Type M error in practice: A case study

Shravan Vasishth
Linguistics, Universität Potsdam, Germany

Daniela Mertzen, MSc
Linguistics
Universität Potsdam

Dr. Lena Jäger
Computer Science
Universität Potsdam

Prof. Andrew Gelman
Statistics
Columbia University

Type M error in practice: A case study

1. Power is quite low in reading research
2. Low power leads to exaggerated estimates
3. Published claims will not be replicable
4. We demonstrate this with real data

Type M error in practice: A case study

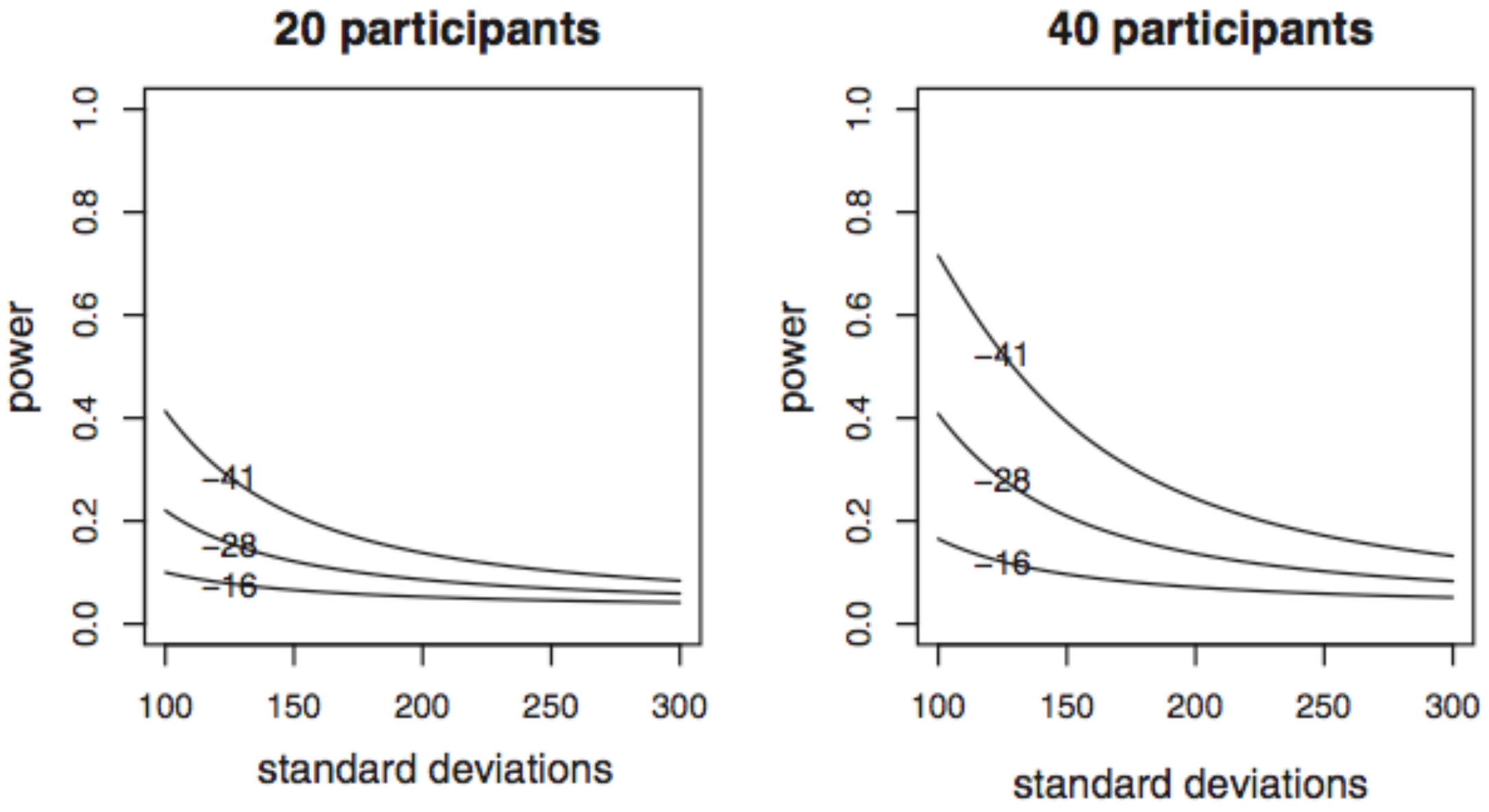
Research area: Reading processes in cognitive psychology

The Marketer: Alchemist, Magician, Sorcerer and Medicine Man

Is marketing an art or a science? Perhaps marketing is more like sorcery. Think of a sorcerer collecting ingredients from different sources and mixing them into a potion, accompanied with the magical effect of a flash of light and the illusion to follow. To some extent this fits with Cullitan's vision of a marketer as a 'mixer of ingredients'. Of course sorcerers are more mythical than real, but if we stay with this myth it may help to dispel some of the myths surrounding 'marketing'.

