Communicating Causal Effect Heterogeneity

- **How to best represent, model, and communicate heterogeneity**
- **How I learned to stop worrying and love the multilevel model**
- **There's no matrix like the $d \times p$ matrix**

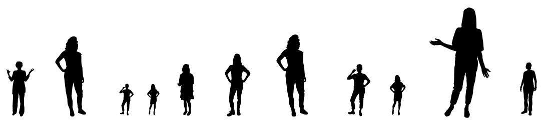
Matti Vuorre (Tilburg University)
https://vuorre.com
@matti.vuorre.com

Effects, sizes, and effect sizes

What you gonna count?

I. Imagine an important treatment and outcome, and an effect size β = x (p < .0001). Is this a big deal?

  I. Depends on x

  II. Depends on distribution of individuals' effects 

II. What does $p(\text{individual effects})$ mean for our interpretation of β

Team & paper

Niall Bolger
Columbia University

Matthew Kay
Northwestern University

Communicating causal effect heterogeneity

Matti Vuorre[1*], Matthew Kay[2] & Niall Bolger[3]
[1]Tilburg University, Department of Social Psychology
[2]Northwestern University, Computer Science and Communication Studies
[3]Columbia University, Department of Psychology

Advances in experimental, data collection, and analysis methods have brought population variability in psychological phenomena to the fore. Yet, current practices for interpreting such heterogeneity do not appropriately treat the uncertainty inevitable in any statistical summary. Heterogeneity is best thought of as a distribution of features with a mean (average person's effect) and variance (between-person differences). This expected heterogeneity distribution can be further summarized e.g. as a heterogeneity interval (Bolger et al., 2019). However, because empirical studies estimate the underlying mean and variance parameters with uncertainty, the expected distribution and interval will underestimate the actual range of plausible effects in the population. Using Bayesian hierarchical models, and with the aid of empirical datasets from social and cognitive psychology, we provide a walk-through of effective heterogeneity reporting and display tools that appropriately convey measures of uncertainty. We cover interval, proportion, and ratio measures of heterogeneity and their estimation and interpretation. These tools can be a spur to theory building, allowing researchers to widen their focus from population averages to population heterogeneity in psychological phenomena.

Keywords: heterogeneity, uncertainty, variation, multilevel model, statistics, visualization
Words: 7734

