

Meta Analysis

Psych 251

11/8/17

Hotel towel reuse

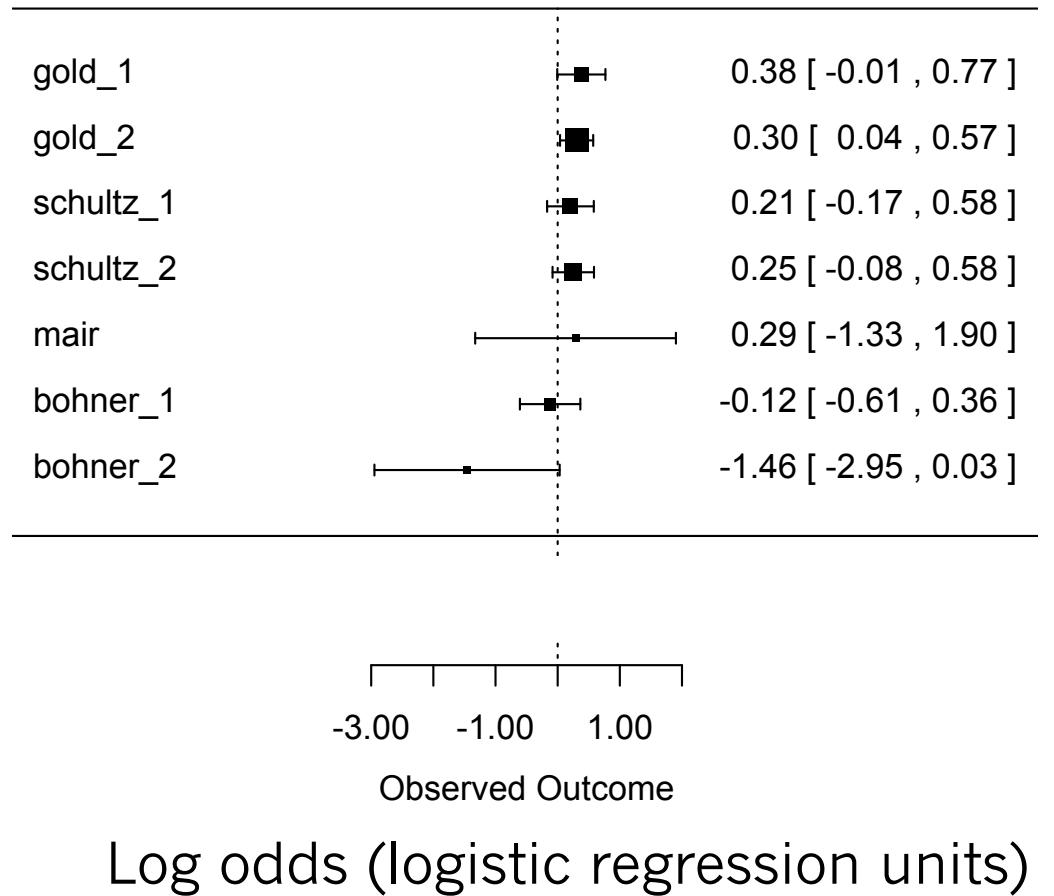
- How to get guests to save the earth?
- Messages
 - benefits of environmental protection (control)
 - majority of guests actually reused their towels in the past (descriptive-social-norm)
- Data:
 - Experiment 1: $p = .05$
 - norm: 44.1%
 - control: 35.1%
 - Experiment 2: $p = .03$
 - norm: 44.5%
 - control: 37.2%



Goldstein, Cialdini, & Griskevicius (2008)

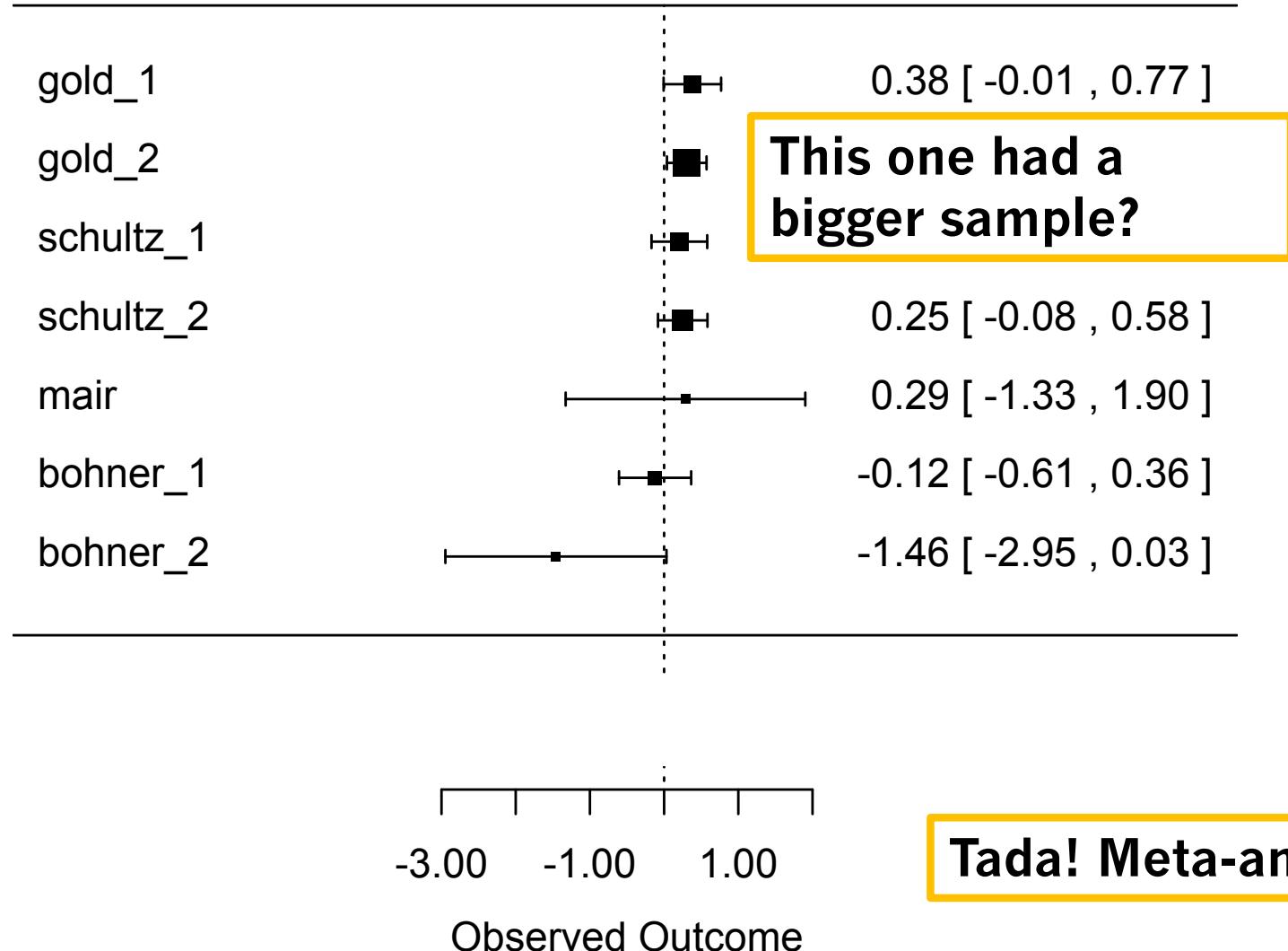
But oh the replications...

- Five replication experiments that assessed the proportion of hotel guests who reused their towels
- N total = 2,466
- All five experiments arguably failed to replicate the original finding (all $p > .14$)



How do we aggregate this evidence?

5 / 7 ain't bad!



Components of MA

- Sample of papers
 - From a literature review
 - Issues of inclusion/exclusion/coding
- Effect size
 - Scale-free measure, with common units
 - d (standard deviations), r , or log odds
- Model
 - Fixed effects
 - Random effects
 - Multi-level

Outline

- Why meta-analysis?
 - History
 - Use cases: summary MA, internal MA
- Technical details
 - Coding issues
 - Effect sizes
 - Models: fixed, random
 - Funnel plots and publication bias
- MetaLab: A fun meta-research project

Outline

- **Why meta-analysis?**
 - History
 - Use cases: summary MA, internal MA
- Technical details
 - Coding issues
 - Effect sizes
 - Models: fixed, random
 - Funnel plots and publication bias
- MetaLab: A fun meta-research project

Originally: The Great Debate

- 1952: Hans J. Eysenck concluded that there were no favorable effects of psychotherapy, starting a raging debate
- 20 years of evaluation research and hundreds of studies failed to resolve the debate
- 1978: To proved Eysenck wrong, Gene V. Glass statistically aggregate the findings of 375 psychotherapy outcome studies
- Glass (and colleague Smith) concluded that psychotherapy did indeed work
- Glass called his method “meta-analysis”

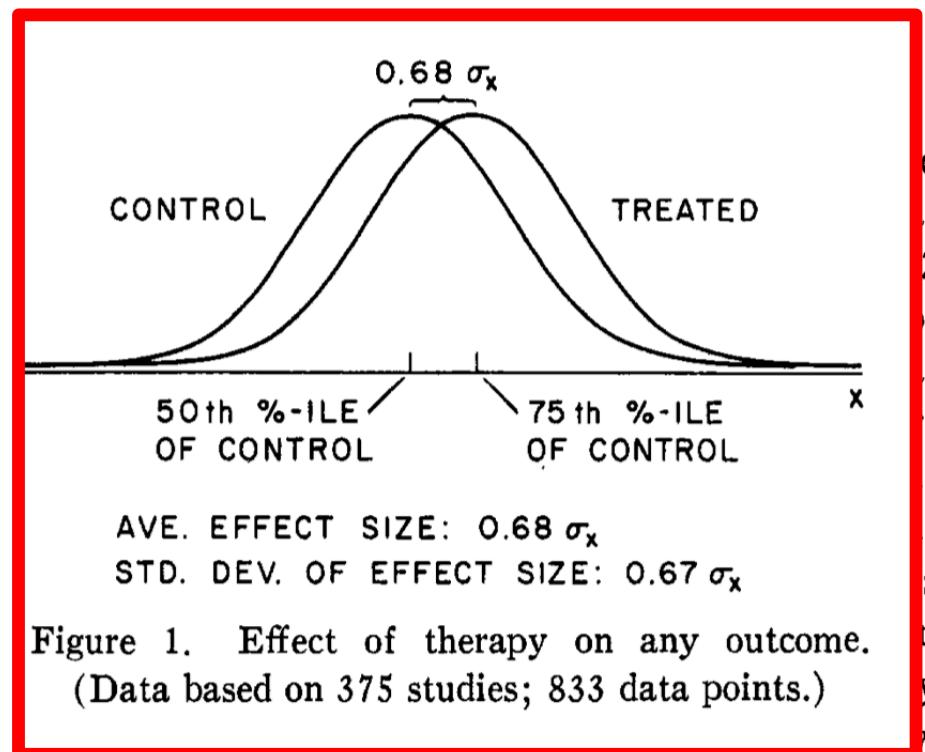
The original paper

Meta-Analysis of Psychotherapy Outcome Studies

MARY LEE SMITH
GENE V GLASS

University of Colorado—Boulder
University of Colorado—Boulder

ABSTRACT: Results of nearly 400 controlled evaluations of psychotherapy and counseling were coded and integrated statistically. The findings provide convincing evidence of the efficacy of psychotherapy. On the average, the typical therapy client is better off than 75% of untreated individuals. Few important differences in effectiveness could be established among many quite different types of psychotherapy. More generally, virtually no difference in effectiveness was observed between the class of all behavioral therapies (systematic desensitization, behavior modification) and the nonbehavioral therapies (Rogerian, psychodynamic, rational-emotive, transactional analysis, etc.).

