The title

First Author[1] & Ernst-August Doelle[1,2]

[1] Wilhelm-Wundt-University

[2] Konstanz Business School

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

The authors made the following contributions. First Author: Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing; Ernst-August Doelle: Writing - Review & Editing, Supervision.

Correspondence concerning this article should be addressed to First Author, Postal address. E-mail: my@email.com

Abstract

The abstract.

Keywords: keywords

Word count: X

The title

The main points

What do we want to argue?

1. When running these elaborate, many-lab style studies, one should make good use of the wealth of data that is obtained. Typically, a frequentist meta-analysis is conducted, in which either a fixed or random effects structure is applied. We would argue that a Bayesian hierarchical model or a model-averaged Bayesian meta-analysis would be better. (Example of good analysis for a "many"-project: ManyBabies The ManyBabies Consortium, 2020).
2. If there are multiple defensible analytic choices, such as different exclusion criteria, different prior settings, one should consider a multiverse analysis to explore the consequences of these choices. This allows for a broad overview and avoid (accidental) cherry-picking.
3. One should conduct a Bayesian analysis, because that allows for obtaining evidence for the null-hypothesis and quantifying uncertainty.
4. Don't exclude labs or countries because they contain few observations and are therefore uncertain. Using a hierarchical model will account for the uncertainty due to sample size and including more observations will always increase the resolution and informativeness of the data.

The choice of a Bayesian model-averaged meta-analysis or full hierarchical analysis can be debated. If participant-level covariates are of interest, a hierarchical model would be better suited. Also, more flexible in terms of testing specific predictions, such as whether the effect is positive across all labs/sites/countries.

Reanalysis

We just run the analyses for all *unique* dataset based on the 72 exclusion criteria constellations. Let's look at the table that shows all possibilities.

```
## [1] 72
```

Table 1
Exclusion constellations and resulting sample sizes

Participant-level	N-based	Protocol	Timing-based	In-House considered	Sample Size	N Studies
All	All	All	All	No	2225	21
White & US-born	All	All	All	No	1880	21
US-Identity > 7	All	All	All	No	1699	21
All	N > 60	All	All	No	2067	17
White & US-born	N > 60	All	All	No	1746	17
US-Identity > 7	N > 60	All	All	No	1593	17
All	N > 80	All	All	No	1866	14
White & US-born	N > 80	All	All	No	1545	14
US-Identity > 7	N > 80	All	All	No	1392	14
All	All	AA	All	No	798	9
White & US-born	All	AA	All	No	453	9
US-Identity > 7	All	AA	All	No	272	9
All	N > 60	AA	All	No	699	7
White & US-born	N > 60	AA	All	No	378	7
US-Identity > 7	N > 60	AA	All	No	225	7
All	N > 80	AA	All	No	699	7
White & US-born	N > 80	AA	All	No	378	7
US-Identity > 7	N > 80	AA	All	No	225	7
All	All	All	After prereg	No	1659	20
White & US-born	All	All	After prereg	No	1314	20
US-Identity > 7	All	All	After prereg	No	1133	20
All	N > 60	All	After prereg	No	1544	17
White & US-born	N > 60	All	After prereg	No	1223	17
US-Identity > 7	N > 60	All	After prereg	No	1070	17
All	N > 80	All	After prereg	No	1343	14
White & US-born	N > 80	All	After prereg	No	1022	14
US-Identity > 7	N > 80	All	After prereg	No	869	14
All	All	AA	After prereg	No	797	9
White & US-born	All	AA	After prereg	No	452	9
US-Identity > 7	All	AA	After prereg	No	271	9
All	N > 60	AA	After prereg	No	698	7
White & US-born	N > 60	AA	After prereg	No	377	7
US-Identity > 7	N > 60	AA	After prereg	No	224	7
All	N > 80	AA	After prereg	No	698	7
White & US-born	N > 80	AA	After prereg	No	377	7
US-Identity > 7	N > 80	AA	After prereg	No	224	7
All	All	All	All	Yes	2211	21
White & US-born	All	All	All	Yes	983	16
US-Identity > 7	All	All	All	Yes	272	9