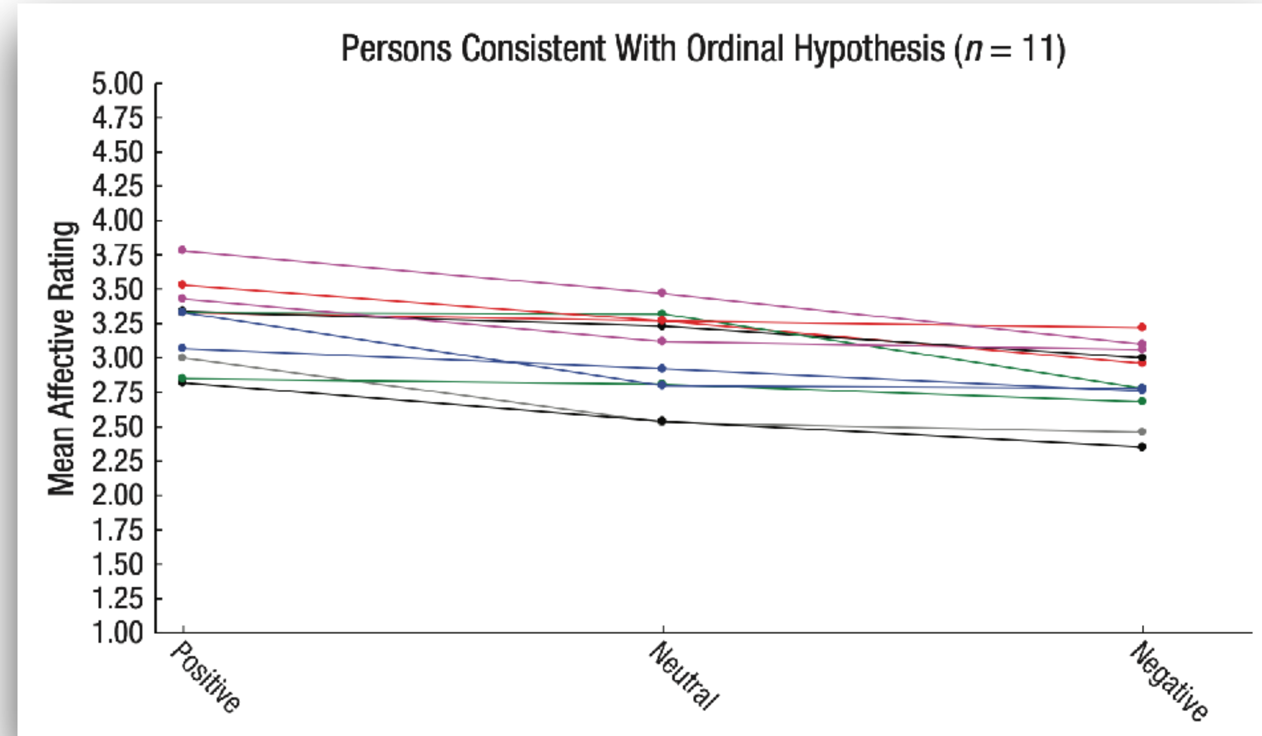
Beyond average treatment effects

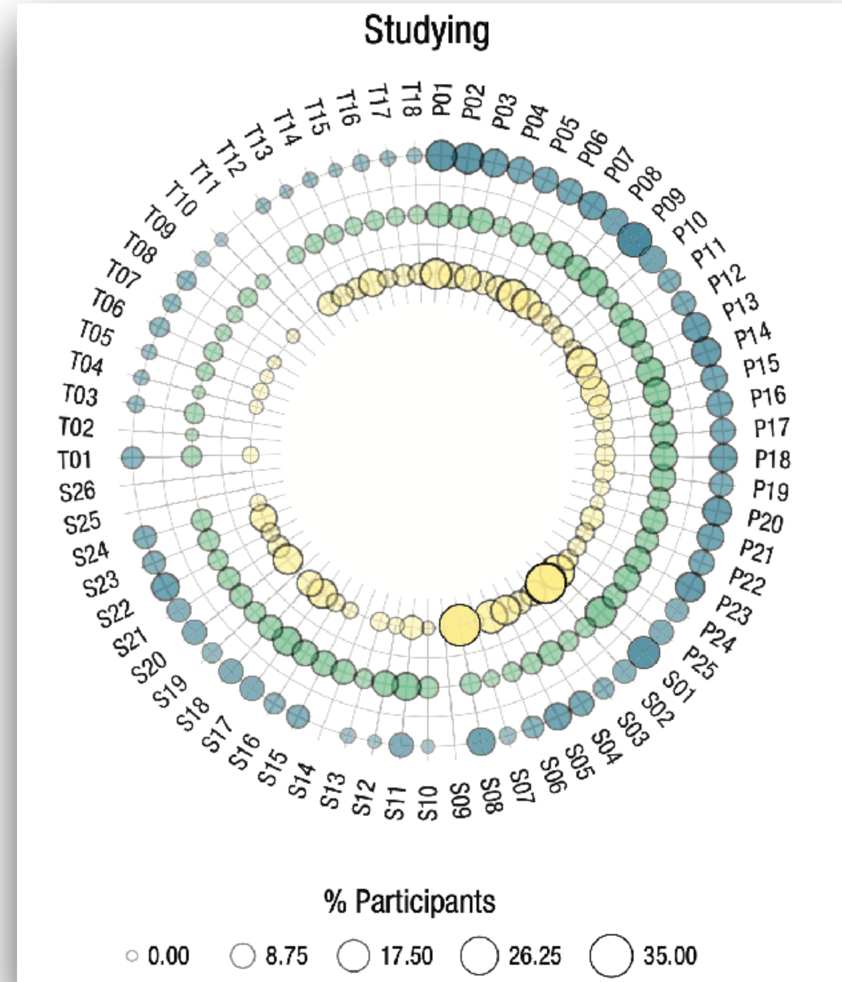
A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever

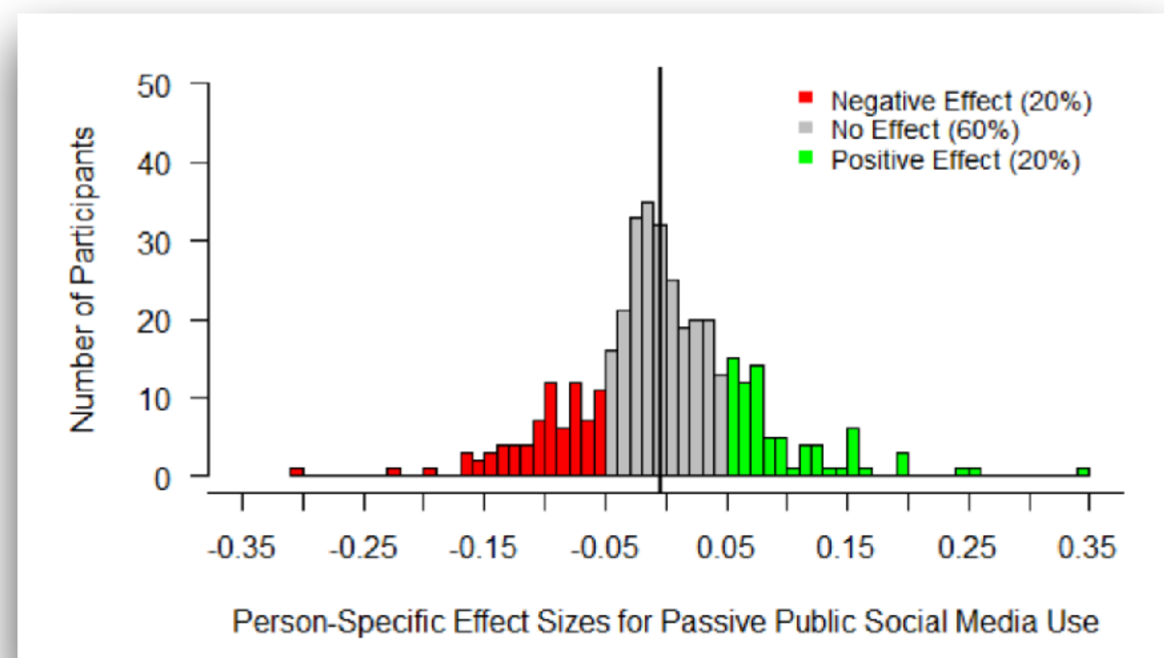
Peter C. M. Molenaar
Department of Psychology
University of Amsterdam