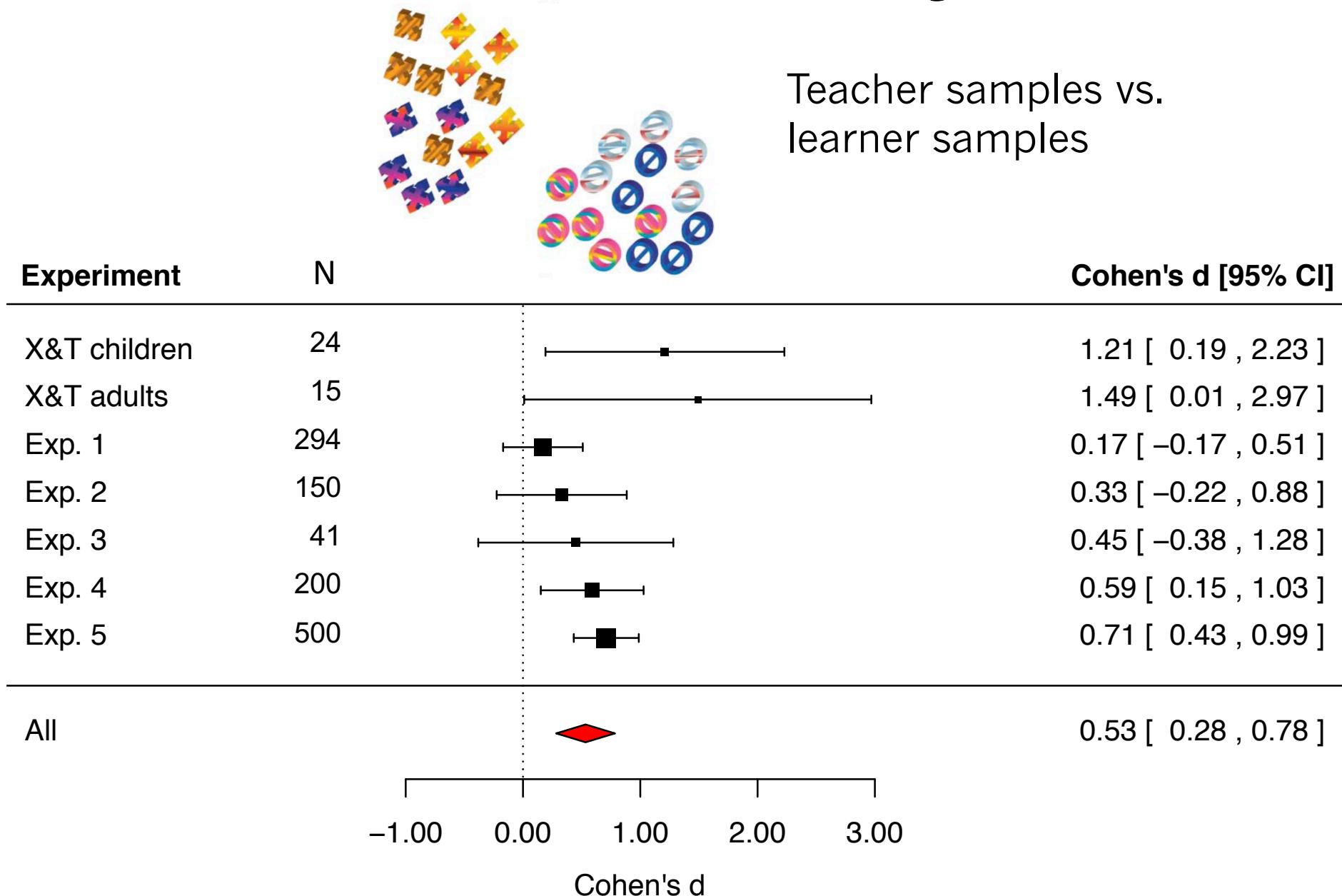


representative the 40 are of the 400 is unknown.

History

- Ideas behind meta-analysis predate Glass' work by several decades
 - Karl Pearson (1904): averaged correlations for studies of the effectiveness of inoculation for typhoid fever
 - R. A. Fisher (1944): “When a number of quite independent tests of significance have been made, it sometimes happens that although few or none can be claimed individually as significant, yet the aggregate gives an impression that the probabilities are on the whole lower than would often have been obtained by chance” (p. 99).
 - W. G. Cochran (1953): Discusses a method of averaging means across independent studies

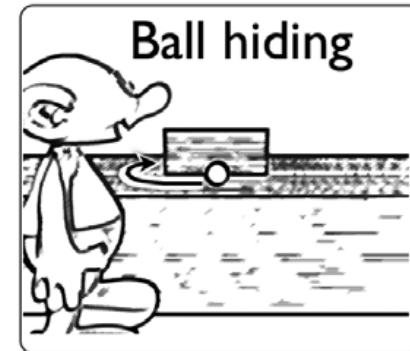
Internal meta-analysis 1



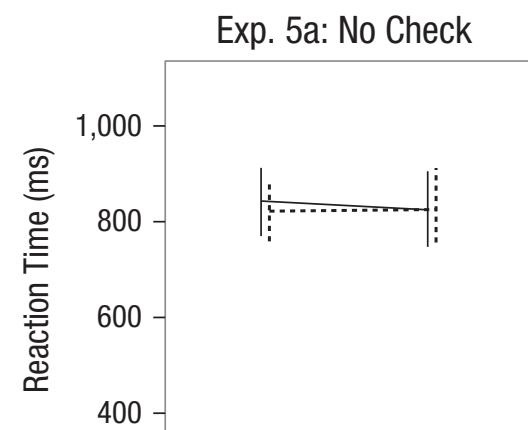
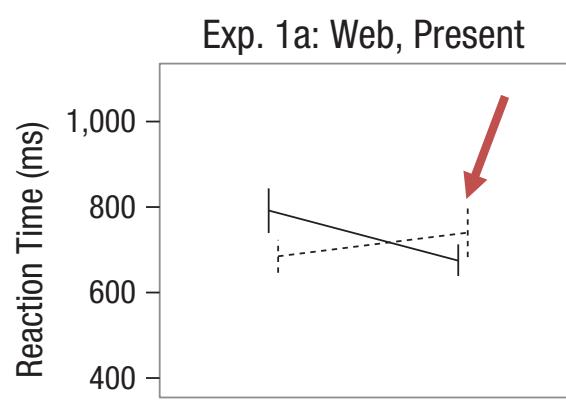
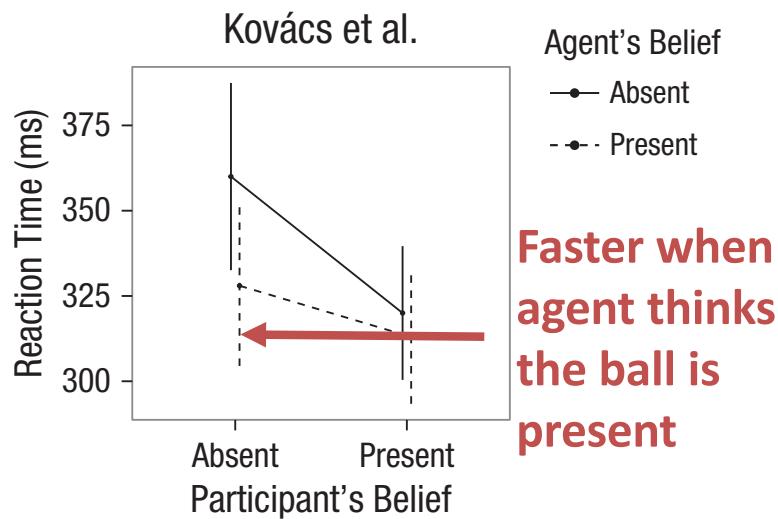
Xu & Tenenbaum (2007b), *Dev Sci*; Lewis & Frank (2016), *JEP:G*