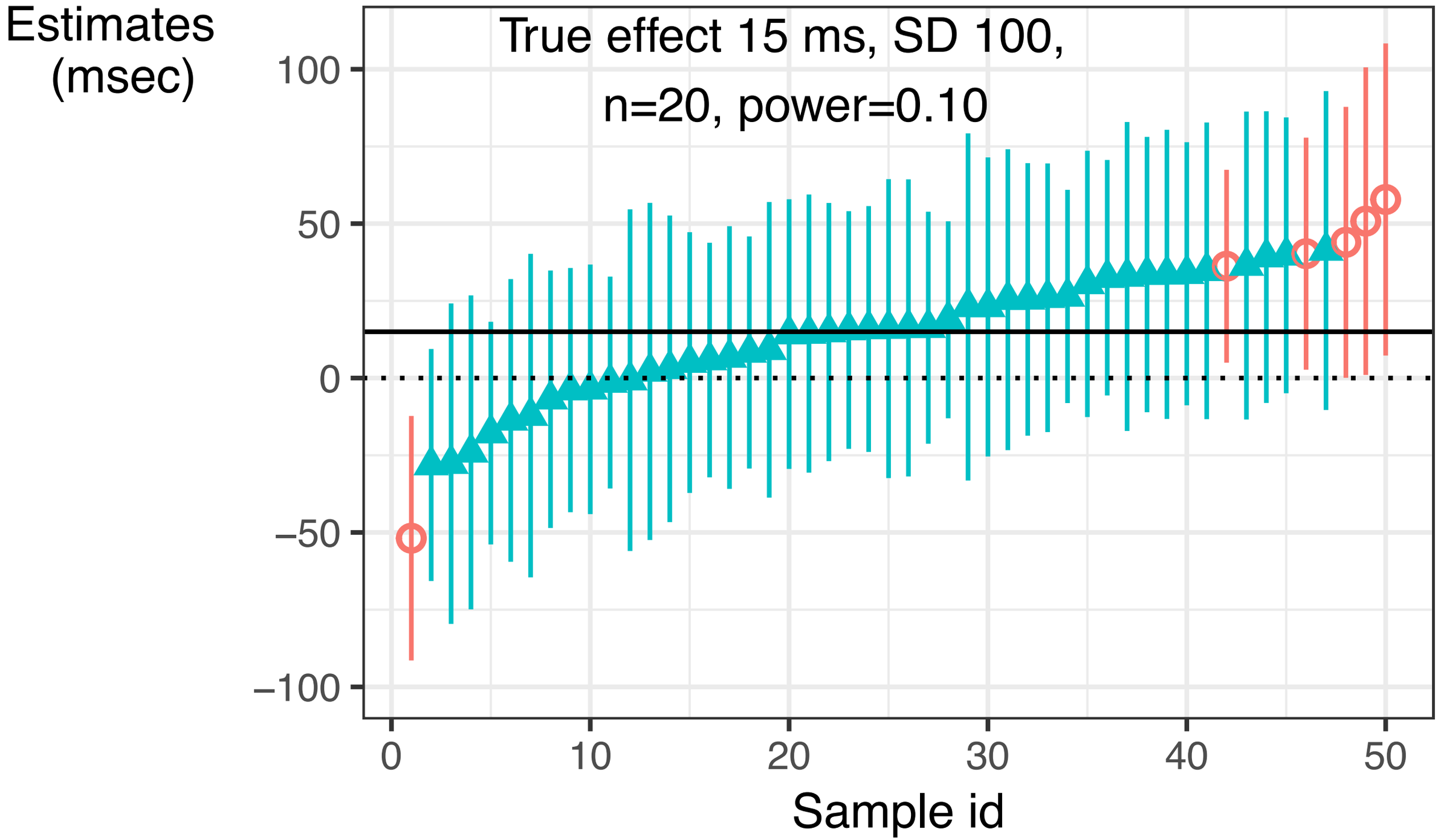
Though mythical, sorcerers were far from perfect. Not all their potions and spells succeeded. When they tried to cure diseases, the patient often died through severe poisoning -- and the fate of the sorcerer was anyone's guess. Perhaps the same could be said of alchemists. Alchemy was the medieval dream of using a 'philosopher's