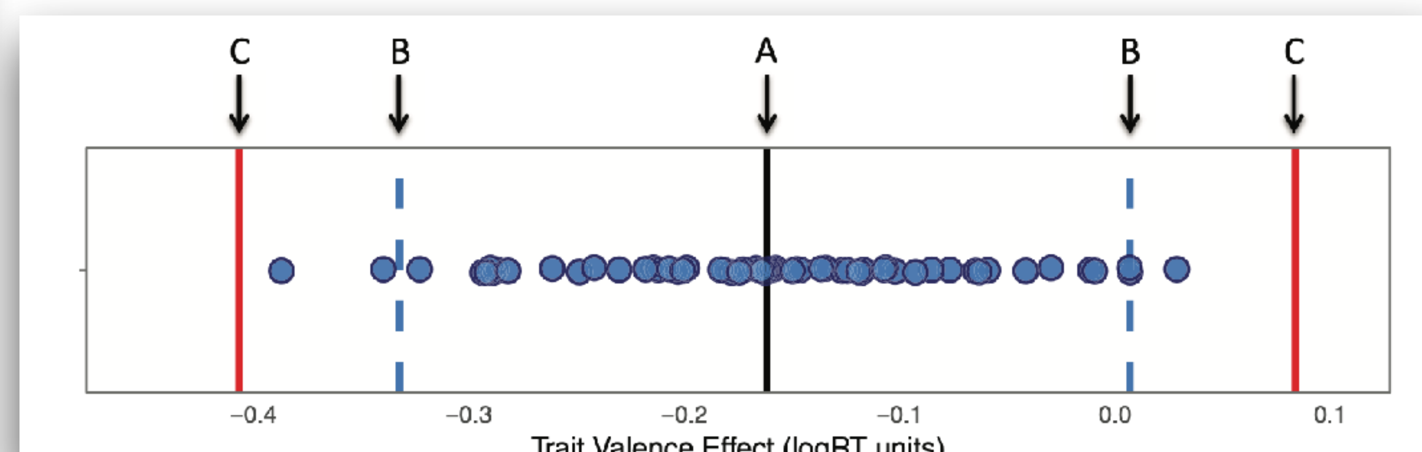
Beyond average treatment effects

How others have thought about it







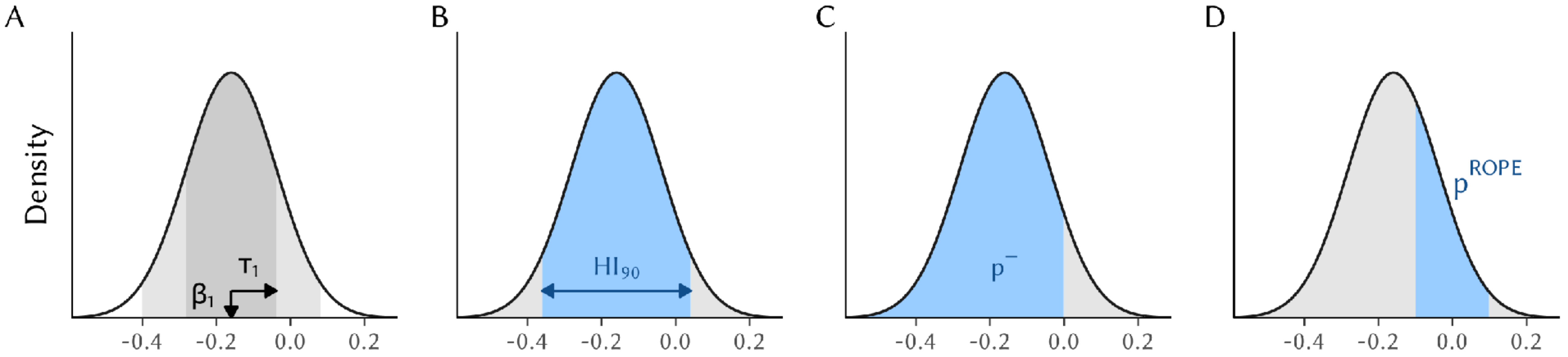


 Grice et al., 2020; Valkenburg et al., 2024; Beck & Jackson 2022; Bolger et al., 2019

Beyond average treatment effects

How we think about it

1. Heterogeneity is best thought of as a distribution of features with a mean (“average person”) and dispersion (person differences)
2. This distribution lends itself to informative metrics of heterogeneity
3. Uncertainty: bayes (and the $d \times p$ matrix)



Communicating heterogeneity

Example data and the valence effect in self-evaluations

Inflating and deflating the self: Sustaining motivational concerns through self-evaluation☆

Abigail A. Scholer [a,*], Yuka Ozaki [b], E. Tory Higgins [c]

[a] University of Waterloo, Canada
[b] Toyo University, Japan
[c] Columbia University, United States

Person	Trial	Valence	ln(RT)
01	7	Negative	6.93
01	8	Negative	6.61
01	9	Positive	7.11
01	15	Negative	6.81
01	18	Negative	6.79
01	23	Negative	6.61

$$
\begin{aligned}
\mathrm{logRT}_{ij} &\sim \mathrm{Normal}\left(\eta_{ij}, \sigma^2\right), \\
\eta_{ij} &= \beta_0 + \gamma_{0j} + \left(\beta_1 + \gamma_{1j}\right) \mathrm{V}_{ij}, \\
\begin{bmatrix} \gamma_0 \\ \gamma_1 \end{bmatrix} &\sim \mathrm{MVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{pmatrix} \tau_0 & \\ \rho & \tau_1 \end{pmatrix} \right).
\end{aligned}
$$

Communicating heterogeneity

Example data and the valence effect in self-evaluations

Person	Trial	Valence	ln(RT)
01	7	Negative	6.93
01	8	Negative	6.61
01	9	Positive	7.11
01	15	Negative	6.81
01	18	Negative	6.79
01	23	Negative	6.61

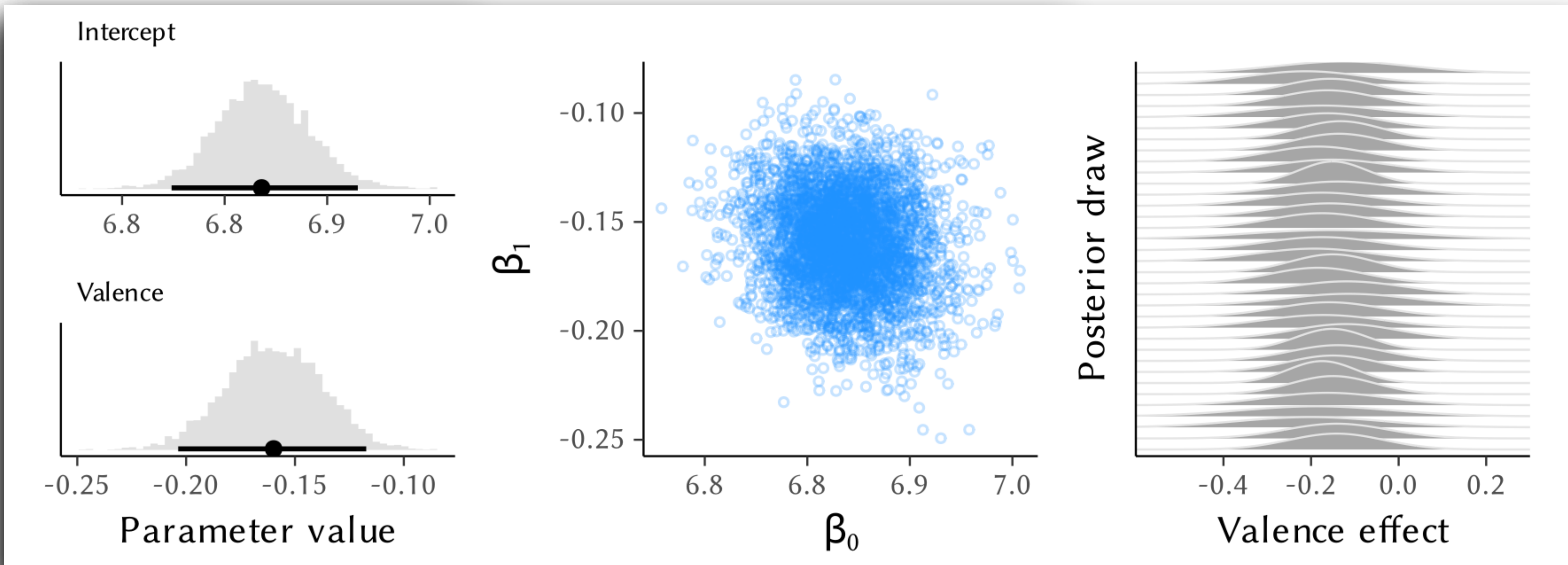
$$
\begin{aligned}
\text{logRT}_{ij} &\sim \text{Normal}(\eta_{ij}, \sigma^2), \\
\eta_{ij} &= \beta_0 + \gamma_{0j} + (\beta_1 + \gamma_{1j})\, \text{V}_{ij}, \\
\begin{bmatrix} \gamma_0 \\ \gamma_1 \end{bmatrix} &\sim \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{pmatrix} \tau_0 & \\ \rho & \tau_1 \end{pmatrix}\right).
\end{aligned}
$$

