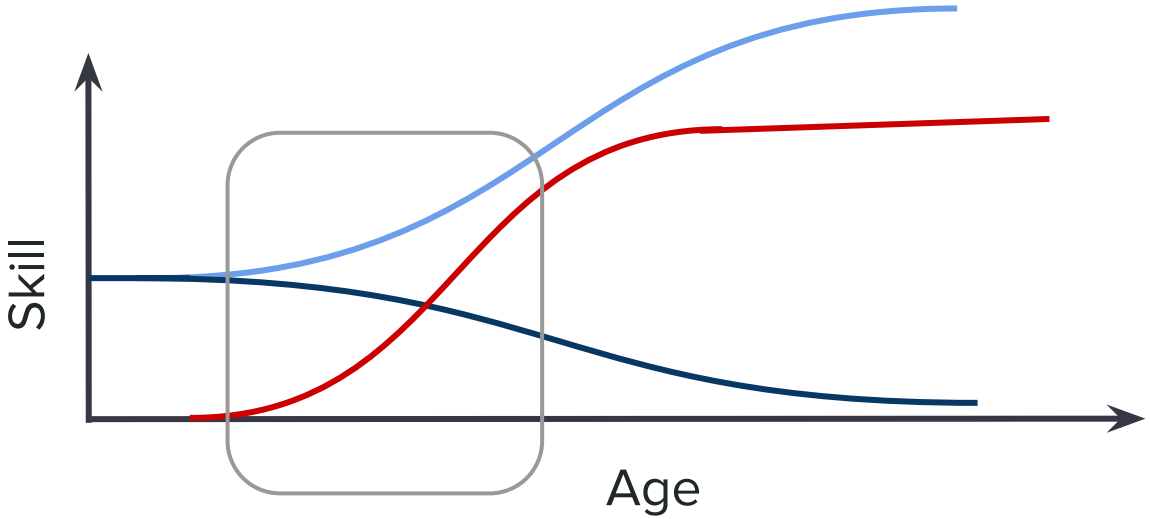
Theory 1: Sounds before words



Theory 2: Words before sounds



Theory 3: Parallel development



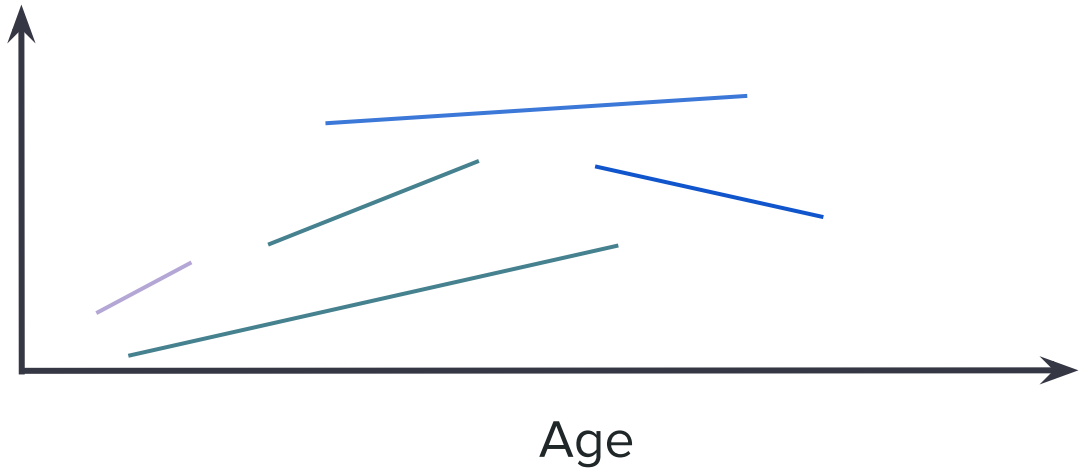
Preprocessing

Data quality check: Is the expected trajectory present?