Power is generally quite low in reading research



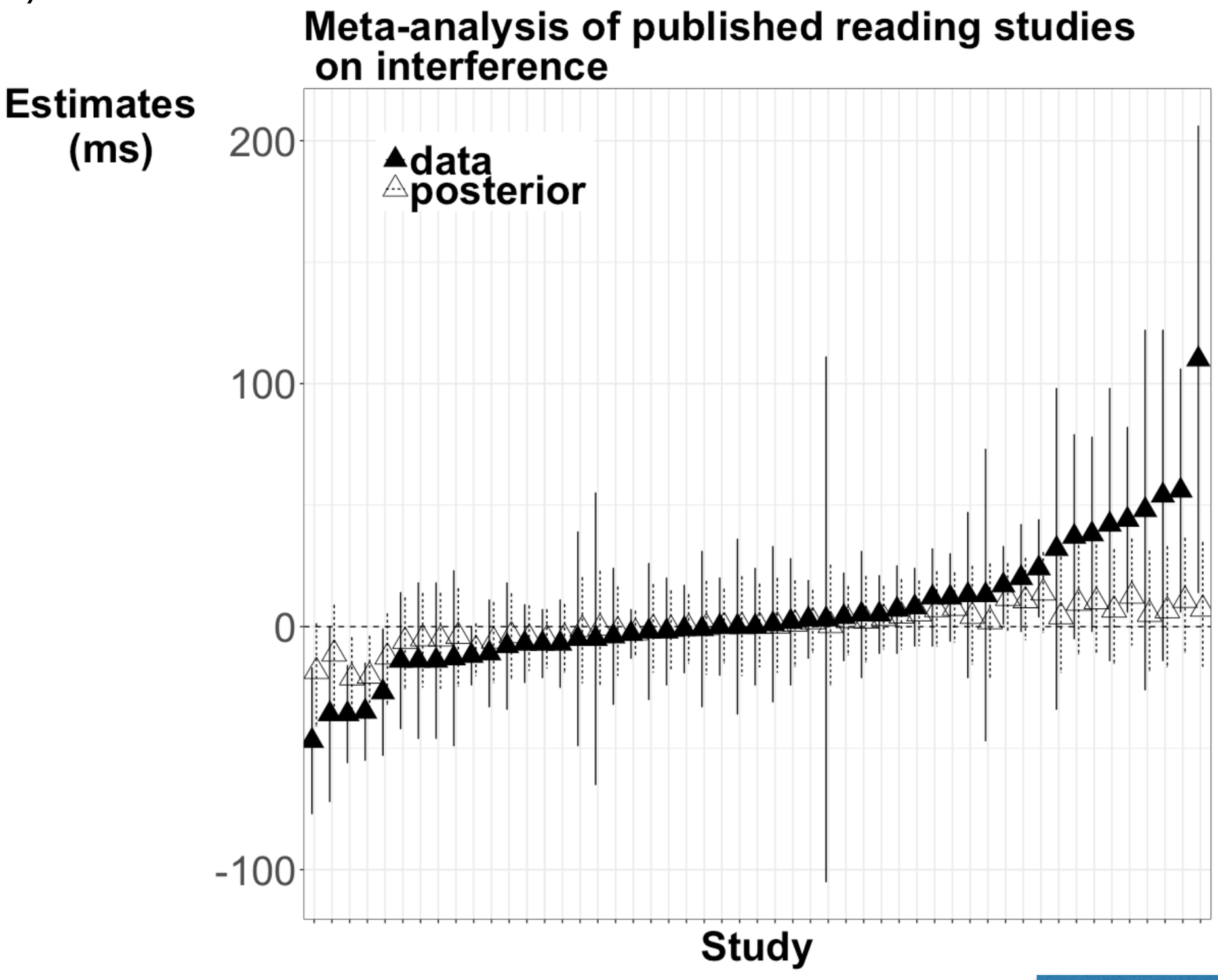
Jäger, Engelmann & Vasishth, 2017

Low power leads to exaggerated estimates: Type M error (simulated data)



Gelman & Carlin, 2014

Low power leads to exaggerated estimates: Type M error (published data)



Jäger, Engelmann & Vasishth, 2017

A puzzle: Most psychologists are aware of the replication crisis, but few think they are affected

Some frequent reactions:

- "In *our* field, we *always* replicate our results."
- "My *own* sub-field doesn't have problems."
- "We replicate, we just don't *publish* the data."

The first principle is that you must not fool yourself and you are the easiest person to fool.