Internal meta-analysis 2

The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults

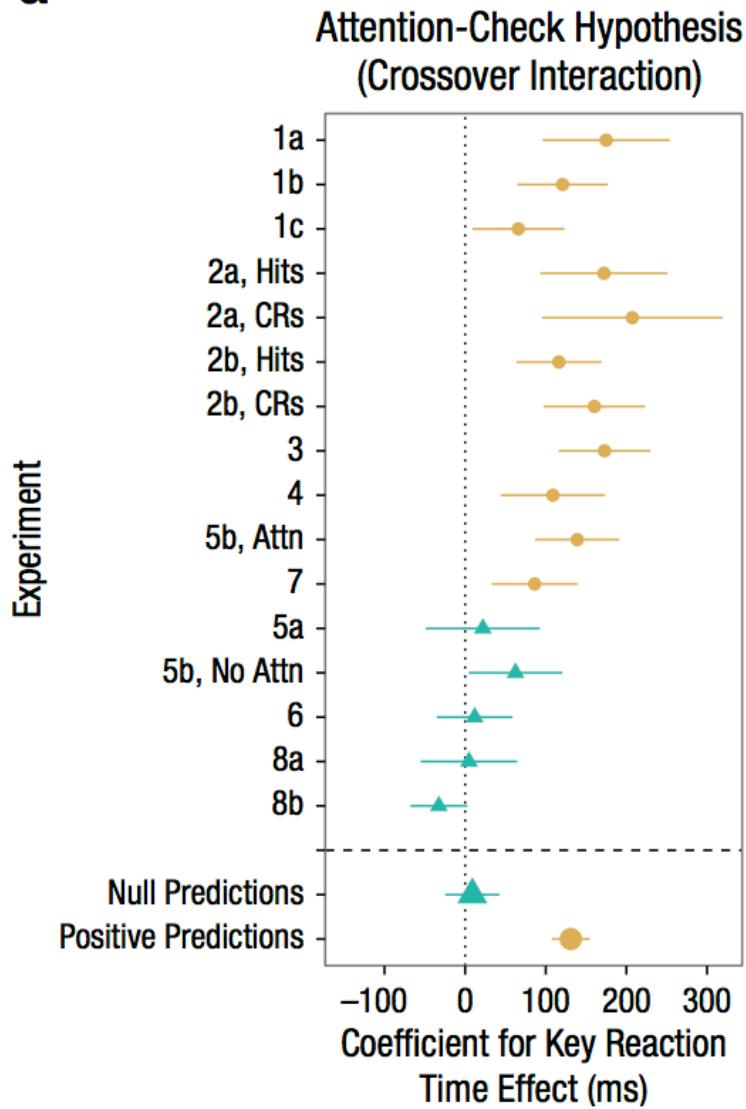


Ágnes Melinda Kovács,^{1,2,3*} Ernő Téglás,^{1,2,3} Ansgar Denis Endress^{3,4}

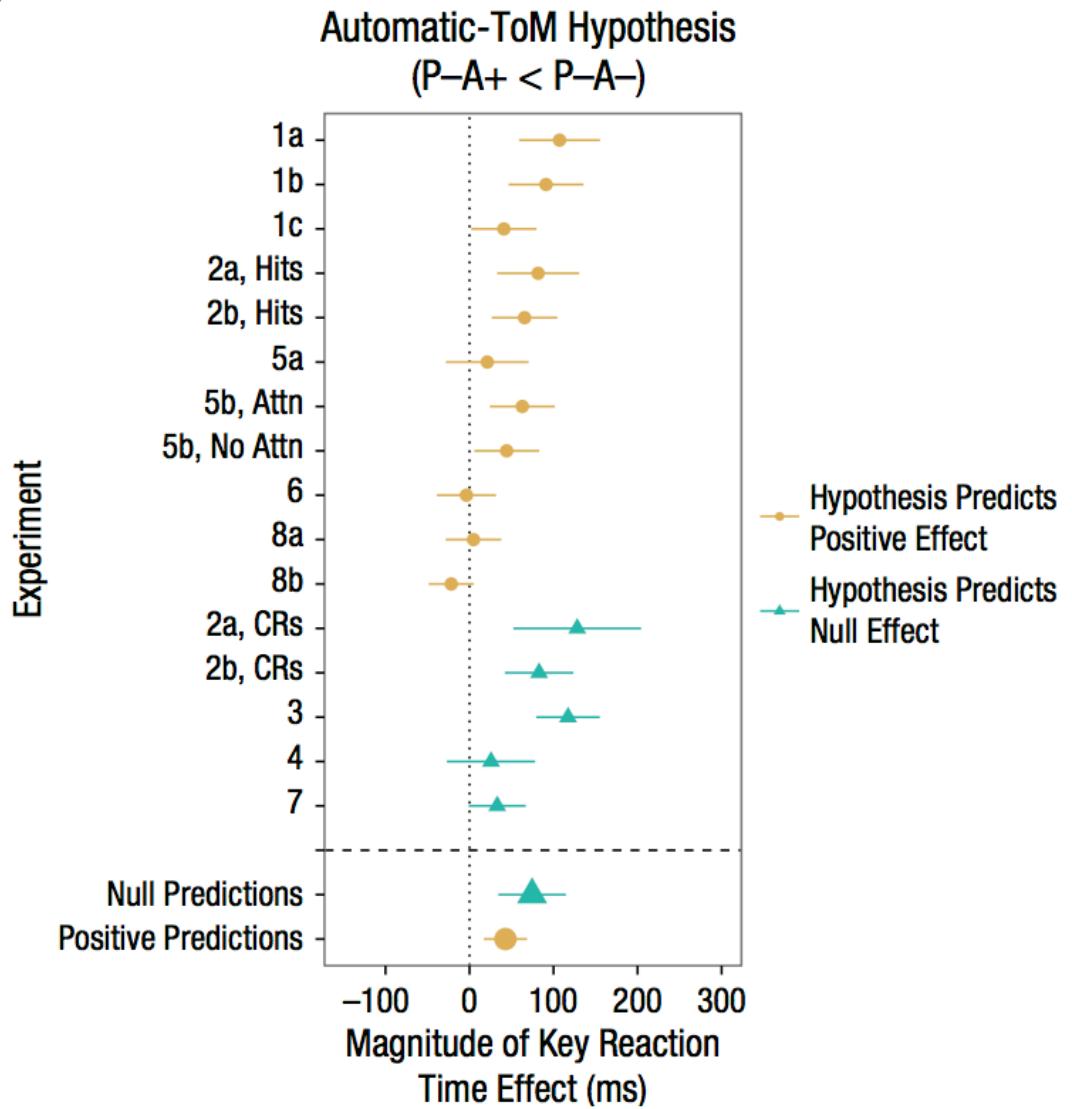


Internal meta-analysis 2

a



b



Outline

- Why meta-analysis?
 - History
 - Use cases: summary MA, internal MA
- **Technical details**
 - Coding issues
 - Effect sizes
 - Models: fixed, random
 - Funnel plots and publication bias
- MetaLab: A fun meta-research project

Structure of a meta-analysis

- Build initial study list
 - E.g. via systematic literature search
- Screen papers according to objective criteria
- Decide on data structure and create template
- Retrieve paper, identify and calculate effect size
 - This is where it gets tricky...

Effect sizes

- Cohen's d common for experimental interventions
 - Think “change in IQ due to training”
- Correlation (r) common for observational studies
 - Think “association between conscientiousness and grades”
- Log odds common for binary outcomes (logistic regression)
 - Think “increase in risk of death”
- Package `compute.es` for R will convert between these

Coding individual studies

- What's the key test?
- Do they have means and SD?
- No? How about a t-value?
- Or a regression coefficient?

Step 1: Estimate Effect Size

This is the most important bit. So let's go thorough it step by step.

This is a repeated measure, akin to a pre-post contrast therefore, we use the *standardized mean gain* set of formulas whenever possible

```
db$LTDif <- ifelse(db$formula2use == "means&SD", db$LTfam - db$LTNov, NA)
```

Compute pooled SD to use in Effect Size measures. Source: Lipsey & Wilson, page 44 - from Becker 1988

```
db$PooledSD <- ifelse(db$formula2use == "means&SD", sqrt((db$SD.LTfam^2 + db$SD.LTNov^2)/2), NA)
```

Compute Effect Size based on means and SD of dependent variables where possible.
Source: Lipsey & Wilson, formula 3.14

```
db$ES <- ifelse(db$formula2use == "means&SD", db$LTDif/db$PooledSD, NA)
```

This is an approximation of the same ES using exact t values We do NOT use formula 2/3 from Table B10 in Lipsey & Wilson because it is based on independent samples Instead, we follow Dunlap et al. 1996, page 171

```
db$ES <- ifelse(db$formula2use == "tValue", db$exact.t * sqrt(2 * (1 - db$correlationFamNov. as.numeric(as.character(db$ES))))
```

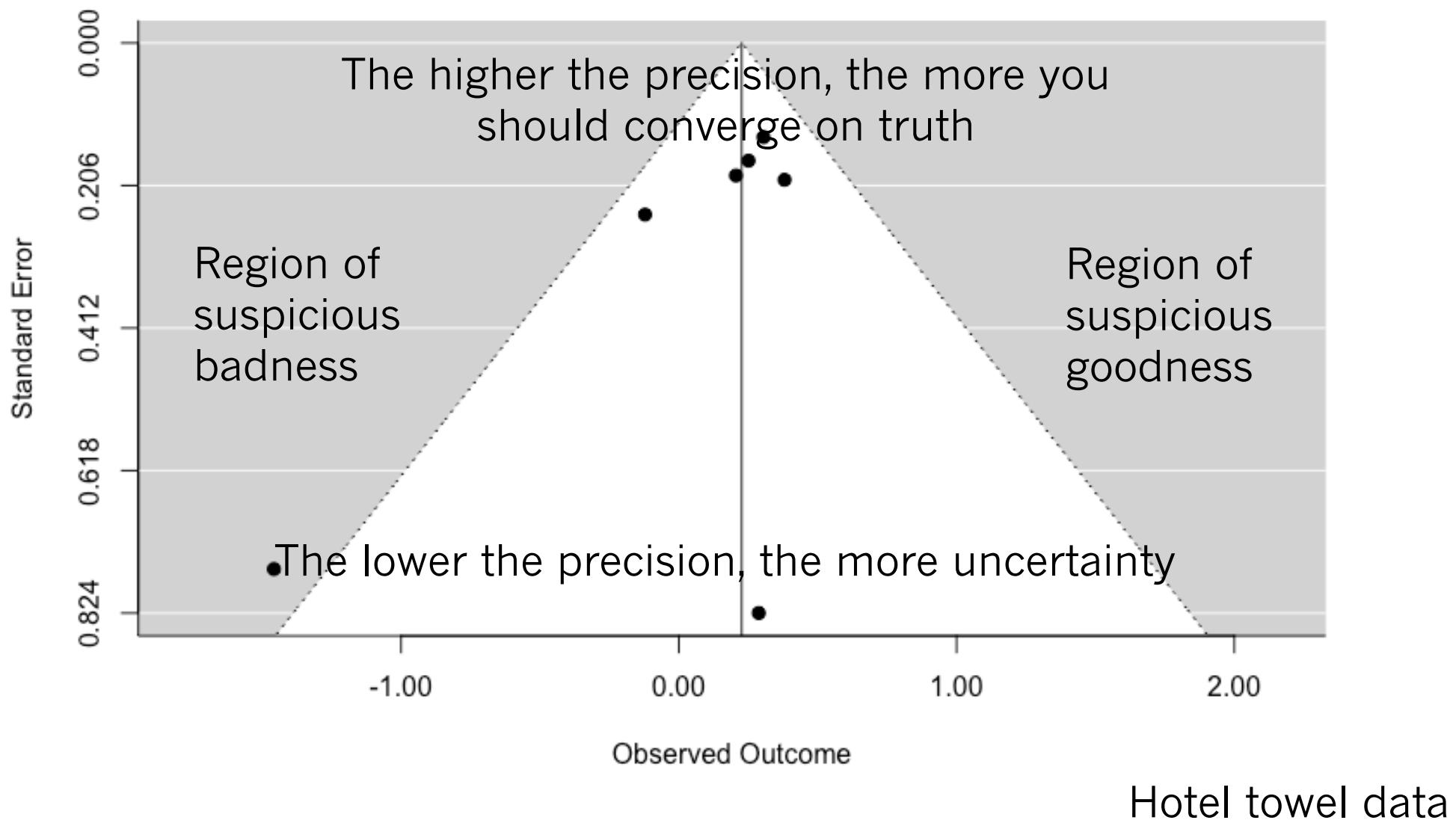
Coding issues

- Issues
 - How systematic is the sampling of studies?
 - How comparable are studies to one another?
- What goes into the MA dictates what comes out
 - Direct replications vs. widely disparate experiments
 - Biased sample vs. all available studies
- Include published work? Or search for unpublished findings too?
- **A priori specification of standards critical**
 - Even easier to p-hack a meta-analysis than an empirical study (samples small, details are detailed...)

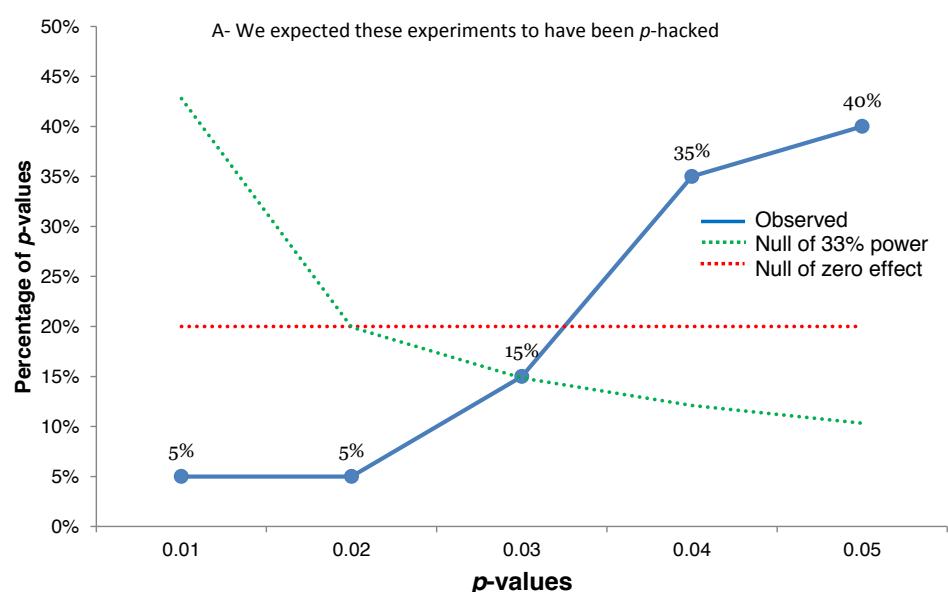
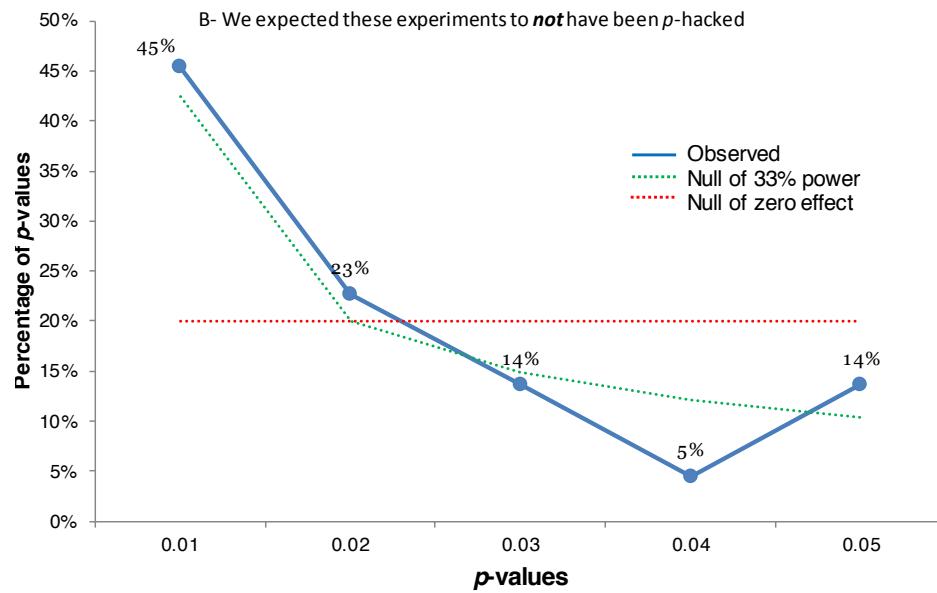
Publication bias

Tendency for researchers to publish positive findings and “file-drawer” the negative ones.