β_1	τ_1	$p(\gamma_1)$
-0.14	0.118	N(-0.14, 0.014)
-0.18	0.133	N(-0.18, 0.018)
-0.18	0.148	N(-0.18, 0.022)
-0.17	0.098	N(-0.17, 0.0097)
-0.14	0.104	N(-0.14, 0.011)
-0.17	0.147	N(-0.17, 0.022)

 Scholer, Ozaki, & Higgins (2014)

Communicating heterogeneity

Example data and the valence effect in self-evaluations



Communicating heterogeneity

Heterogeneity interval

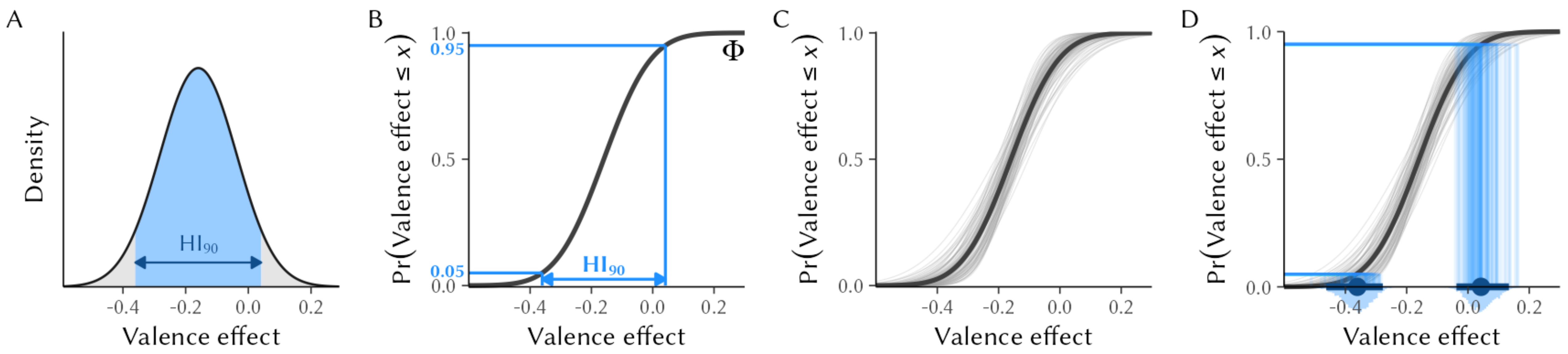
β_1	τ_1	$p(\gamma_1)$	HI_{90}
-0.14	0.118	N(-0.14, 0.014)	[-0.33, 0.05]
-0.18	0.133	N(-0.18, 0.018)	[-0.40, 0.04]
-0.18	0.148	N(-0.18, 0.022)	[-0.42, 0.06]
-0.17	0.098	N(-0.17, 0.0097)	[-0.33, 0.00]
-0.14	0.104	N(-0.14, 0.011)	[-0.31, 0.03]
-0.17	0.147	N(-0.17, 0.022)	[-0.41, 0.07]

I. $HI_{90} = \Phi^{-1}([0.05, 0.95], \beta_1, \tau_1)$

II. Calculate for all posterior draws

III. Summarize (e.g. mean, [95%CI])

1. −0.36 [−0.46, −0.28]
2. 0.04 [−0.04, 0.13]



Communicating heterogeneity

Population proportions: Prevalence

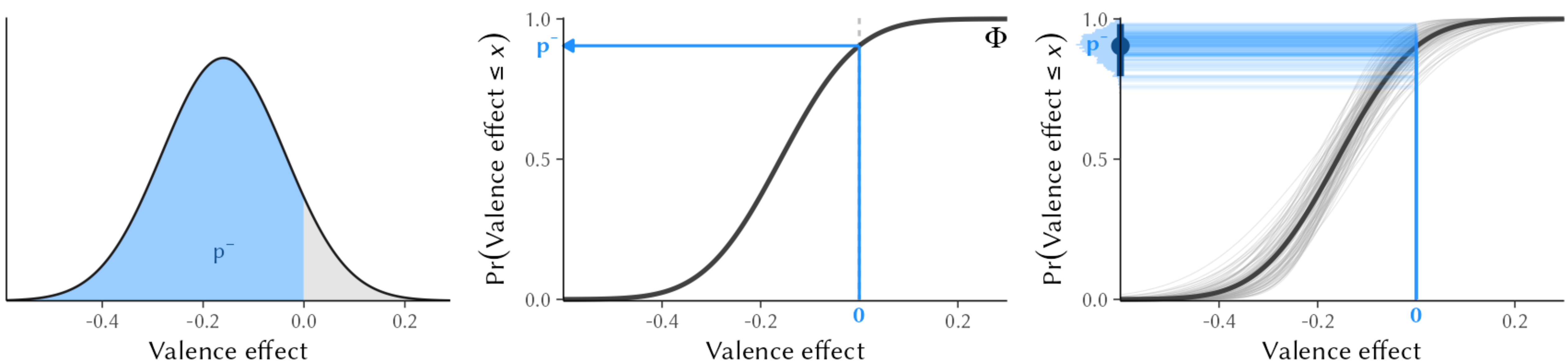
β_1	τ_1	$p(\gamma_1)$	p^+
-0.14	0.118	N(-0.14, 0.014)	0.88
-0.18	0.133	N(-0.18, 0.018)	0.91
-0.18	0.148	N(-0.18, 0.022)	0.89
-0.17	0.098	N(-0.17, 0.0097)	0.95
-0.14	0.104	N(-0.14, 0.011)	0.92
-0.17	0.147	N(-0.17, 0.022)	0.88