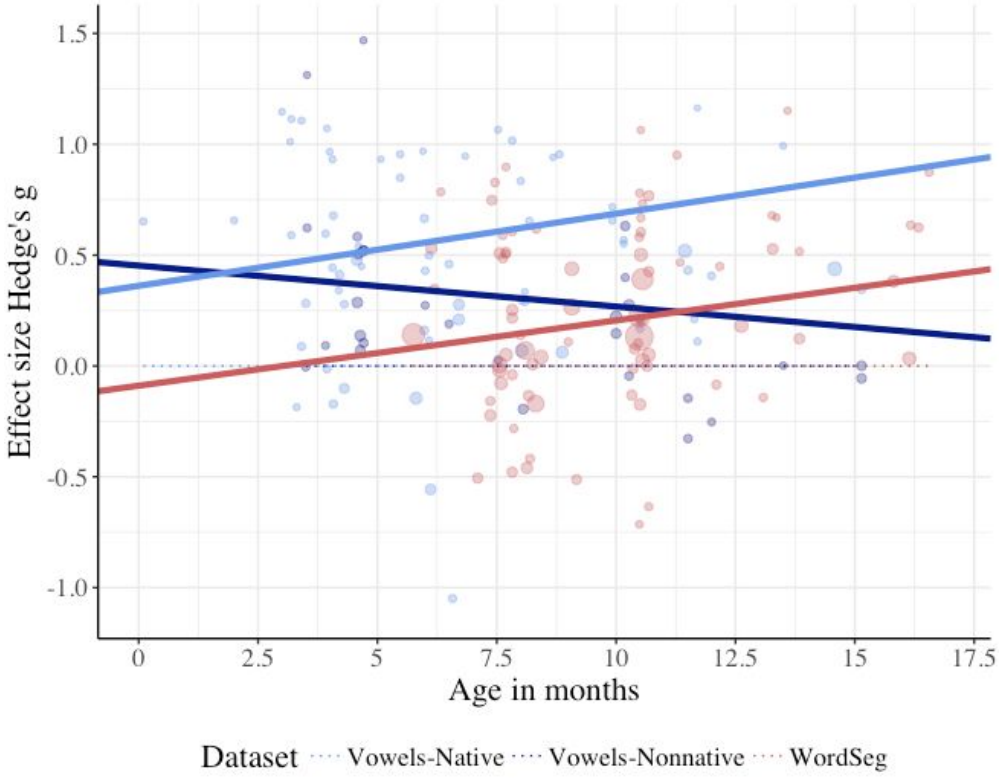
Preprocessing

Data quality check: Is the expected trajectory present?