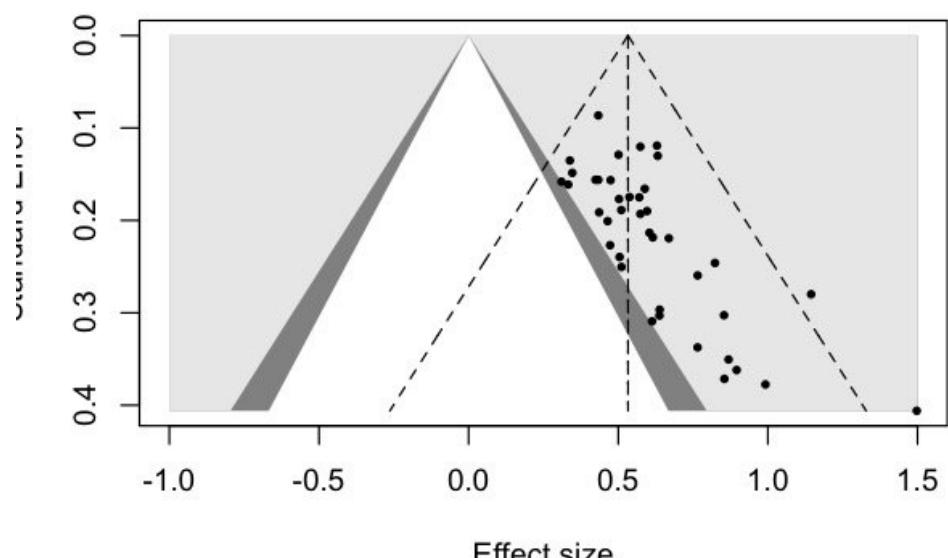
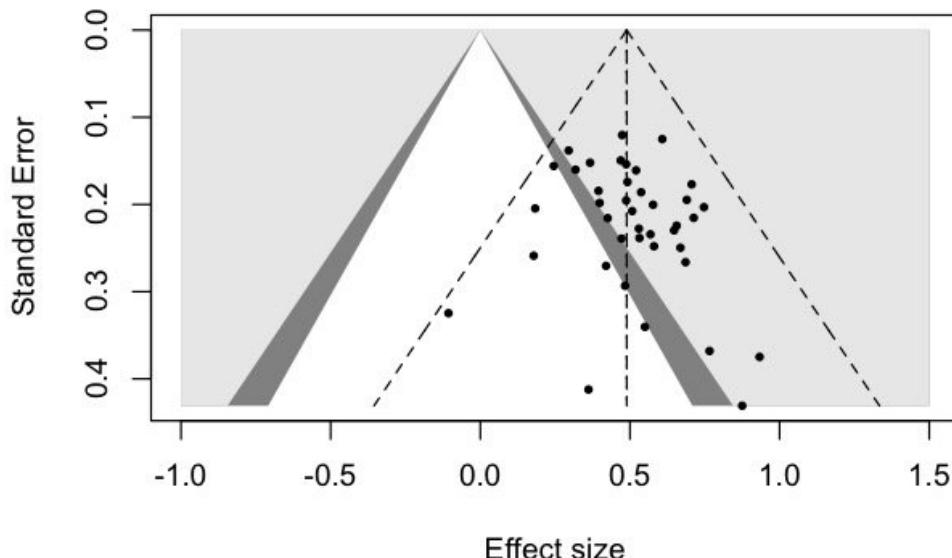
Related to (perhaps hard to distinguish from) p-hacking...



Forensics for bias P-curve



Funnel plots



Light & Pillemer (1984); Simonsohn, Nelson, & Simmons (2014)

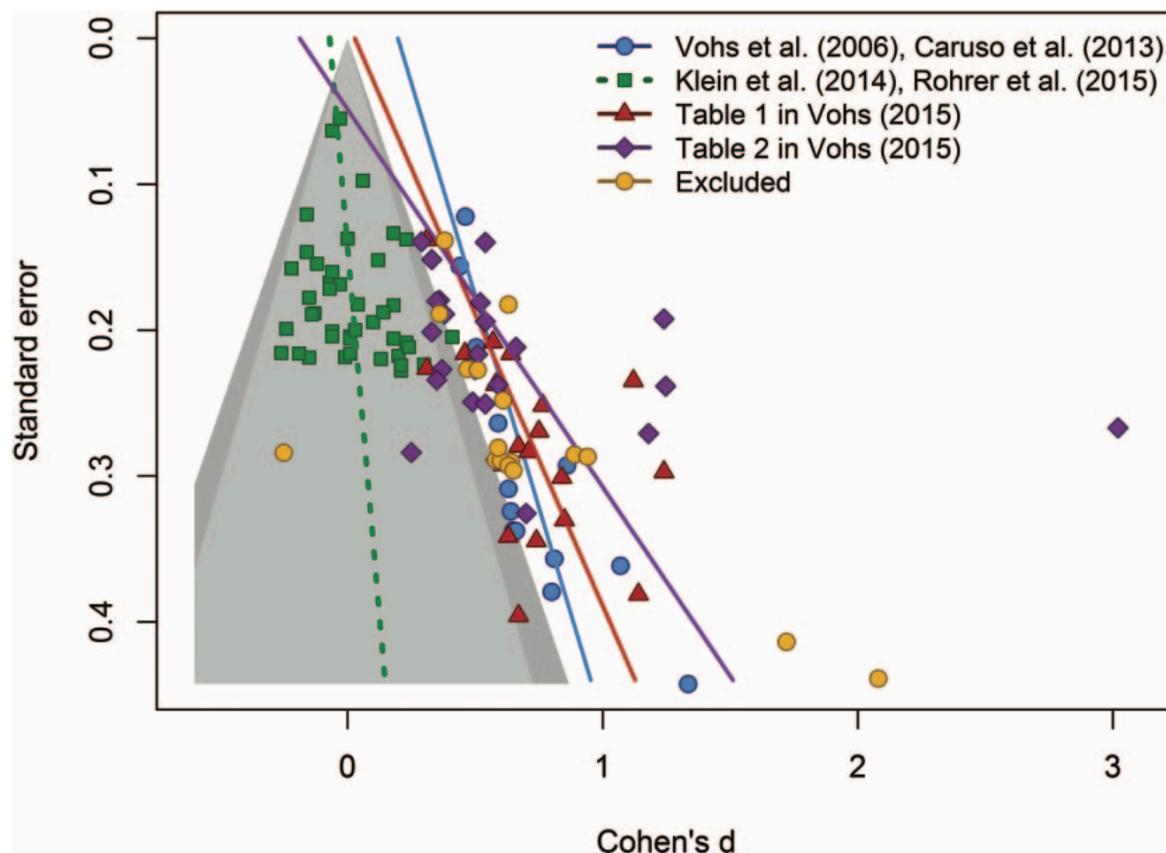
Funnel plot showing obvious bias

COMMENT

Selection Bias, Vote Counting, and Money-Priming Effects: A Comment
on Rohrer, Pashler, and Harris (2015) and Vohs (2015)

Miguel A. Vadillo
King's College London

Tom E. Hardwicke and David R. Shanks
University College London

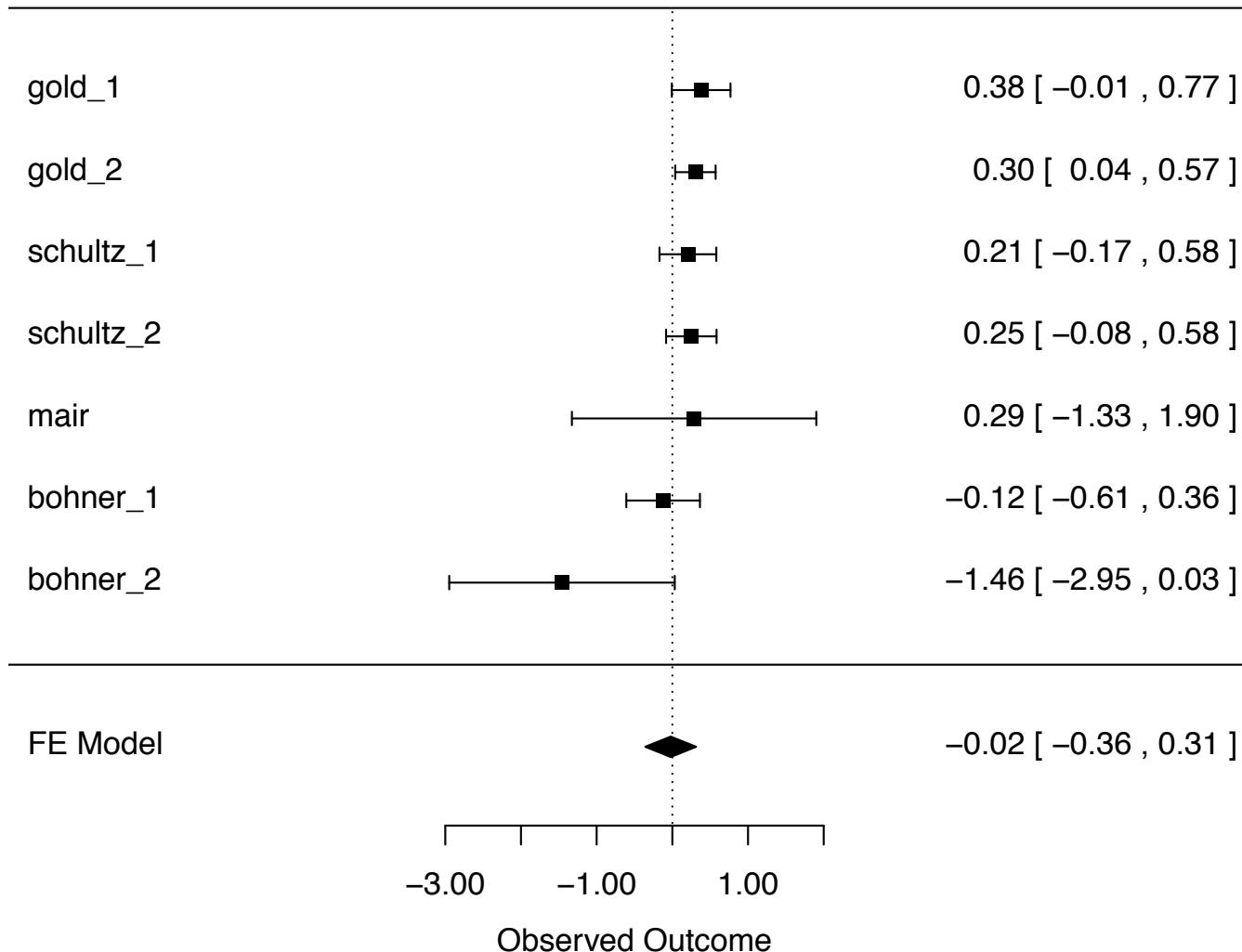


Models: Fixed vs. Random Effects

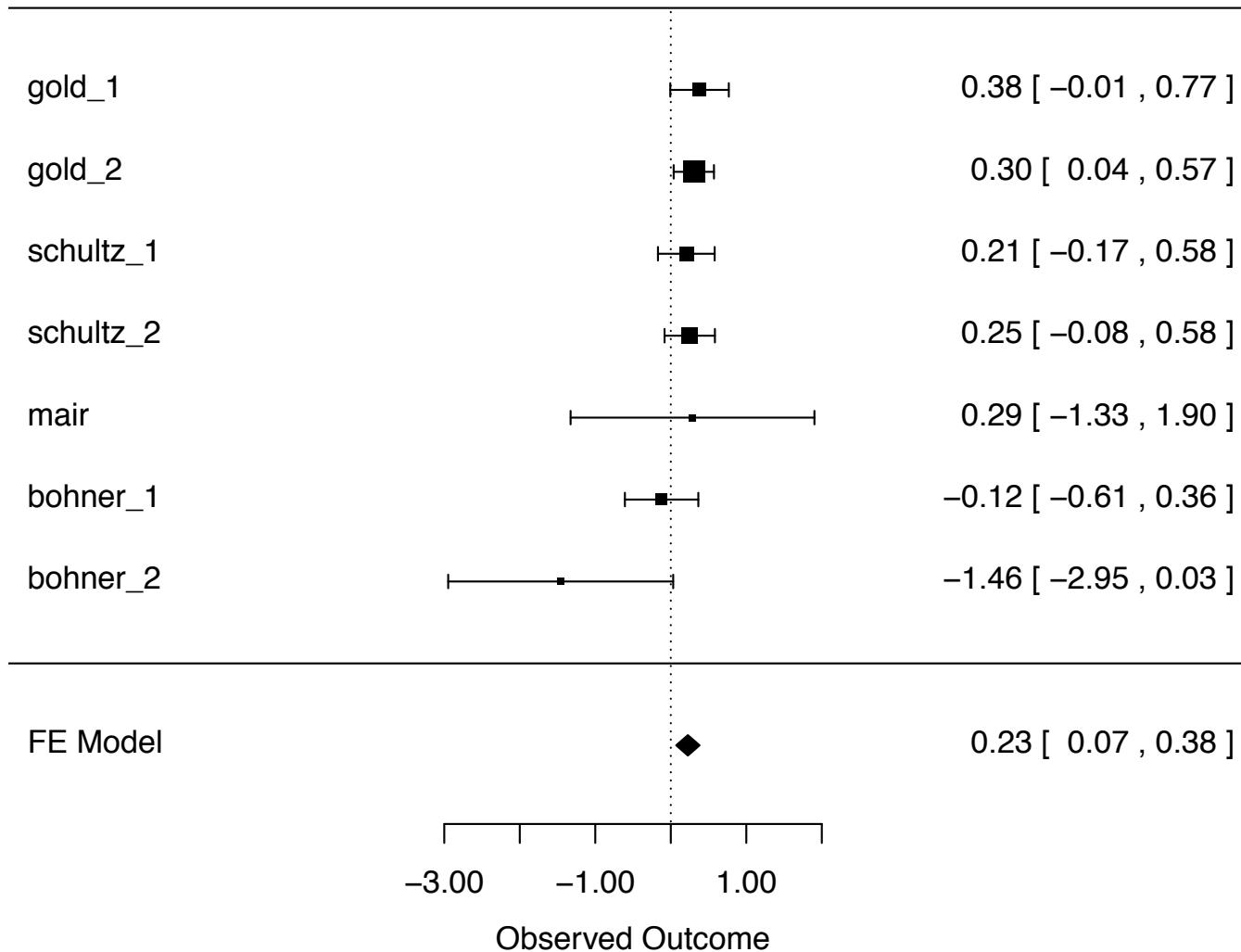
- Meta-analytic models are really just regression
 - With data weighted by precision: Inverse of variance (more variance, less weight)
- Basic MA: $es \sim 1$ (no predictors)
- **Fixed effects:** one true effect, all error is sampling error
- **Random effects:** studies are different from one another