I. $p^- = Pr(VE \leq 0) = \Phi(0,\beta_1, \tau_1)$

II. Calculate for all posterior draws

III. Summarize (e.g. mean, [95%CI])

1. 89.9% [79.6%, 98.2%]



Communicating heterogeneity

Population proportions: Practically zero

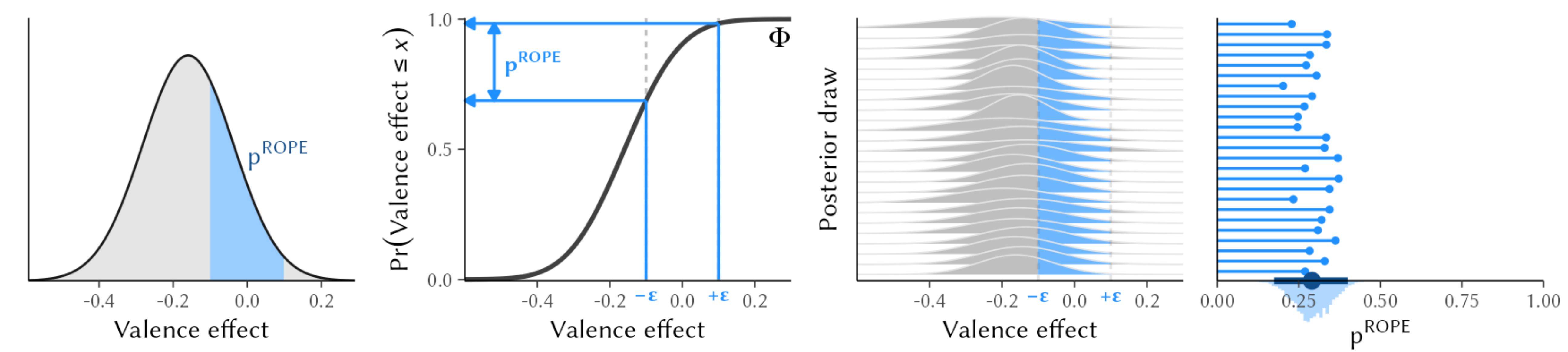
β_1	τ_1	$p(\gamma_1)$	p^{ROPE}
-0.14	0.118	N(-0.14, 0.014)	0.35
-0.18	0.133	N(-0.18, 0.018)	0.26
-0.18	0.148	N(-0.18, 0.022)	0.26
-0.17	0.098	N(-0.17, 0.0097)	0.25
-0.14	0.104	N(-0.14, 0.011)	0.33
-0.17	0.147	N(-0.17, 0.022)	0.29

I. $p^{ROPE} = \Phi(0.1; \beta_1, \tau_1) - \Phi(-0.1; \beta_1, \tau_1)$

II. Calculate for all posterior draws

III. Summarize (e.g. mean, [95%CI])

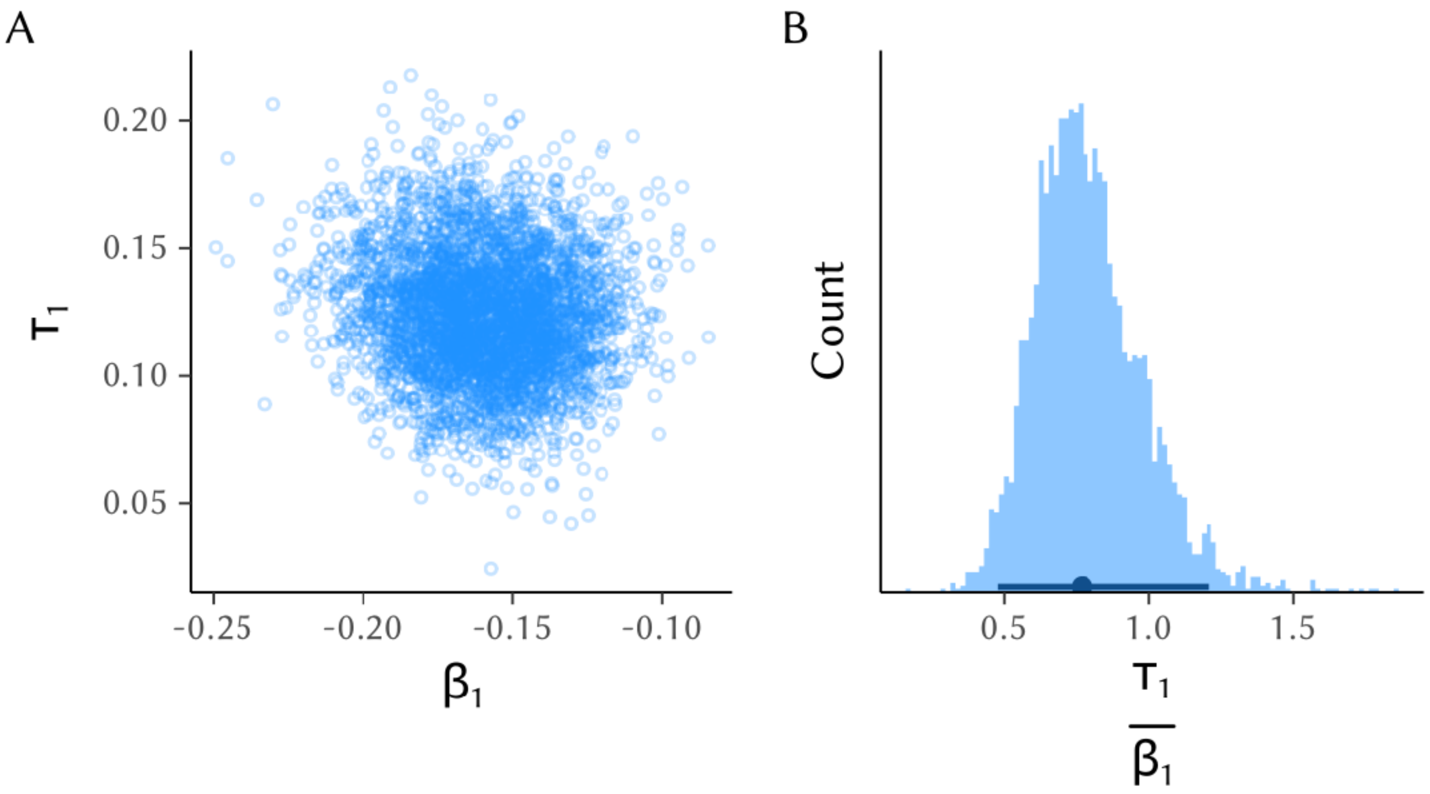
1. 29% [17.4%, 40%]