Table 1 continued

Participant-level	N-based	Protocol	Timing-based	In-House considered	Sample Size	N Studies
All	N > 60	All	All	Yes	2053	17
White & US-born	N > 60	All	All	Yes	897	13
US-Identity > 7	N > 60	All	All	Yes	225	7
All	N > 80	All	All	Yes	1852	14
White & US-born	N > 80	All	All	Yes	864	12
US-Identity > 7	N > 80	All	All	Yes	225	7
All	All	AA	All	Yes	799	9
White & US-born	All	AA	All	Yes	453	9
US-Identity > 7	All	AA	All	Yes	272	9
All	N > 60	AA	All	Yes	700	7
White & US-born	N > 60	AA	All	Yes	378	7
US-Identity > 7	N > 60	AA	All	Yes	225	7
All	N > 80	AA	All	Yes	700	7
White & US-born	N > 80	AA	All	Yes	378	7
US-Identity > 7	N > 80	AA	All	Yes	225	7
All	All	All	After prereg	Yes	1650	20
White & US-born	All	All	After prereg	Yes	777	15
US-Identity > 7	All	All	After prereg	Yes	271	9
All	N > 60	All	After prereg	Yes	1535	17
White & US-born	N > 60	All	After prereg	Yes	702	13
US-Identity > 7	N > 60	All	After prereg	Yes	224	7
All	N > 80	All	After prereg	Yes	1334	14
White & US-born	N > 80	All	After prereg	Yes	669	12
US-Identity > 7	N > 80	All	After prereg	Yes	224	7
All	All	AA	After prereg	Yes	798	9
White & US-born	All	AA	After prereg	Yes	452	9
US-Identity > 7	All	AA	After prereg	Yes	271	9
All	N > 60	AA	After prereg	Yes	699	7
White & US-born	N > 60	AA	After prereg	Yes	377	7
US-Identity > 7	N > 60	AA	After prereg	Yes	224	7
All	N > 80	AA	After prereg	Yes	699	7
White & US-born	N > 80	AA	After prereg	Yes	377	7
US-Identity > 7	N > 80	AA	After prereg	Yes	224	7

Note. Blue rows refer to Klein et al.'s key analyses; pink rows refer to Chatard et al.'s key analyses; green rows refer to the Inclusive analyses; grey rows are repeated data sets and not included in the multiverse analysis; AA = Author-Advised. 'In-House considered' indicates that the participant-level exclusion criteria are also applied to the In-House studies (missing data excluded).

Reanalysis with Exclusion Criterion .1.1.1.2

This is our all-inclusive analysis. It includes data from all participants that have completed the relevant measures, all studies, and applied the participant-level exclusion criteria to both Author-Advised (AA) protocols and In-House (IH) protocols. Note that for many IH protocols, the relevant information for participant-level exclusion criteria 3 (and 2 to a lesser degree) is missing. We decided to exclude participants for whom nationality and country of origin (exclusion criterion 2) or identification with American culture (exclusion criterion 3) is unknown as we cannot assume that people met this requirement. For exclusion criterion 1 – completeness of the measures –, we did retain participants for labs where no information was available, as long as they were assigned to an experimental condition and answered both items of the dependent variable.

Reanalysis with Exclusion Criterion .2.1.2.1

This is the original main analysis that is the basis for the key claims of the Many Labs 4 project, as reported in the published paper (Klein et al., 2022). The authors included participants whose data was collected after the lead team posted their preregistration, only studies that featured more than 60 observations (before participant-level exclusions). The participant-level exclusion criteria were only applied to AA-studies, which means that for exclusion criteria 2 and 3 all participants from the IH-studies were retained, indicating that the authors implicitly assumed they were all American, born in the US, and strongly identified with American culture. Exclusion of the observations collected by labs prior to the lead team's preregistration was uploaded caused the authors to discard 566 observations (25.44%). Note that the timing-based and the study-level the N-based exclusions are applied in the published article but not in the preprint that appeared in 2019.

Table 2
Bayes factors for key analyses.

			Evidence		
Participant-level	Sample size	Labs	BF_{0f}	BF_{01}	BF_{0+}
Inclusive					
All	2211	21	35.21	10.33	10,490.12
White & US-born	983	16	21.46	13.99	2,538.61
US-Identity > 7	272	9	8.44	2.77	11.88
Klein et al. (2022)					
All	1544	17	12.44	5.35	628.62
White & US-born	1223	17	9.35	4.21	204.55
US-Identity > 7	1070	17	6.95	4.50	157.52
Chatard et al. (2020)					
All	699	7	14.61	2.16	13.32
White & US-born	378	7	6.76	0.95	2.61
US-Identity > 7	225	7	4.47	0.79	1.42

Note. All Bayes factors are reported in favor of the null model.