Unweighted

Back to the hotel towel data...

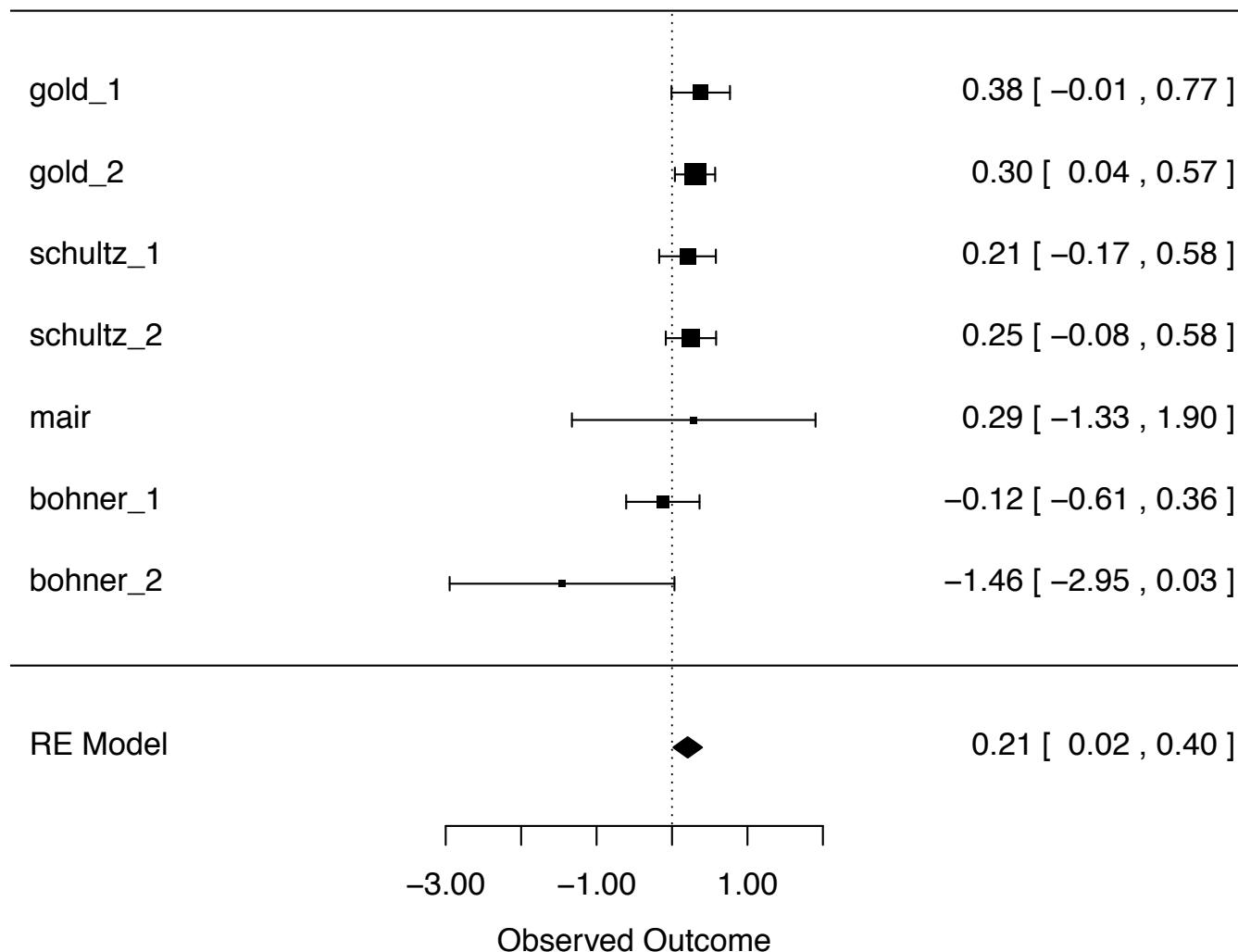


Fixed

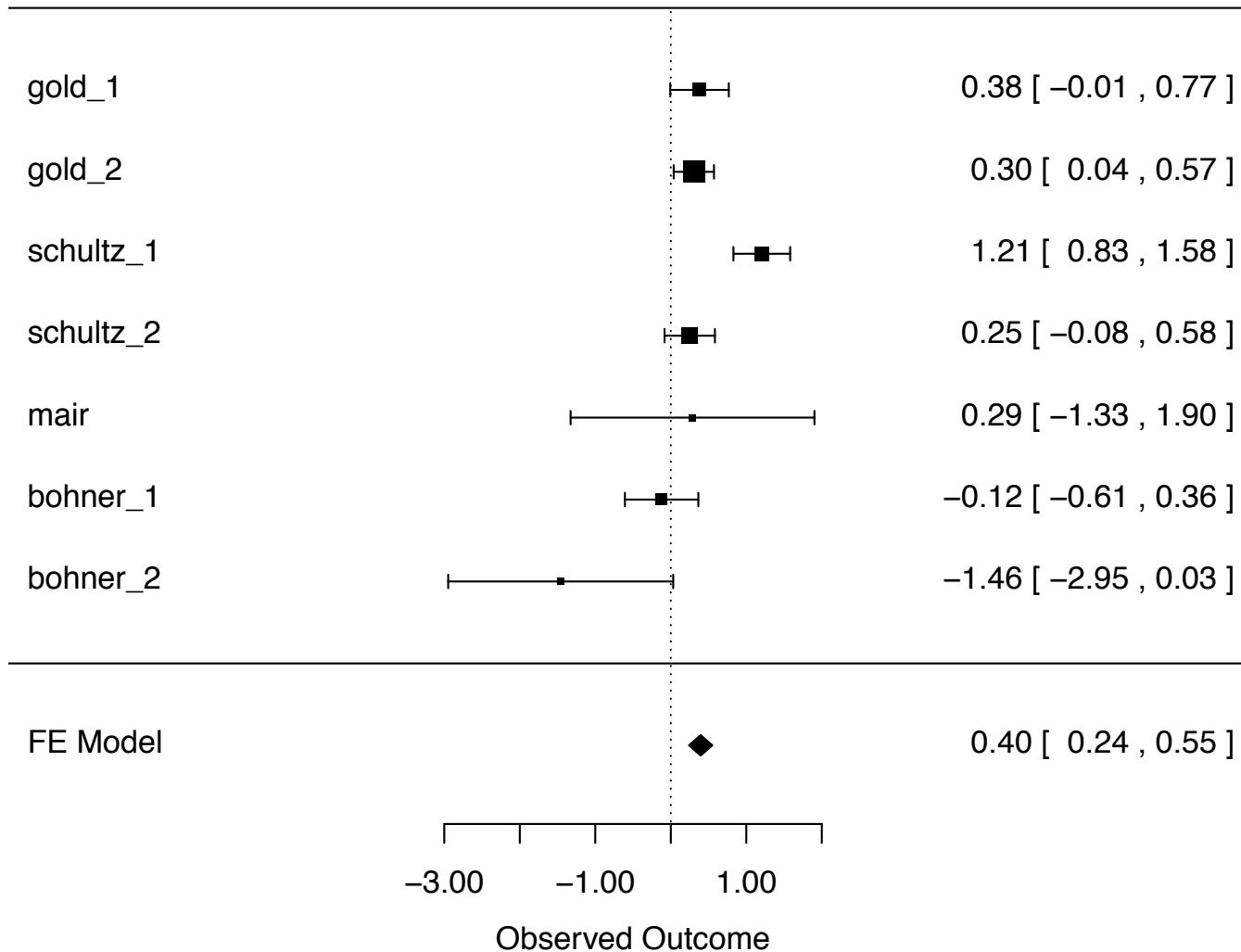


Random

In this case fixed and random effects give relatively similar answers.

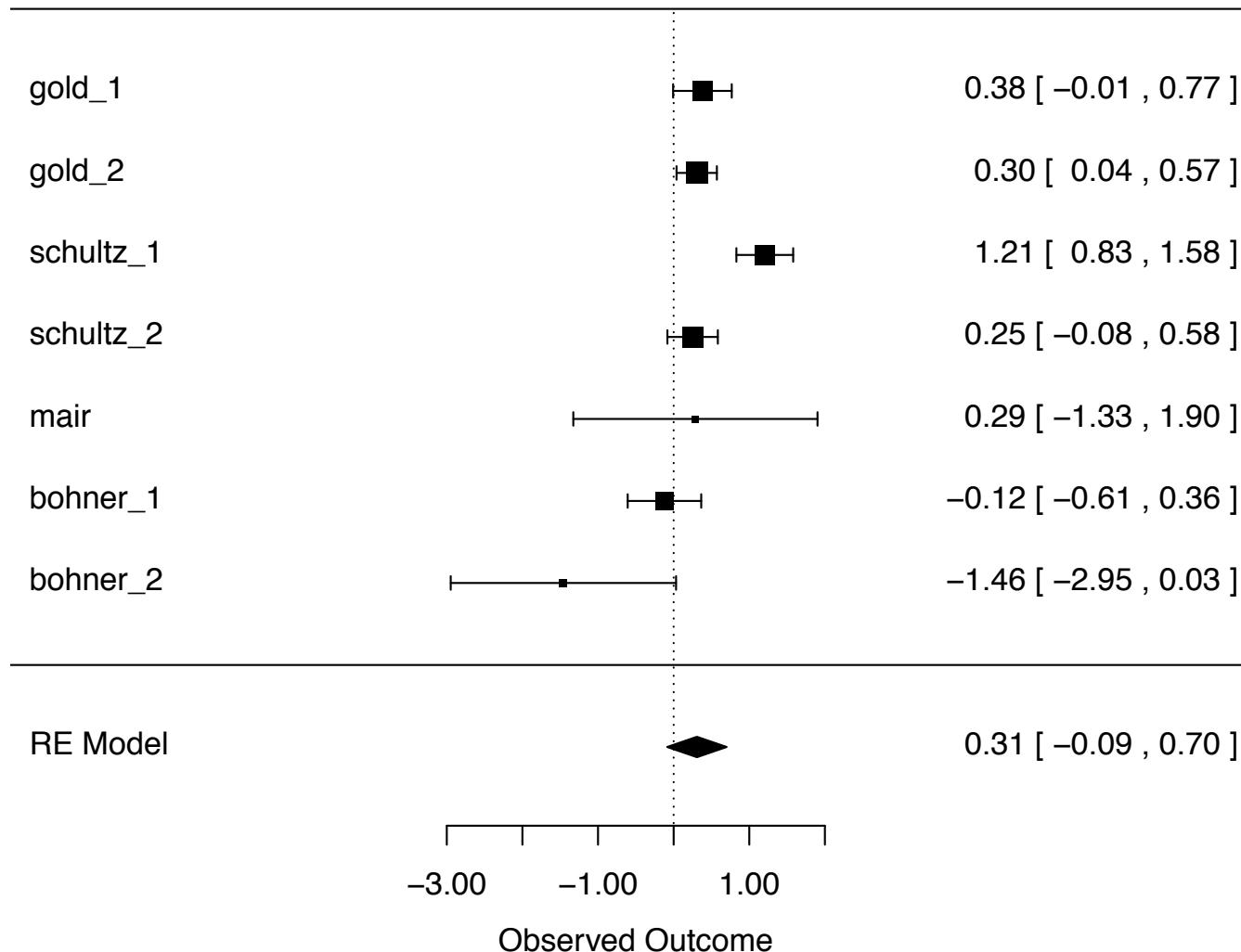


Fixed (fake)



Random (fake)

Here the random effects model is much more conservative.