Feynman

We demonstrate Type M error in published data

Journal of Memory and Language 68 (2013) 199–222

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Expectation and locality effects in German verb-final structures

Roger P. Levy [a,*], Frank Keller [b,1]

[a] Department of Linguistics, UC San Diego, 9500 Gilman Drive #0108, La Jolla, CA 92093-0108, USA
[b] School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, UK

We demonstrate Type M error in published data

The original eye tracking (reading) experiments:

- 2x2 repeated measures factorial design
  Two main effects and one interaction
- 28 subjects, 24 items, Latin square design
- Reading time in milliseconds

Seven replication attempts of Levy & Keller, 2013, using eyetracking and self-paced reading.

Self-paced reading

Self-paced reading

The _____ _____

Self-paced reading

____ boy ____

Four replication attempts

- Two self-paced reading studies, two eye tracking
- Prospective power for Levy and Keller experiments:

Effect (ms)	Power (percentage)
30	11
50	28
80	51

[Full details in paper: bit.ly/TypeMError]

Hierarchical linear models in Stan

i=1,…,I subjects
j=1,…,J items
n data points

$$\log rt = \underbrace{X\beta}_{\text{fixed effects}} + \underbrace{Z_u b_u}_{\text{subjects random effects}} + \underbrace{Z_w b_w}_{\text{items random effects}} + \varepsilon$$

$$X_{n\times p} = \begin{bmatrix} 1 & -1 & -1 & +1 \\ 1 & +1 & +1 & +1 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \qquad \beta_{p\times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \begin{matrix} \\ \text{Main Effect 1} \\ \text{Main Effect 2} \\ \text{Interaction} \end{matrix}$$

Hierarchical linear models in Stan

$$\log rt = \underbrace{X\beta}_{\text{fixed effects}} + \underbrace{Z_u b_u}_{\text{subjects random effects}} + \underbrace{Z_w b_w}_{\text{items random effects}} + \varepsilon$$

$$X_{n\times p} = \begin{bmatrix} 1 & -1 & -1 & +1 \\ 1 & +1 & +1 & +1 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} = Z_u = Z_w$$

Hierarchical linear models in Stan

$$\log rt = \underbrace{X\beta}_{\text{fixed effects}} + \underbrace{Z_u b_u}_{\text{subjects random effects}} + \underbrace{Z_w b_w}_{\text{items random effects}} + \varepsilon$$

Priors:

$$\beta_0 \sim Normal(0,10)$$

$$\beta_{1,2,3} \sim Normal(0,1)$$

$$\sigma \sim Normal_+(0,1)$$

$$\rho \sim LKJ(\nu = 2)$$

$$b_u \sim MVN_4(\mathbf{0}, \Sigma_u)$$

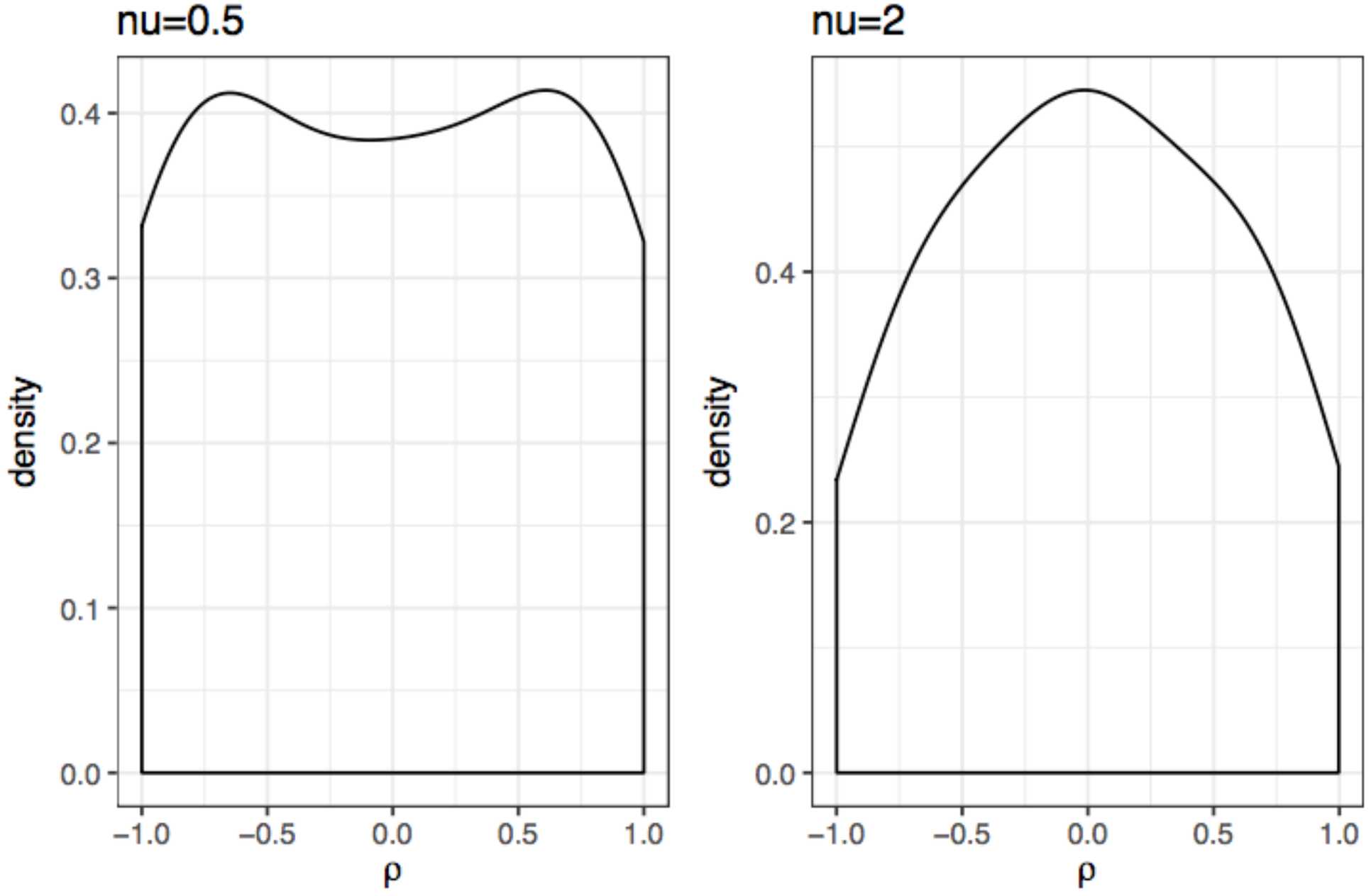
$$b_w \sim MVN_4(\mathbf{0}, \Sigma_w)$$

$$\varepsilon \sim Normal(0,\sigma)$$

Hierarchical linear models in Stan

Priors:

$$\rho \sim LKJ(2)$$



Hierarchical linear models in Stan

$$\log rt = \underbrace{X\beta}_{\text{fixed effects}} + \underbrace{Z_u b_u}_{\text{subjects random effects}} + \underbrace{Z_w b_w}_{\text{items random effects}} + \varepsilon$$

Priors:

$$\beta_0 \sim Normal(0,10)$$

$$\beta_{1,2,3} \sim Normal(0,1)$$

$$\sigma \sim Normal_+(0,1)$$

$$\rho \sim LKJ(\nu = 2)$$

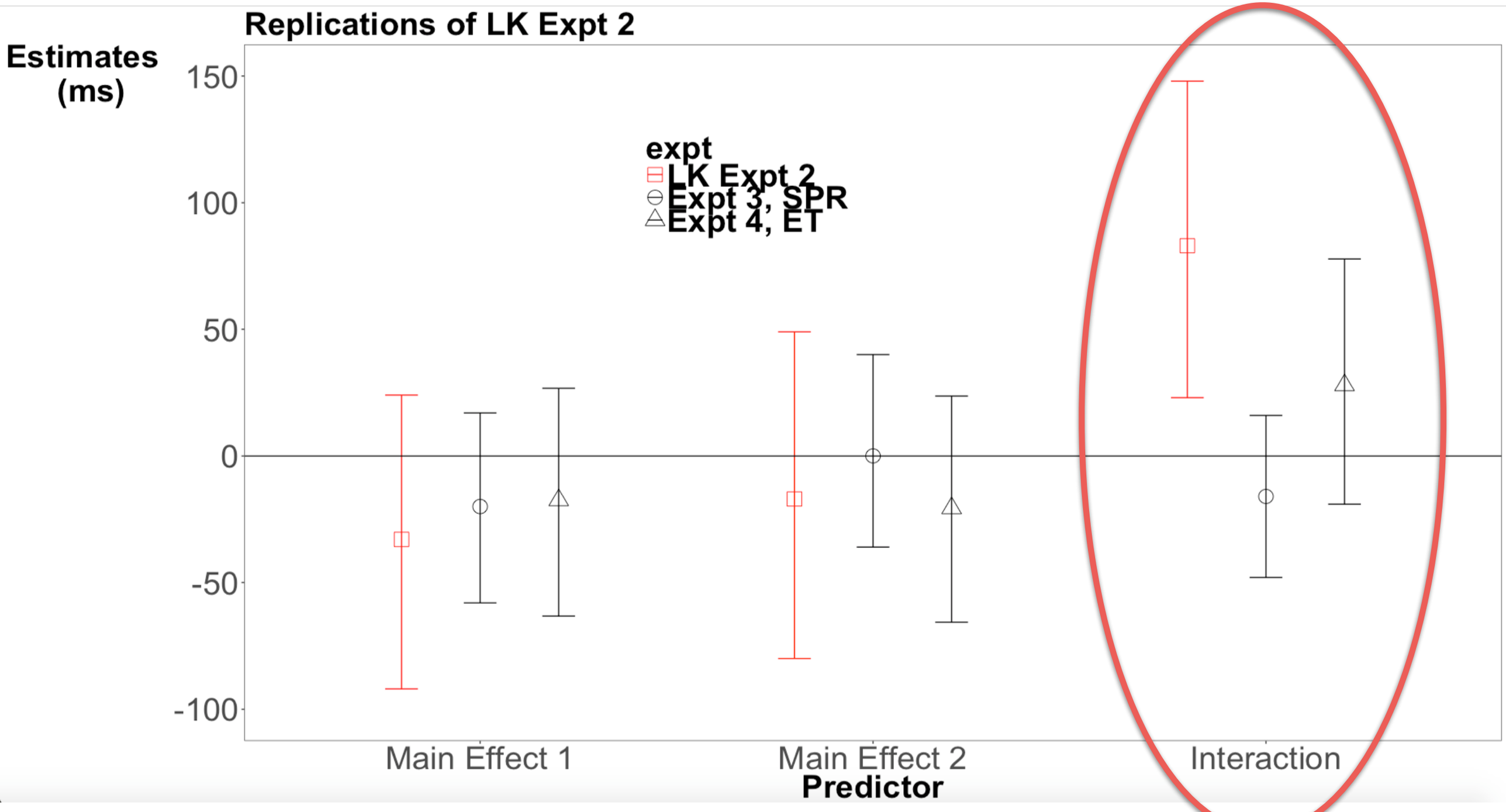
$$b_u \sim MVN_4(\mathbf{0}, \Sigma_u)$$