Reanalysis with Exclusion Criterion .3.2.1.1

(1,3,2,1,1), (2,3,2,1,1), and (3,3,2,1,1) from the comment by Chatard, Hirschberger, and Pyszczynski (2020). The authors argued that a valid test of the theory as formulated by the original authors would only include the AA-studies. They additionally read the preregistration as stating that only labs that collected data from at least 80 participants would be included in the analysis. Following Klein et al. (2022), they applied the participant-level exclusion criteria only the AA-studies, although in this case that does not matter as all IH-data is excluded anyway. No timing-based exclusion criteria were applied.

Summary

We want to create a figure that shows the evidence against heterogeneity on the x-axis and evidence against the effect on the y-axis. The size of the points will reflect N. The effect-evidence will be reflected by a weigthed average (model average) of the common effect

Table 3
Bayes factors for key analyses.

			Effect BF_{01}			Heterogeneity BF_{01}
Participant-level	Sample size	Labs	Default	Oosterwijk	Vohs	Default
Inclusive						
All	2211	21	12.60	44.69	16.64	2.28
White & US-born	983	16	19.42	67.73	25.90	2.03
US-Identity > 7	272	9	4.13	3.90	2.44	1.79
Klein et al. (2022)						
All	1544	17	4.45	10.71	4.18	1.89
White & US-born	1223	17	2.79	5.02	2.14	1.45
US-Identity > 7	1070	17	3.34	5.90	2.57	1.33
Chatard et al. (2020)						
All	699	7	4.04	6.36	2.97	2.63
White & US-born	378	7	1.43	0.90	0.66	2.06
US-Identity > 7	225	7	1.44	0.72	0.62	1.88

Note. All Bayes factors are reported in favor of the null model. The different column names for the effect BF_{01} refer to the different priors used.

and unconstrained effect model. The heterogeneity-evidence will be simple the evidence for the fixed model vs. the unconstrained model.

We also want to include some forest plots. Let's do our most inclusive one (set 11112), Klein et al.'s primary one (set 12121) and Chatard et al.'s chosen one (set 33211). We want these for both the hierarchical analysis and the meta-analysis.