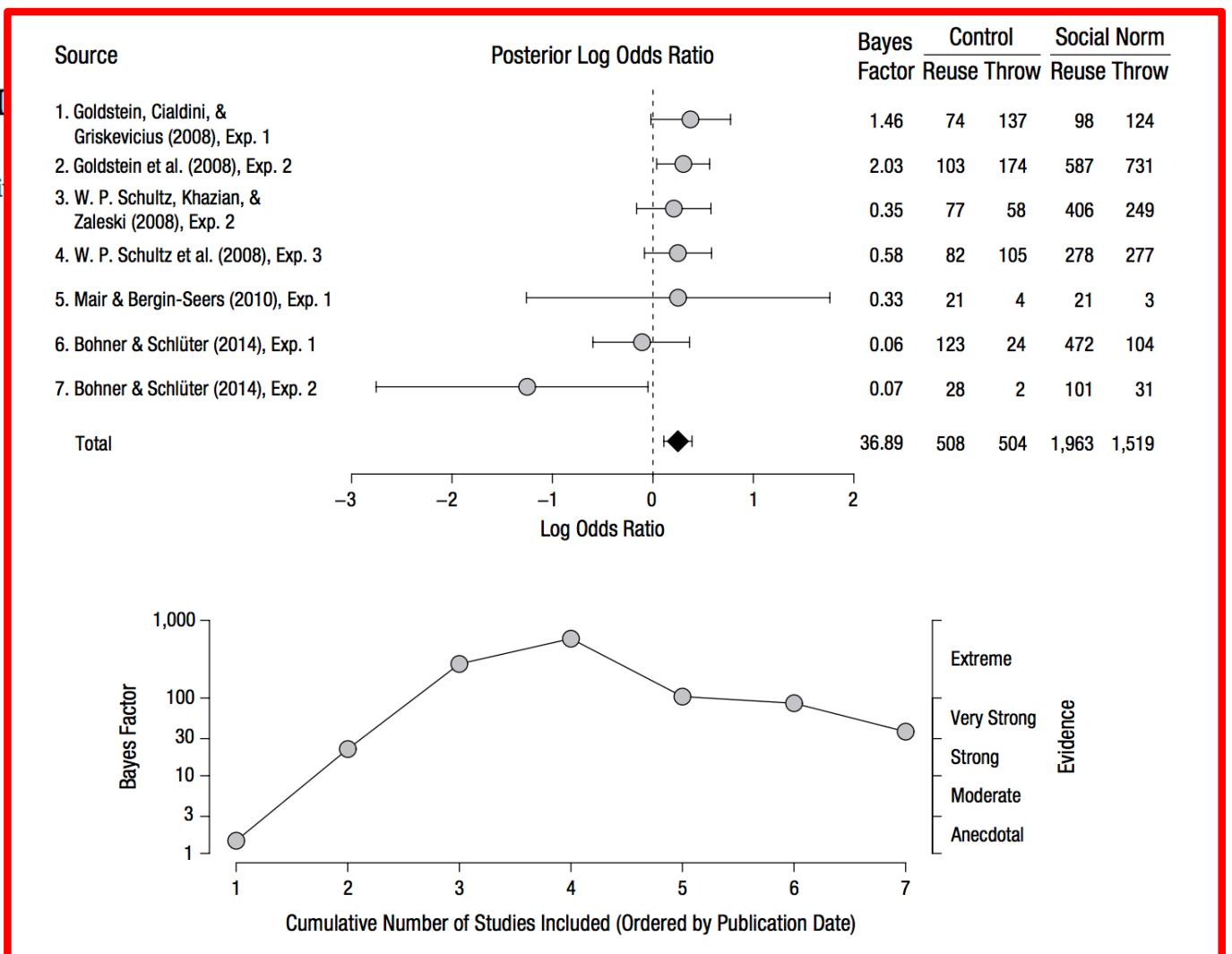
More models: Bayes 1

Bayesian Evidence Synthesis Can Reconcile Seemingly Inconsistent Results: The Case of Hotel Towel Reuse



Benjamin Scheibehenne¹, T
Eric-Jan Wagenmakers²

¹Faculty of Economics and Management, Uni
University of Amsterdam



More models: Bayes 2

Bayesian Evidence
Synthesis is No Substitute for
Meta-analysis

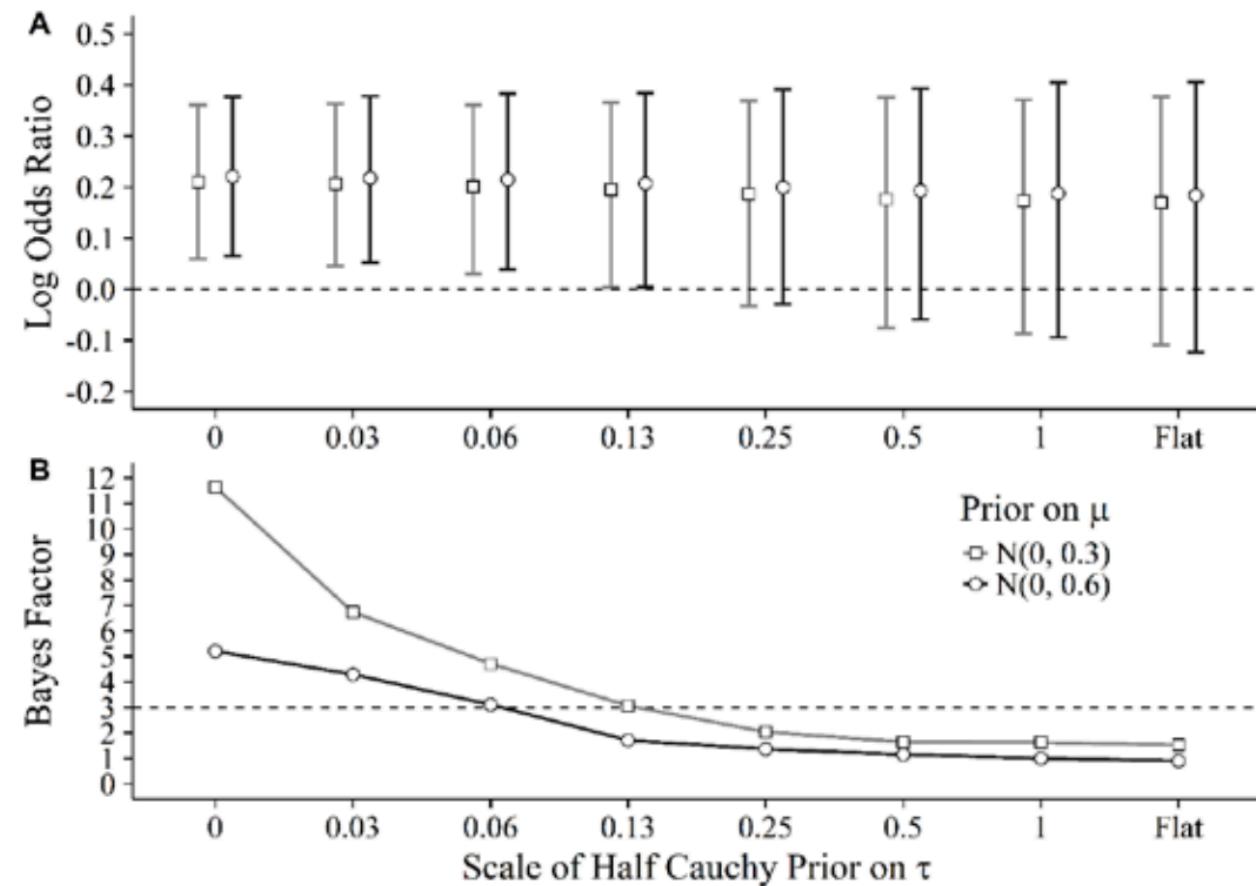
– A Re-analysis of Scheibehenne, Jamil and
Wagenmakers (2016)



Carlsson et
al. (2016)

**Moral of the story: you can
always fight about models.**

Figure 1.

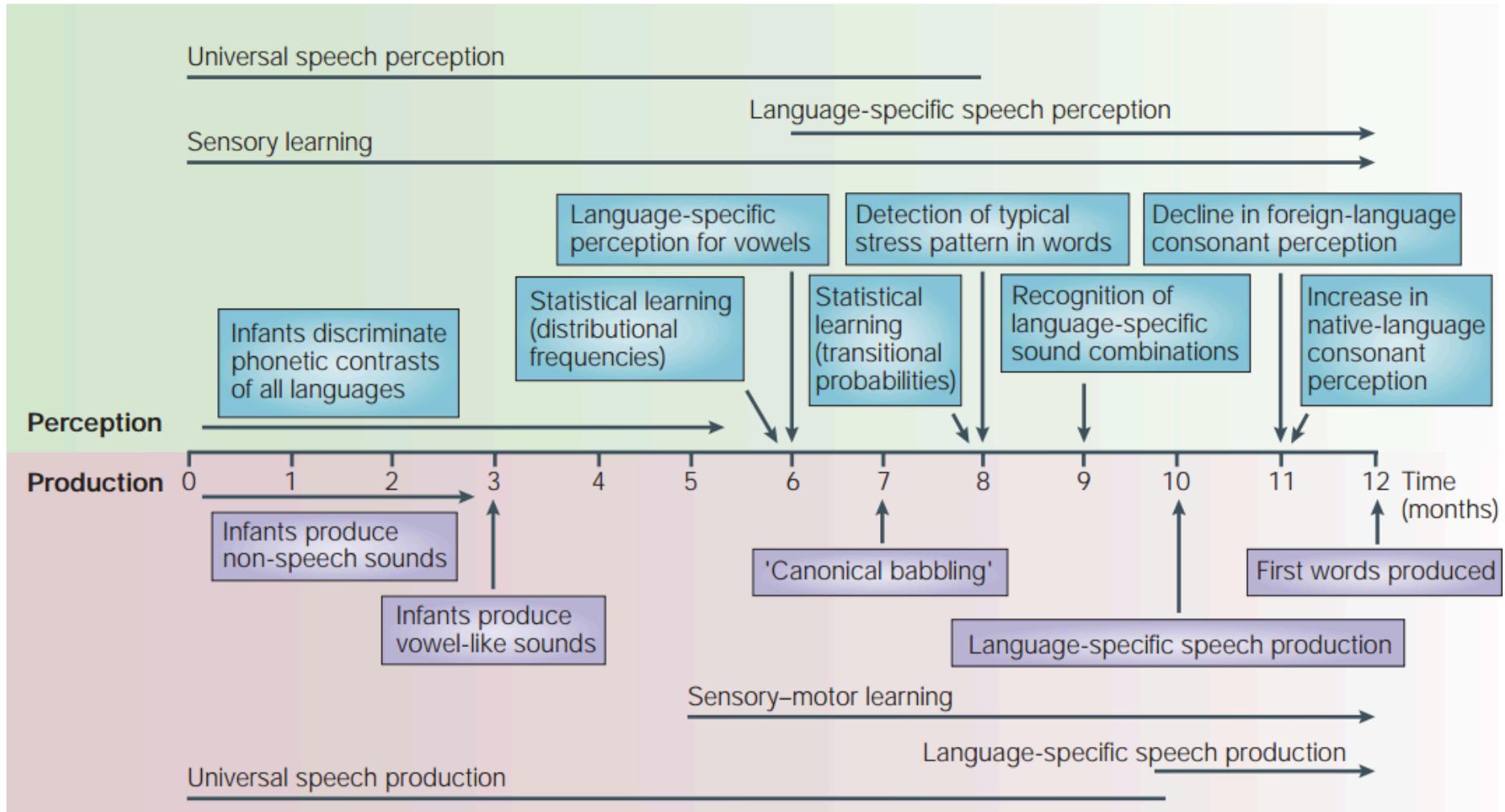


(A) Credible intervals of the meta-analytic log odds ratio μ , as well as (B) Bayes factors measuring evidence in favor of a non-zero effect for different prior distributions of μ and τ .