➔ Subset to studies with multiple age groups to control variation

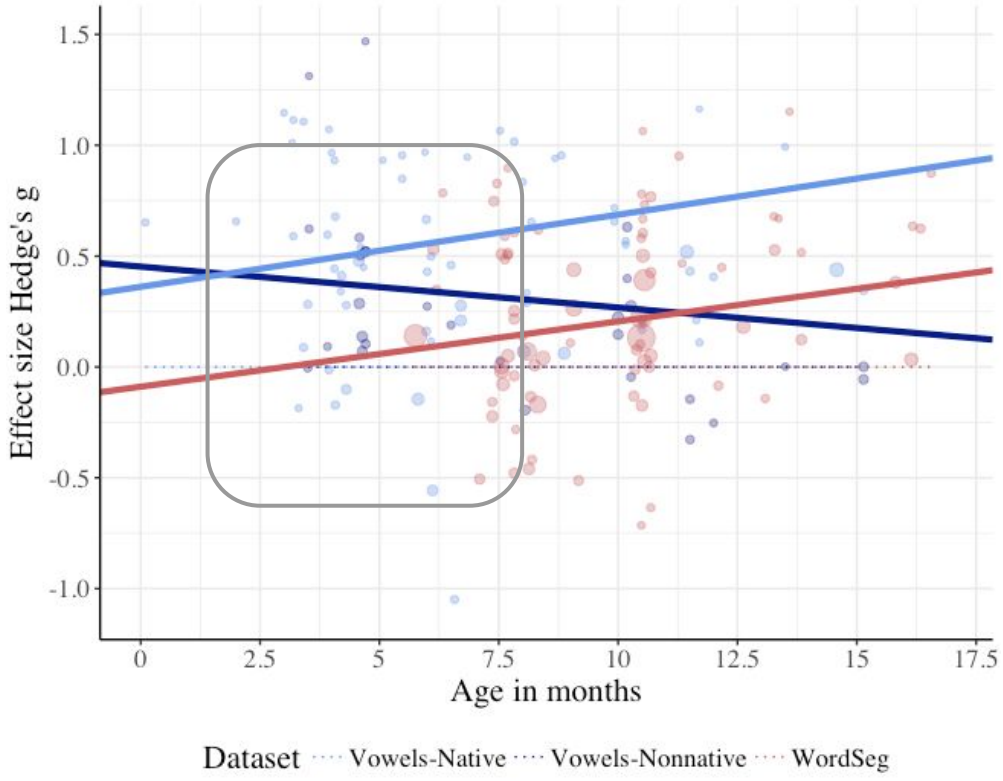


Qualitative comparison



Bergmann, Cristia, & Tsuji (in prep/2017)

Qualitative comparison



Support for Theory 3: parallel development

➜ No clear developmental order

Bergmann, Cristia, & Tsuji (in prep/2017)

Open questions

Divergence versus emergence?

Can we really compare studies that different?

Open questions

Divergence versus emergence?

Can we really compare studies that different?

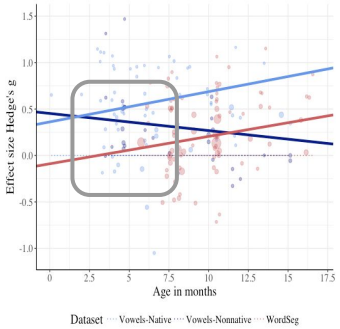
Replication distance and method effects as "noise"?

...or meaningful task effects?

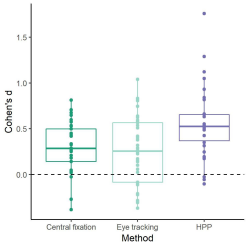
➜ Prerequisite for useful model comparison

Summary

Why CAMAs?

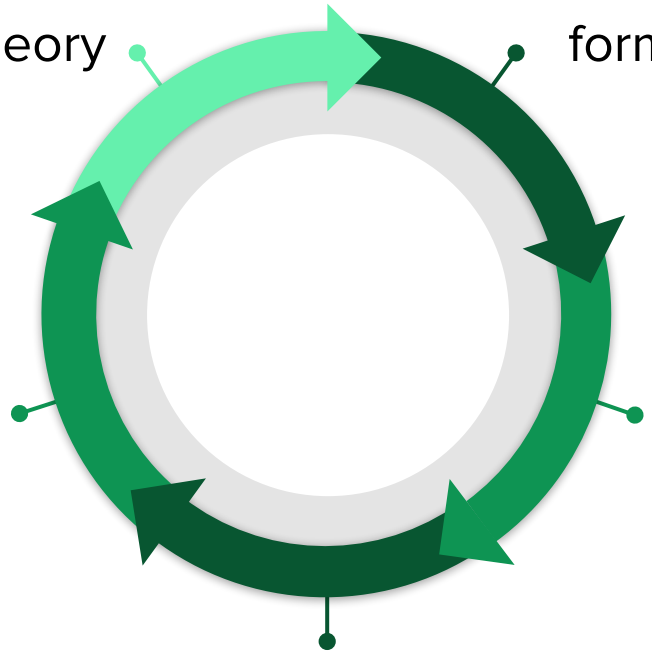


study_ID	long_cite	short_cite	doi	peer_reviewed
Cooper1990	Cooper, R. P. & Aslin, N. R.	Cooper & Aslin (1990)	10.1111/j.1467-8624.1990.tb02	yes
Cooper1990	Cooper, R. P. & Aslin, N. R.	Cooper & Aslin (1990)	10.1111/j.1467-8624.1990.tb02	yes
Cooper1994	Cooper, R. P. & Aslin, N. R.	Cooper & Aslin (1994)	10.1111/j.1467-8624.1994.tb00	yes
Cooper1994	Cooper, R. P. & Aslin, N. R.	Cooper & Aslin (1994)	10.1111/j.1467-8624.1994.tb00	yes
Cooper1994	Cooper, R. P. & Aslin, N. R.	**Cooper & Aslin (1994)**	10.1111/j.1467-8624.1994.tb00	yes
Cooper1994	Cooper, R. P. & Aslin, N. R.	**Cooper & Aslin (1994)**	10.1111/j.1467-8624.1994.tb00	yes
Cooper1997	Cooper, R. P.; Abraham, J.;	Cooper et al. (1997)	10.1016/S0163-6383(97)90037-0	yes
Cooper1997	Cooper, R. P.; Abraham, J.;	Cooper et al. (1997)	10.1016/S0163-6383(97)90037-0	yes
Cooper1997	Cooper, R. P.; Abraham, J.;	Cooper et al. (1997)	10.1016/S0163-6383(97)90037-0	yes