Communicating heterogeneity

Relative heterogeneity

β_1	τ_1	$p(\gamma_1)$	$\frac{\tau_1}{\beta_1}$
-0.14	0.118	N(-0.14, 0.014)	-0.85
-0.18	0.133	N(-0.18, 0.018)	-0.75
-0.18	0.148	N(-0.18, 0.022)	-0.81
-0.17	0.098	N(-0.17, 0.0097)	-0.59
-0.14	0.104	N(-0.14, 0.011)	-0.72
-0.17	0.147	N(-0.17, 0.022)	-0.87



Comparing heterogeneity

Comparing heterogeneity

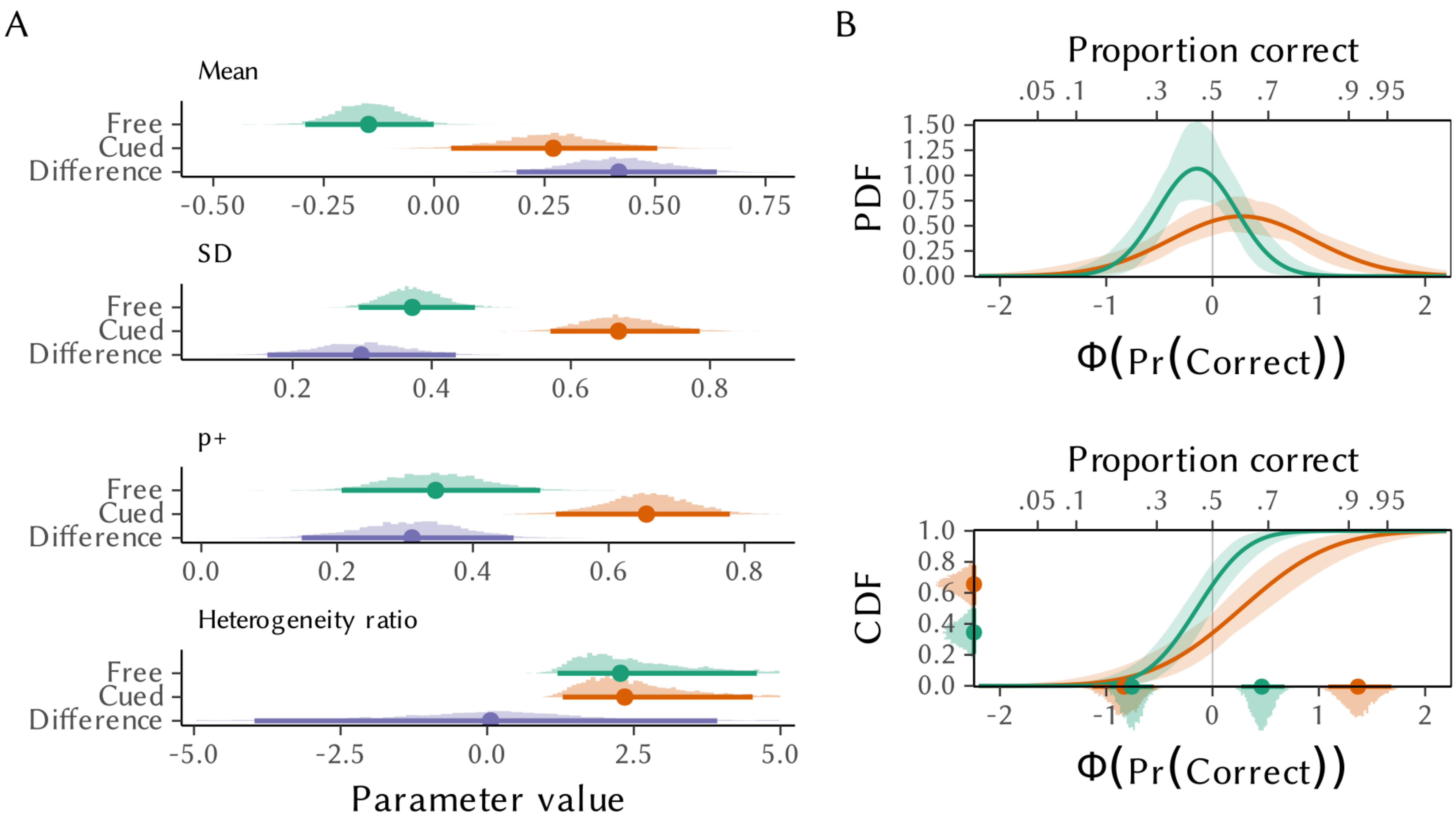
Recall performance example

1. People recall better when cued with related information
2. Mah and Lindsay compared recall performance under free recall (hard) and cued recall (easier)
3. To what extent is memory performance more variable between people in the cued recall task compared to the free recall task? (✅)
4. To what extent is memory performance more variable between target words in cued-versus-free recall tasks?
5. How consistent is target word heterogeneity across the two tasks: Are target words associated with good memory performance in cued recall experiments the same words that are associated with good memory performance in free recall experiments?