Outline

- Why meta-analysis?
 - History
 - Use cases: summary MA, internal MA
- Technical details
 - Coding issues
 - Effect sizes
 - Models: fixed, random
 - Funnel plots and publication bias
- **MetaLab: A fun meta-research project**

Quantitative theory in infancy

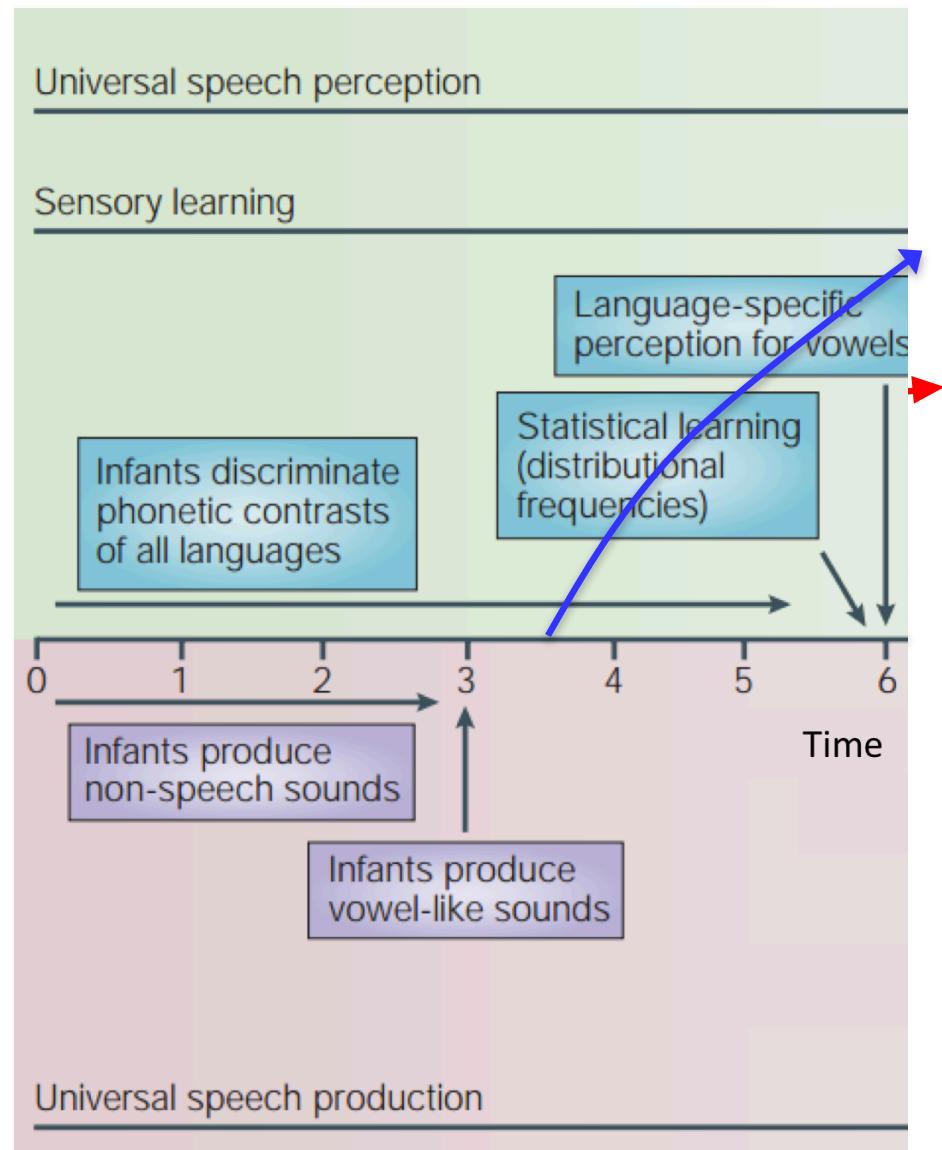


Kuhl (2004)

What would a summary look like?

Issues:

- Binary point of emergence
- Stage-like transitions
- Lack of variability in estimate
- Cross-phenomenon comparisons difficult



Quantitative synthesis via meta-analysis

MetaLab

Interactive tools for community-augmented meta-analysis,
power analysis, and experimental planning in language acquisition research



11



Meta-analyses

258



Papers

938



Effect sizes

11,628



Participants

Meta-analyses currently in MetaLab:



Infant directed speech preference

Looking times as a function of whether infant-directed vs. adult-directed speech is presented as stimulation.



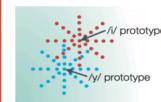
Label advantage in concept learning

Infants' categorization judgments in the presence and absence of labels.



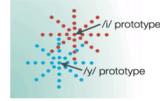
Gaze following

Gaze following using standard multi-alternative forced-choice paradigms.



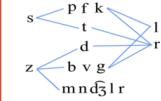
Vowel discrimination (native)

Discrimination of native-language vowels, including results from a variety of methods.



Vowel discrimination (non-native)

Discrimination of non-native vowels, including results from a variety of methods.



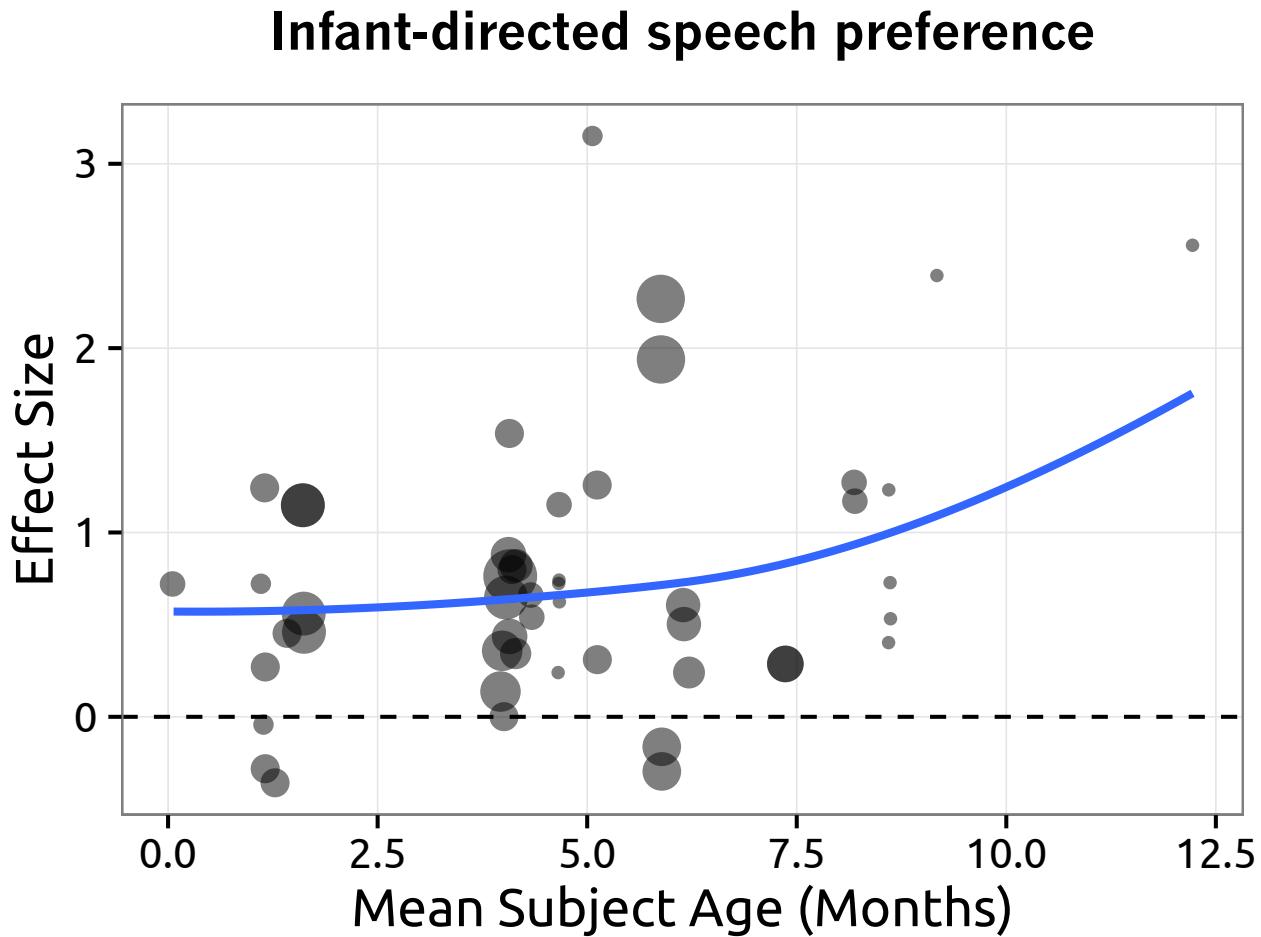
Phonotactic learning

Infants' ability to learn phonotactic generalizations from a short exposure.