Update theory

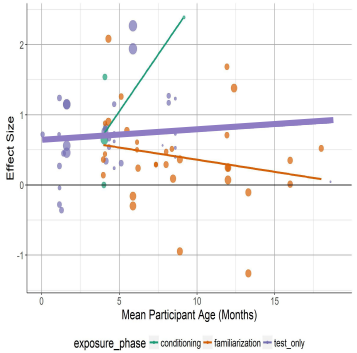
Hypothesis formulation



Analysis and interpretation

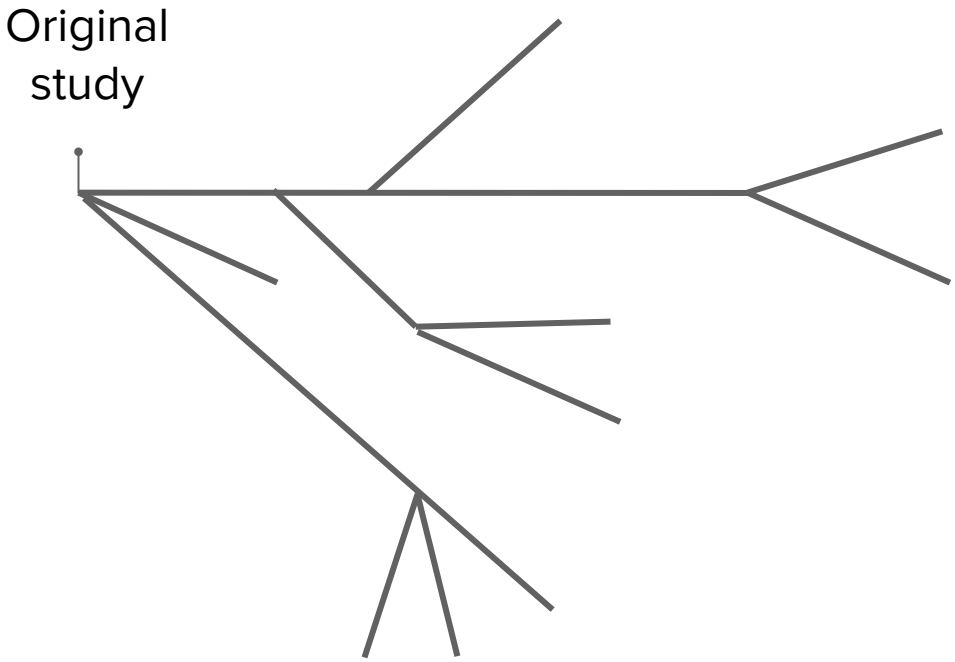
Experiment design

Data collection



CAMAs to understand replication in meta-analyses

Rich, dynamic data

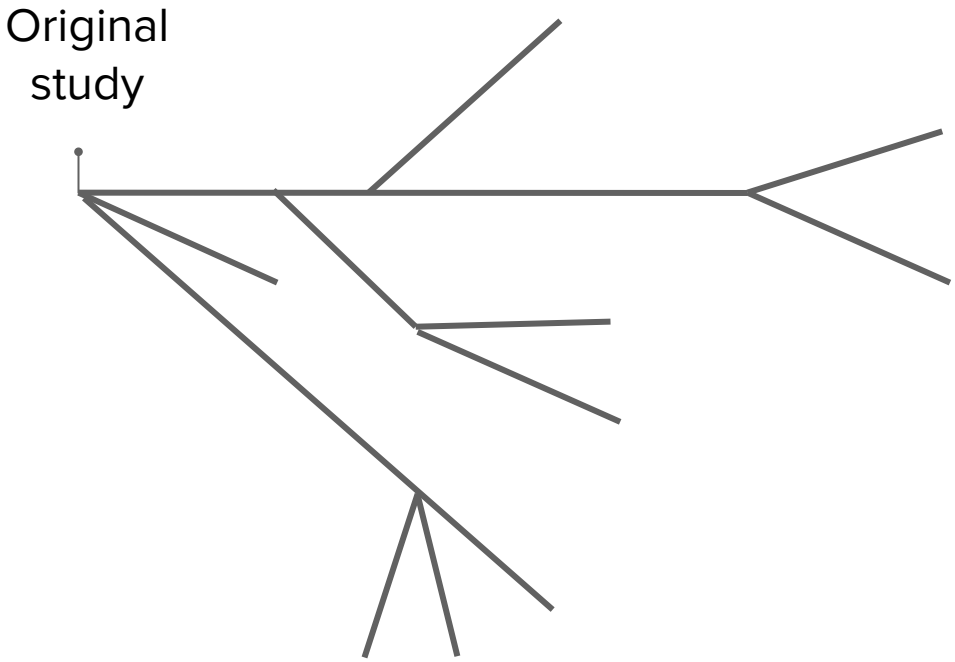


➜ Determine effects of study differences / commonalities on

replication distance

CAMAs to understand replication in meta-analyses

Rich, dynamic data



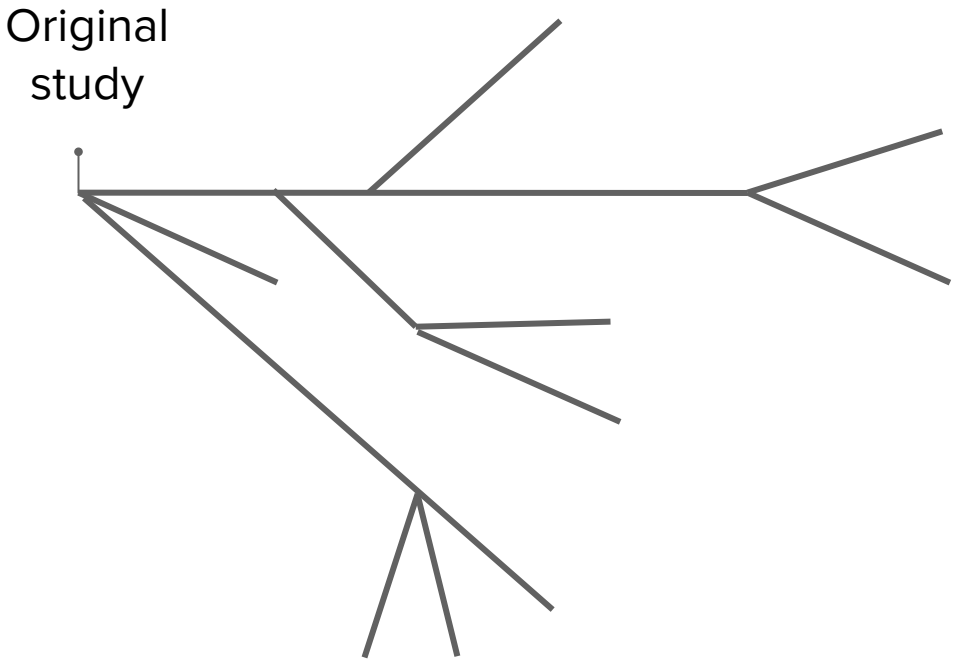
➜ Determine effects of study differences / commonalities on

replication distance

Leverage meaningful variation, control "noise"

CAMAs to understand replication in meta-analyses

Improve our models (statistical and cognitive)



Future directions

Integrate participant-level data

➔ Multi-level approach experts welcome

Future directions

Integrate participant-level data

- ➜ Multi-level approach experts welcome

Improve educational tools

- ➜ Further increase engagement (see also yesterday's talk)

Future directions

Integrate participant-level data

➔ Multi-level approach experts welcome

Improve educational tools

➔ Further increase engagement (see also yesterday's talk)

Facilitate data upload

➔ Automatic integration of new meta-analyses

Thank you