 Mah & Lindsay (2024)

Comparing heterogeneity

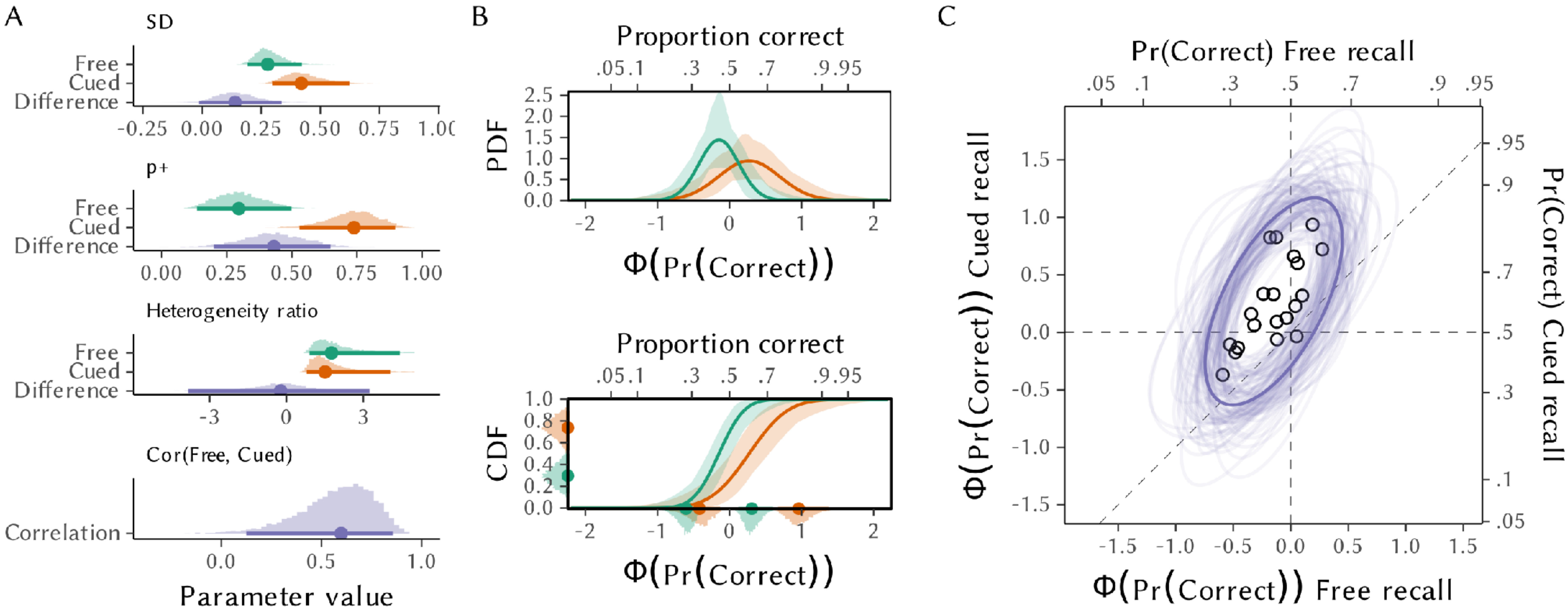
Between two populations of people



 Mah & Lindsay (2024)

Comparing heterogeneity

Between one population of items in two different conditions



 Mah & Lindsay (2024)

So what

Discussion

Multilevel models are ~~un~~necessary

ELSEVIER

Current Research in Neurobiology

Volume 2, 2021, 100024

Unnecessary reliance on multilevel modelling to analyse nested data in neuroscience: When a traditional summary-statistics approach suffices

Carolyn Beth McNabb [a], Kou Murayama [a b c]

ELSEVIER

Current Research in Neurobiology

Volume 3, 2022, 100041

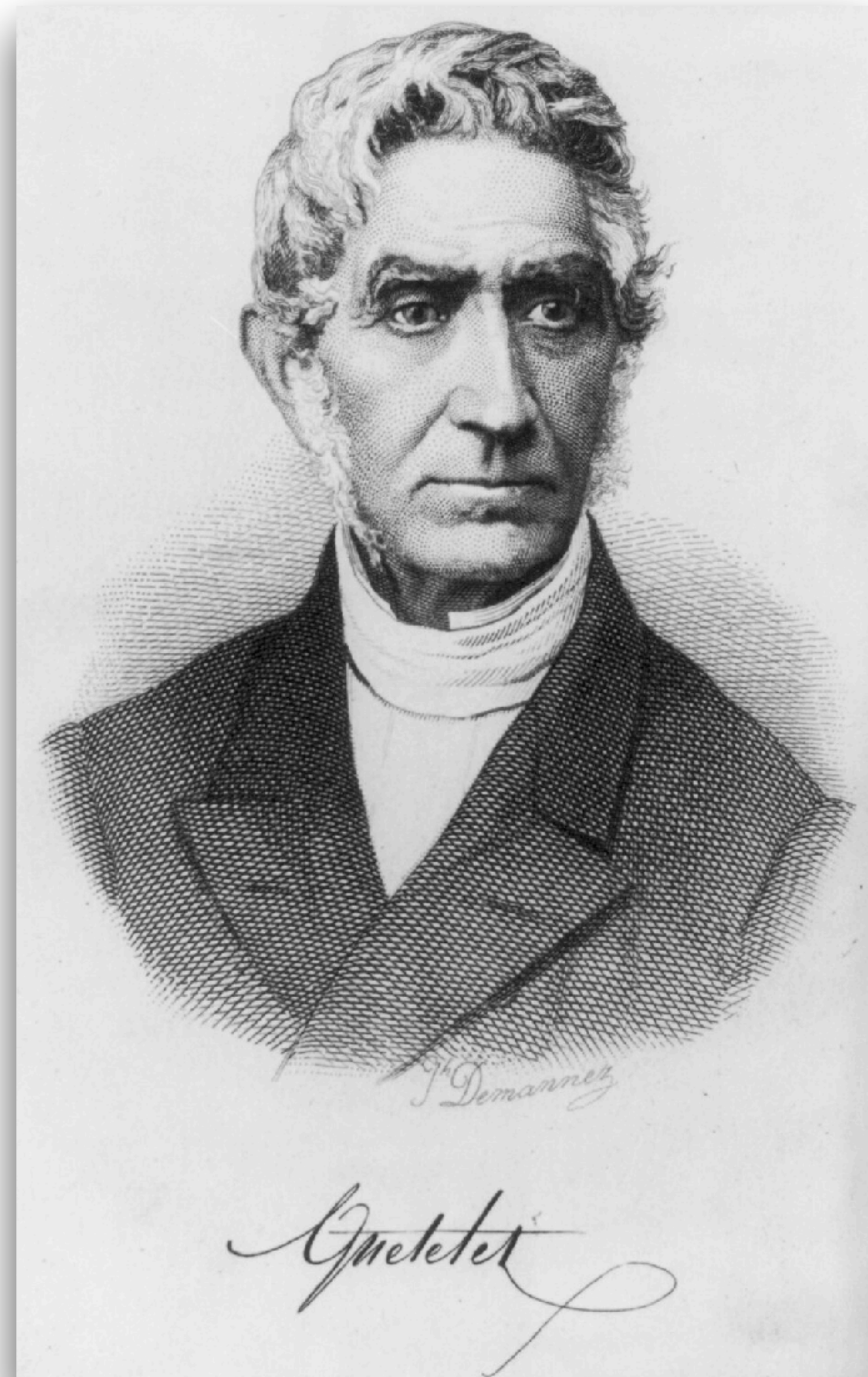
Commentary on Unnecessary reliance on multilevel modelling to analyse nested data in neuroscience: When a traditional summary-statistics approach suffices