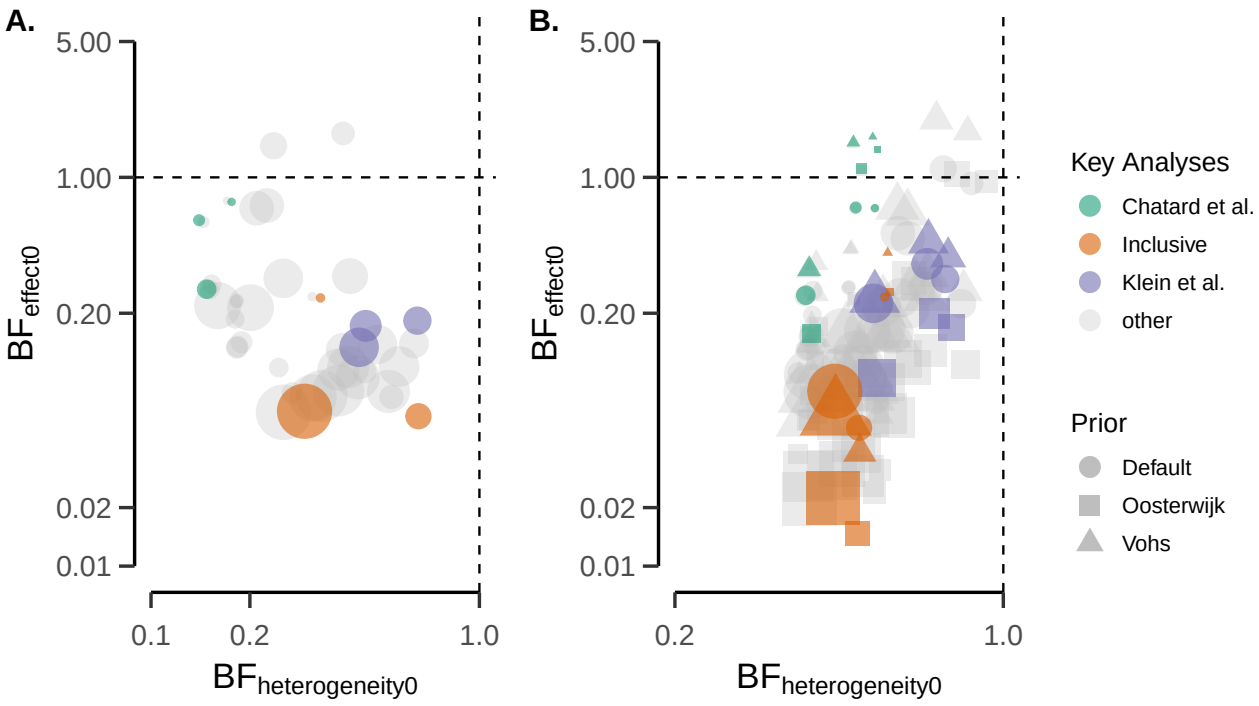

Figure 1. Results from the multiverse analysis: Bayes factors in favor of a mortality salience effect are above the horizontal line, Bayes factors against the mortality salience effect are below the horizontal line. Panel **A.** shows the evidence obtained from the hierarchical analyses and panel **B.** shows the evidence obtained from the model-averaged meta-analysis. All analyses provide evidence against between-study heterogeneity as shown by all heterogeneity Bayes factors are smaller than 1 on the x-axis. The color of the points refers to the different key analyses sets, the shape of the points refers the different prior setting in the meta-analysis, and the size of the points refers to the number of participants the analysis is based on. The majority of analyses provide evidence against the mortality salience effect.

References

Chatard, A., Hirschberger, G., & Pyszczynski, T. (2020). *A Word of Caution about Many Labs 4: If You Fail to Follow Your Preregistered Plan, You May Fail to Find a Real Effect* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/ejubn

Klein, R. A., Cook, C. L., Ebersole, C. R., Vitiello, C., Nosek, B. A., Hilgard, J., … Ratliff, K. A. (2022). Many labs 4: Failure to replicate mortality salience effect with and without original author involvement. *Collabra: Psychology*, *8*(1), 35271. https://doi.org/10.1525/collabra.35271

The ManyBabies Consortium. (2020). Quantifying Sources of Variability in Infancy Research Using the Infant-Directed-Speech Preference: *Advances in Methods and*