$$b_w \sim MVN_4(\mathbf{0}, \Sigma_w)$$

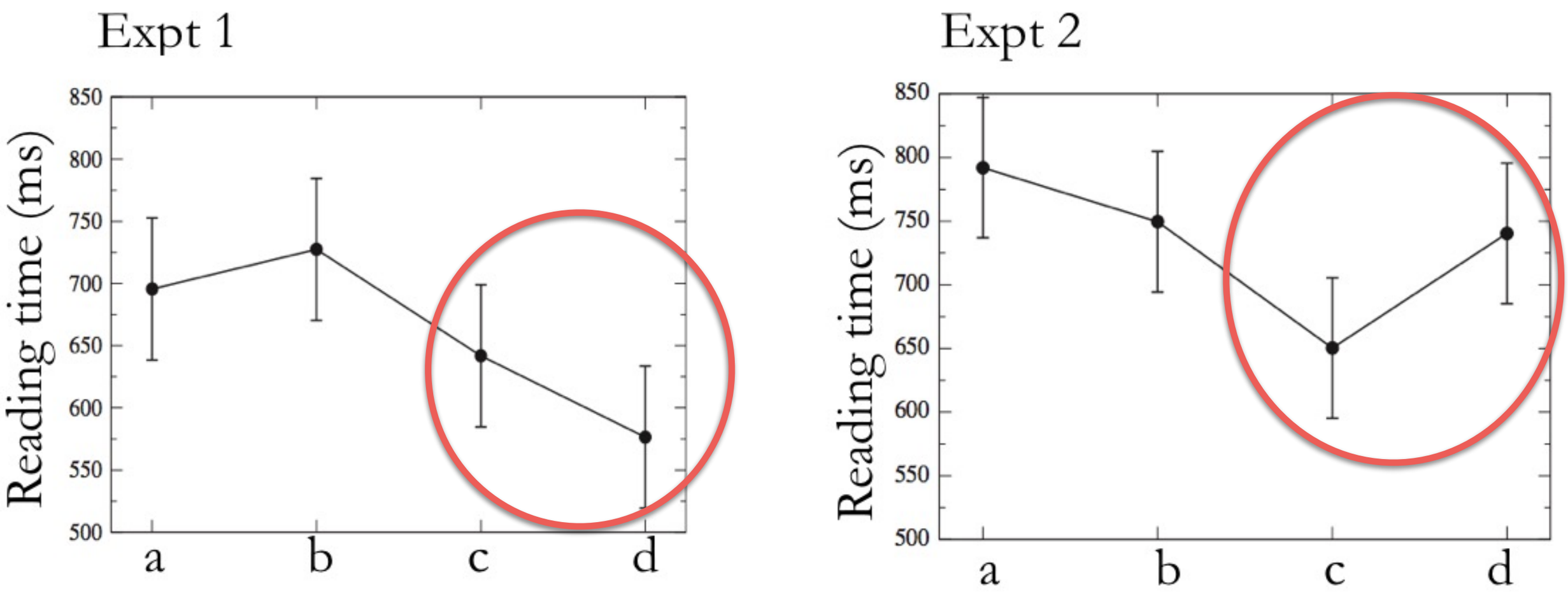
$$\varepsilon \sim Normal(0,\sigma)$$

Levy & Keller's Expt 1 replication attempts

Levy & Keller's Expt 2 replication attempts



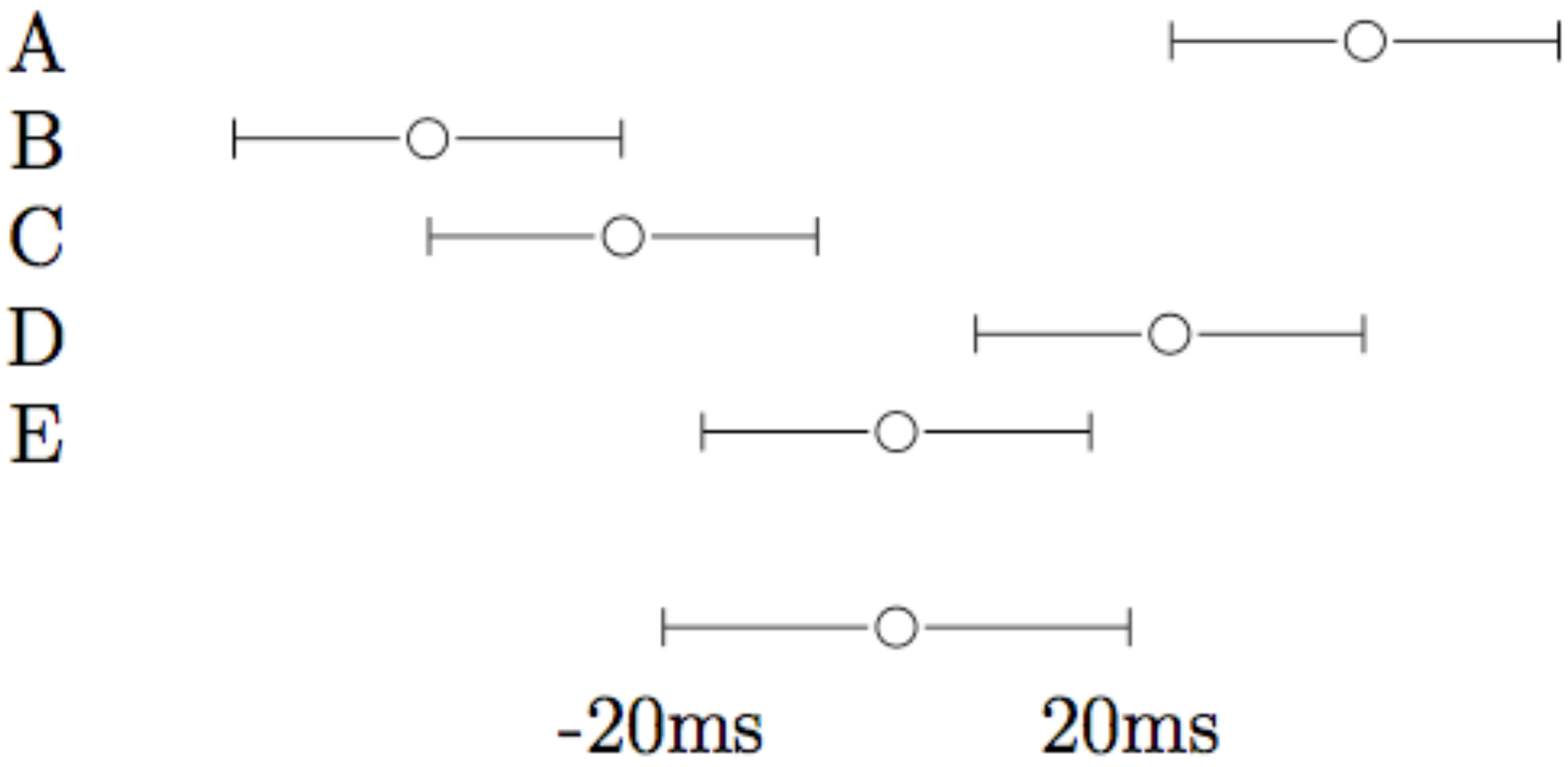
Levy & Keller 2013 claimed an interaction across the two experiments but never checked it statistically