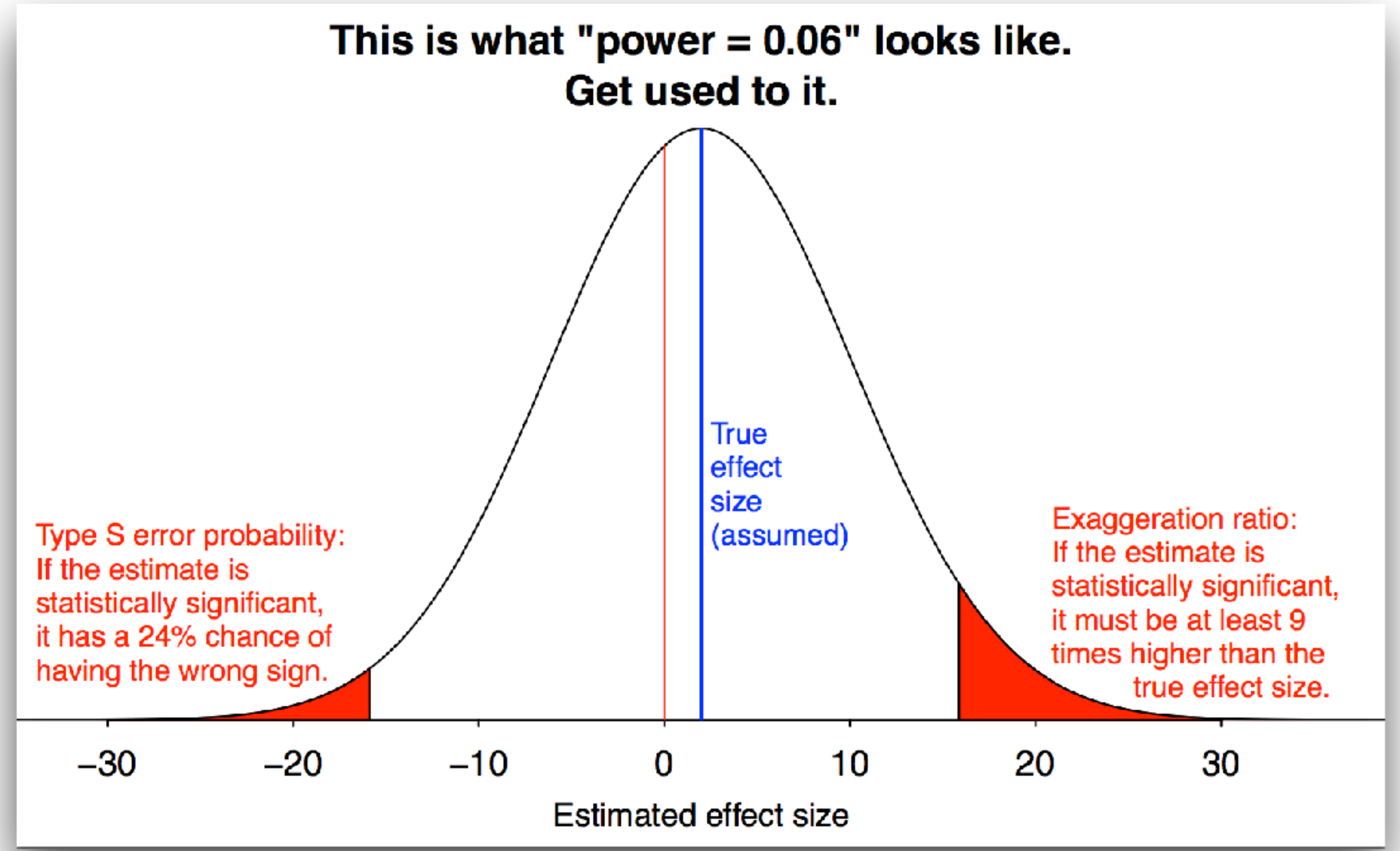
Paul Alexander Bloom PhD [1], Monica Kim Ngan Thieu PhD [1], Niall Bolger PhD

McNabb & Murayama (2021); Bloom et al. (2022)

Discussion

Quetelet's *l'homme moyen* with small averages and large heterogeneity





 Gelman (2014)

Communicating Causal Effect Heterogeneity

- **How to best represent, model, and communicate heterogeneity**
- **How I learned to stop worrying and love the multilevel model**
- **There's no matrix like the $d \times p$ matrix**

Matti Vuorre (Tilburg University)
https://vuorre.com
@matti.vuorre.com