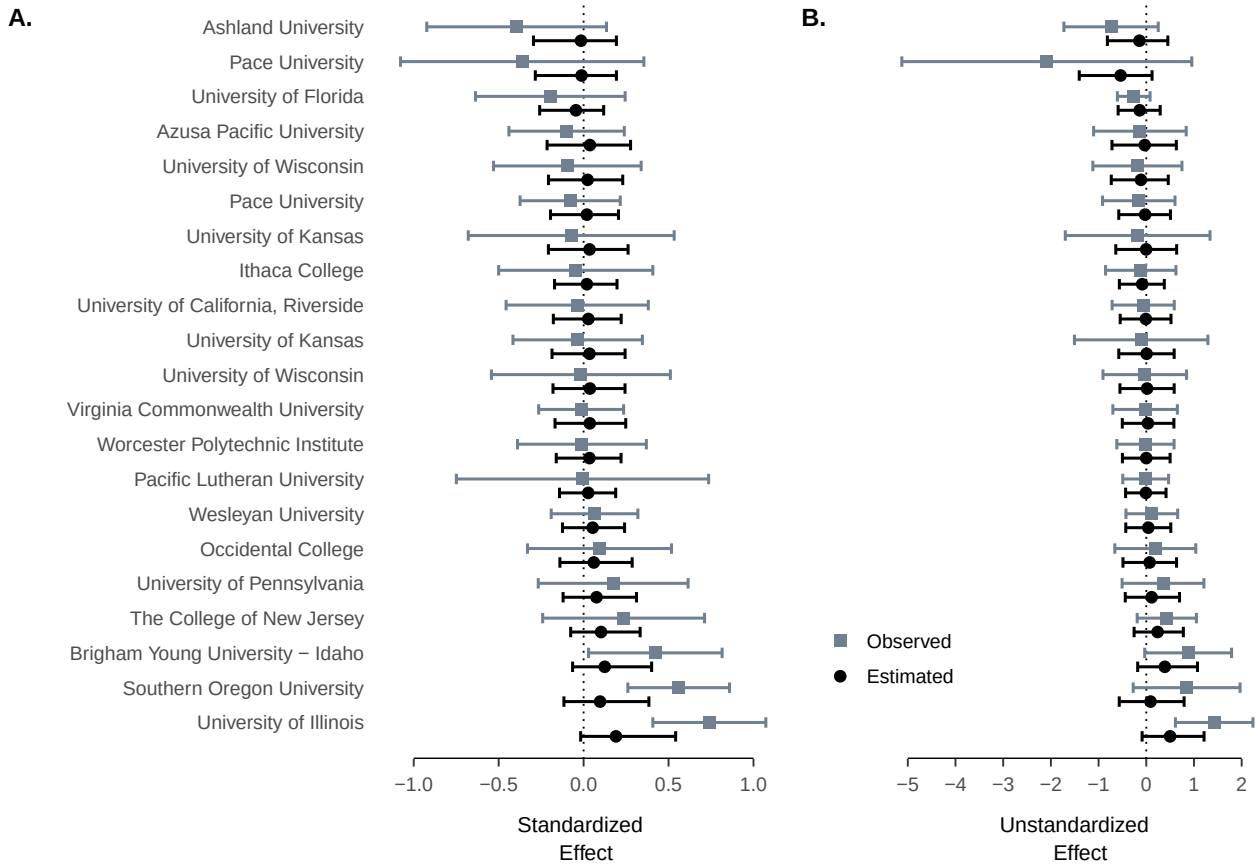

Figure 2. Forest plot with Bayesian parameter estimates for the most inclusive dataset featuring participant-level exclusion set 1 (applied to all participants) for all studies and all complete data. **A.** Bayesian meta-analysis (with two-sided default prior). The grey points represent calculated effect sizes with 95% confidence intervals, the black points represent estimated effect sizes from the random-effects alternative model with 95% credible intervals. **B.** Bayesian hierarchical analysis. The grey points represent unstandardized observed effects for each study with 95% confidence intervals. The black points represent estimated unstandardized effects from the unconstrained model with 95% credible intervals.