<http://metalab.stanford.edu>

Lewis et al. (under review); Bergmann et al. (in press), *Child Dev*

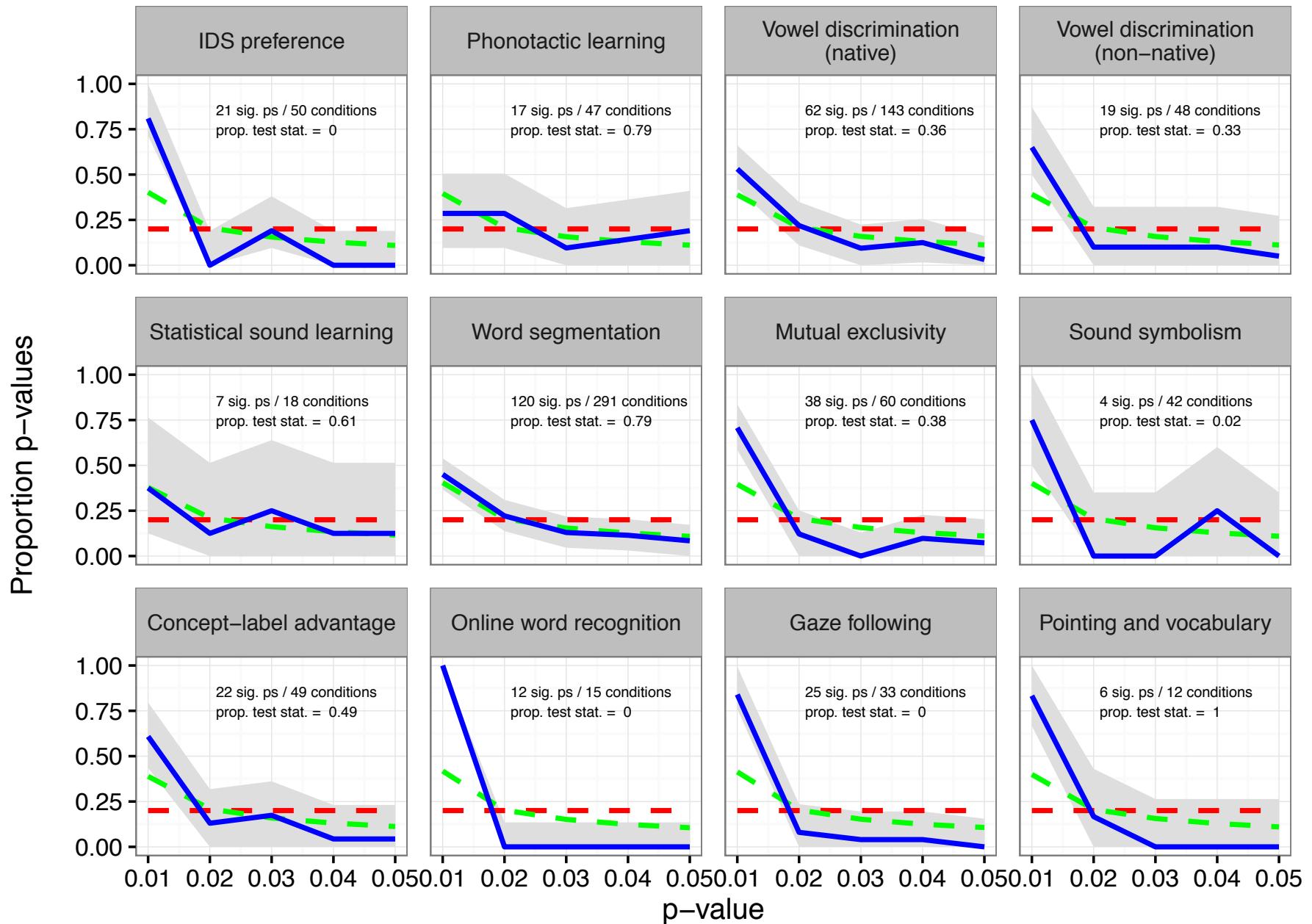
Individual MAs



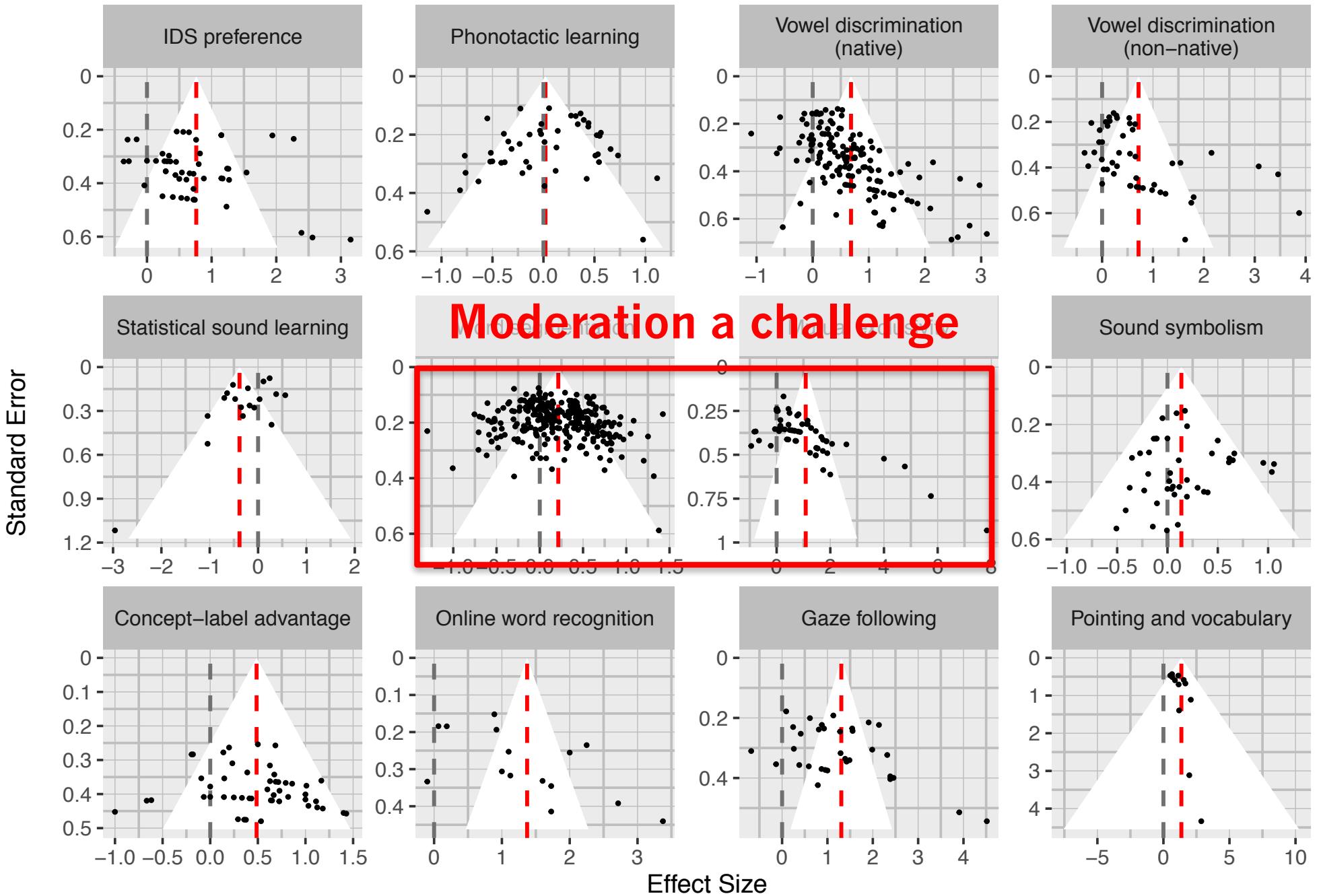
Note use of moderated meta-analytic model

P-curves

— Null of no effect — Null of 33% power — Observed

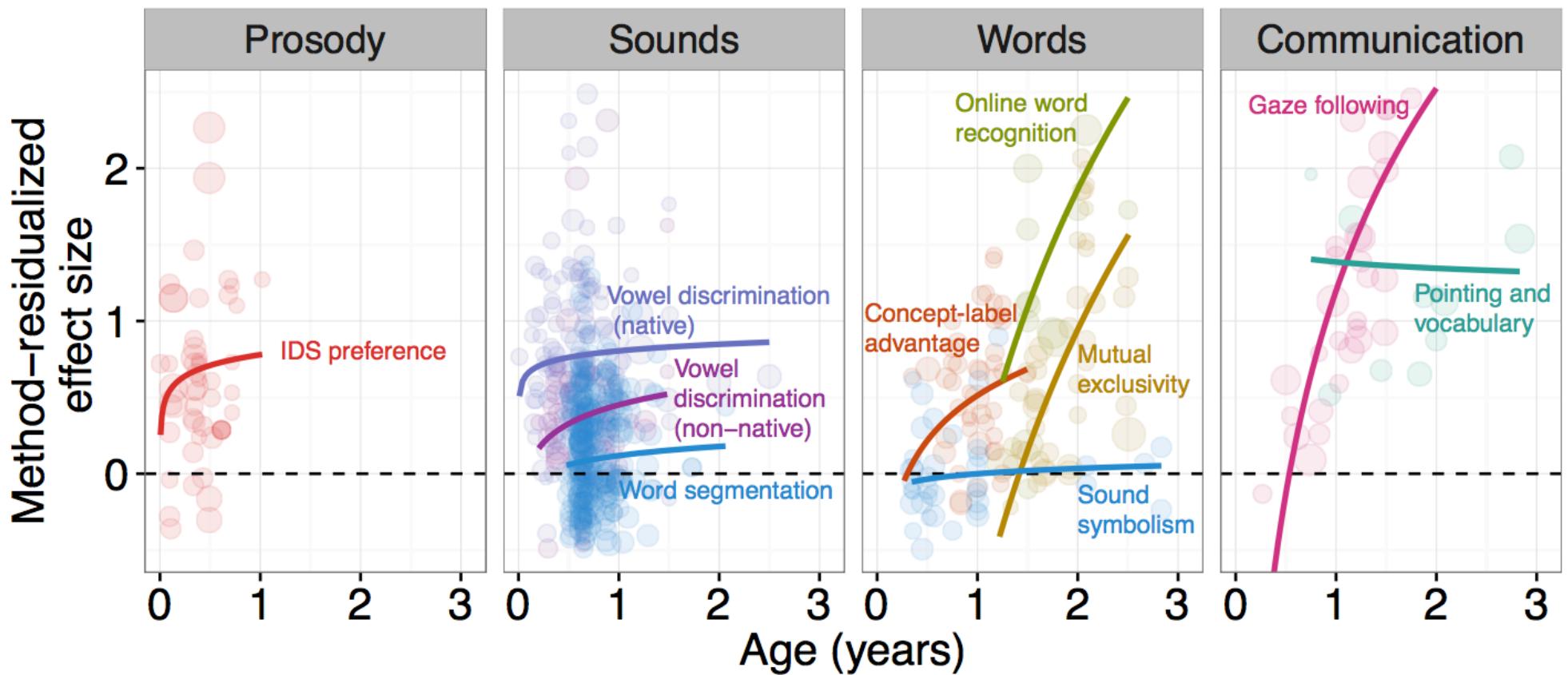


Funnel plots



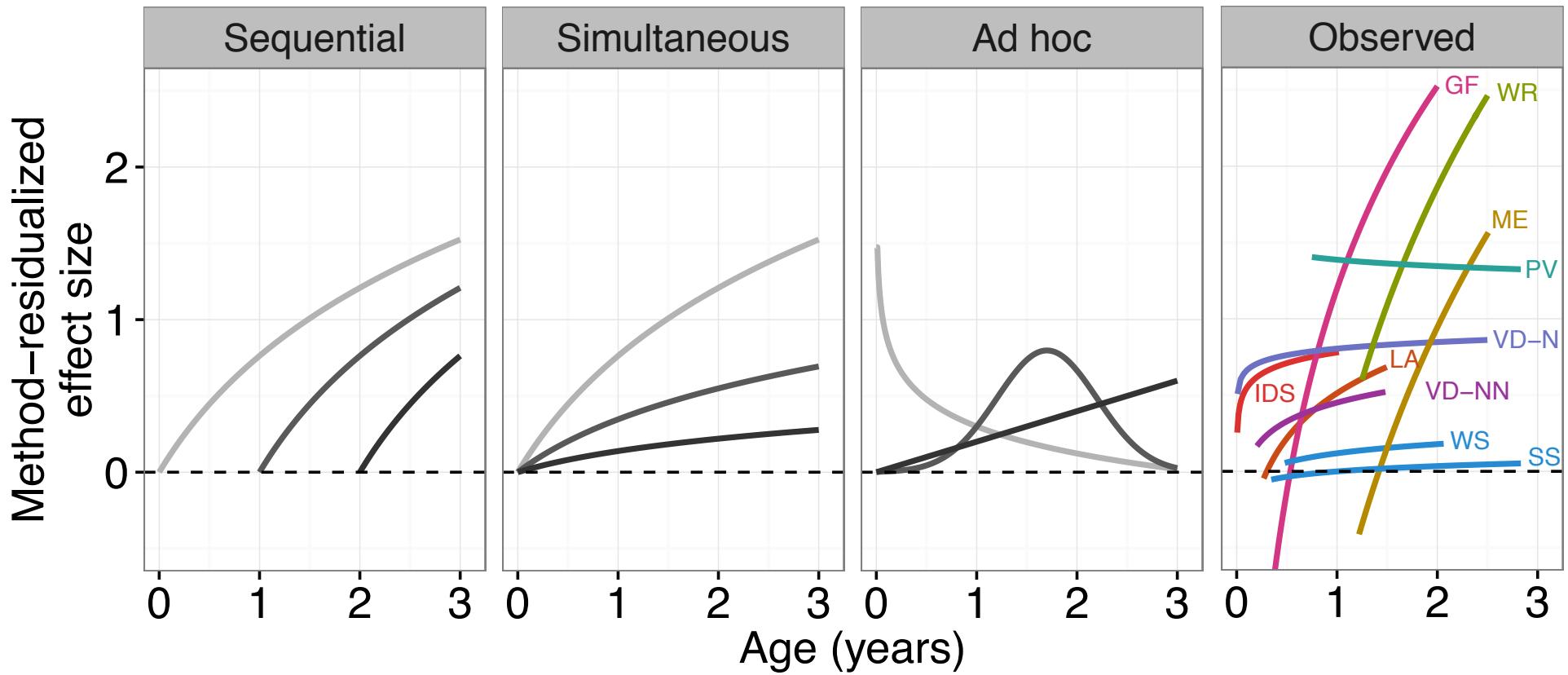
Meta-meta analysis

1. Trajectories vary by domain – explicit behaviors grow, implicit learning stays below the surface
2. Relatively large effects but much variability and many studies still underpowered



Meta-meta analysis

Capacities emerge interactively – no support for stage-like emergence



Overall signs of a relatively healthy literature,
albeit one with some work to do...