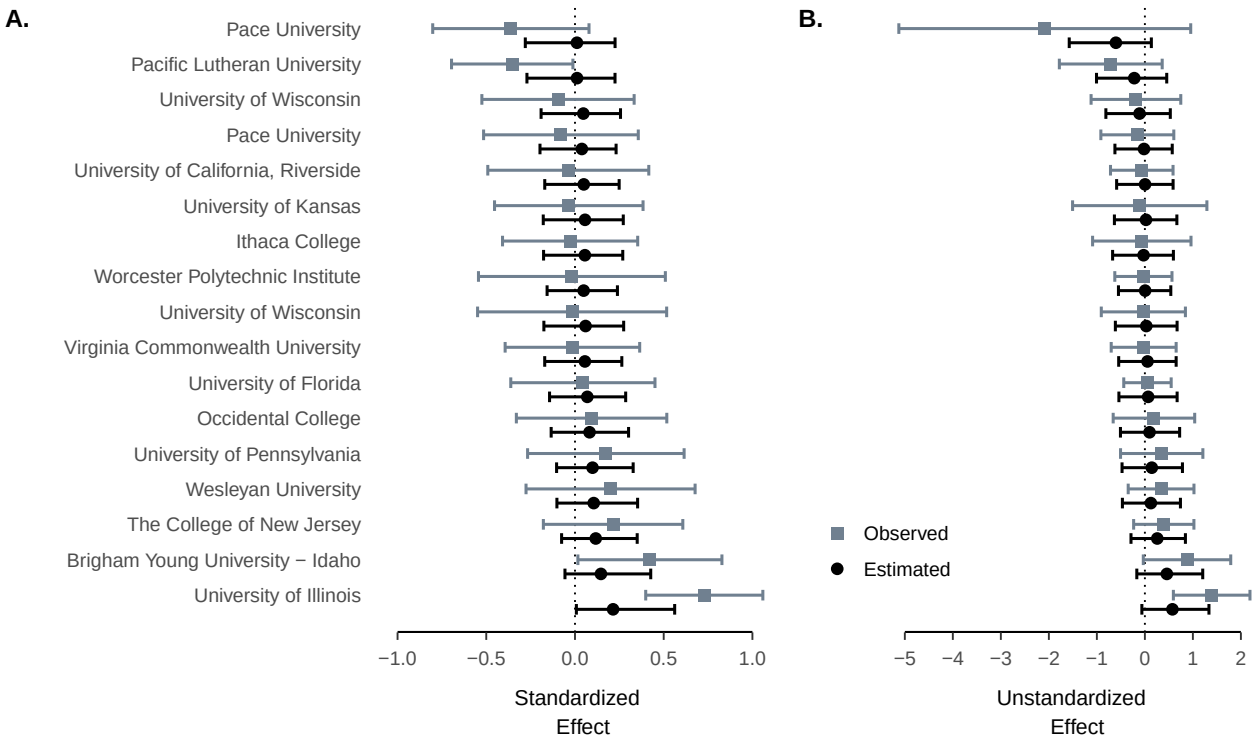

Figure 3. Forest plot with Bayesian parameter estimates for the critical analysis from Klein et al. (2022) featuring participant-level exclusion set 1 (only applied to AA-protocols) for studies with more than 60 particpants and excluding observation collected prior to the analysis preregistration date. **A.** Bayesian meta-analysis (with two-sided default prior). The grey points represent calculated effect sizes with 95% confidence intervals, the black points represent estimated effect sizes from the random-effects alternative model with 95% credible intervals. **B.** Bayesian hierarchical analysis. The grey points represent unstandardized observed effects for each study with 95% confidence intervals. The black points represent estimated unstandardized effects from the unconstrained model with 95% credible intervals.

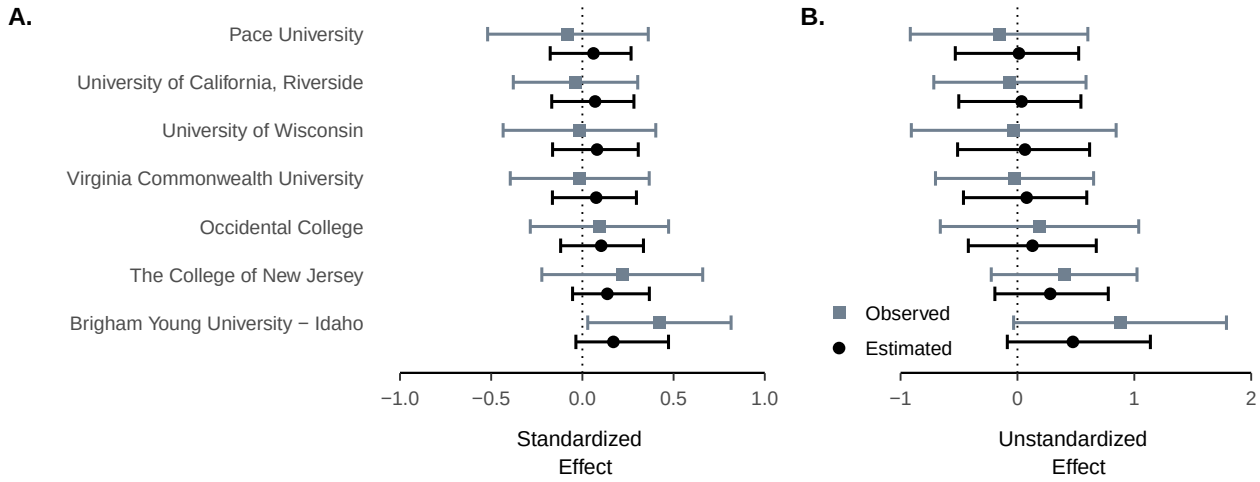

Figure 4. Forest plot with Bayesian parameter estimates for the critical analysis from Chatard et al. (2020) featuring participant-level exclusion set 3 for studies with more than 80 particpants and only author-advised studies included. **A.** Bayesian meta-analysis (with two-sided default prior). The grey points represent calculated effect sizes with 95% confidence intervals, the black points represent estimated effect sizes from the random-effects alternative model with 95% credible intervals. **B.** Bayesian hierarchical analysis. The grey points represent unstandardized observed effects for each study with 95% confidence intervals. The black points represent estimated unstandardized effects from the unconstrained model with 95% credible intervals.