Three replication attempts of the claimed interaction

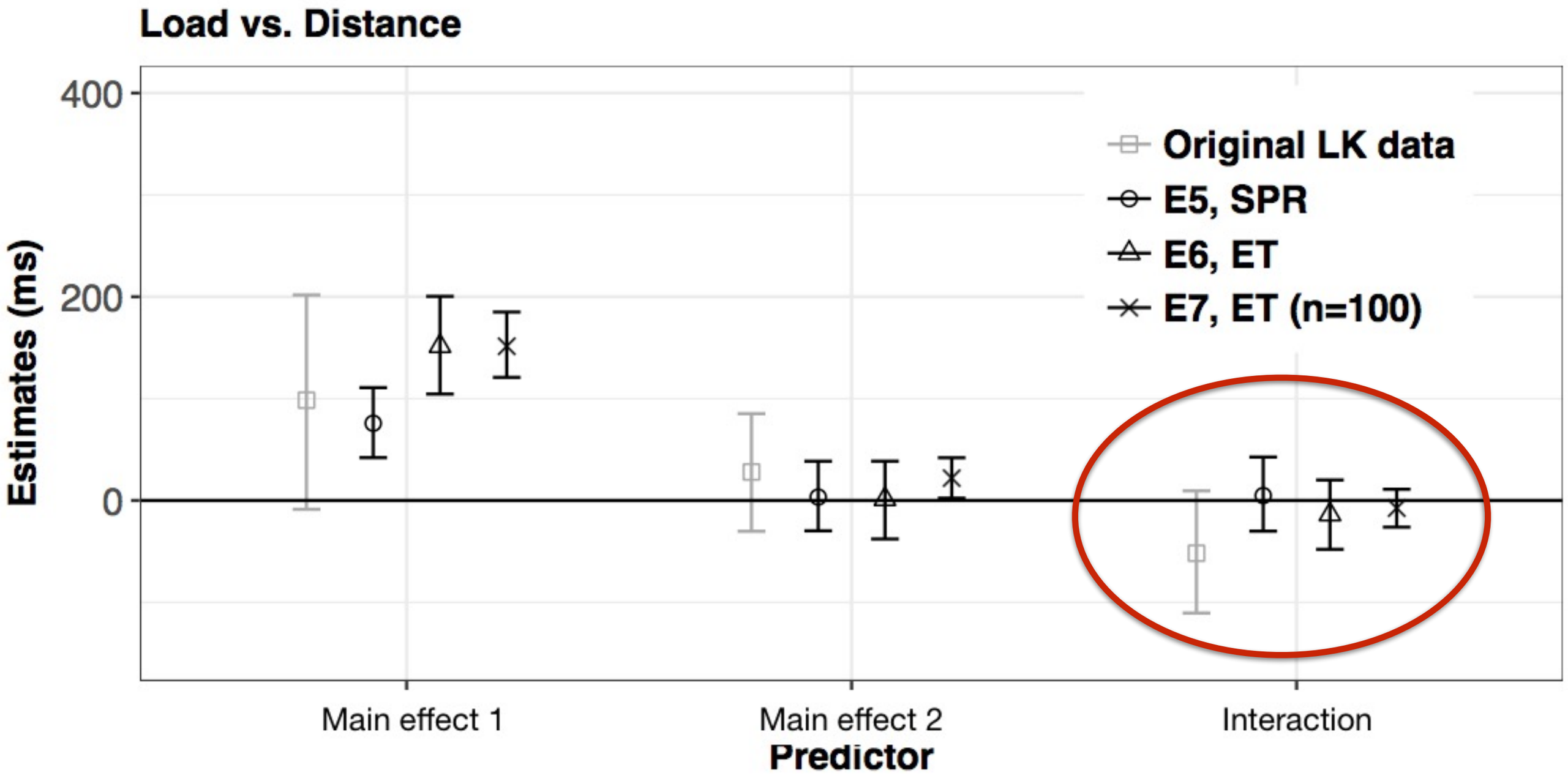
- Expt 5 (SPR): 28 participants, 24 items
- Expt 6 (ET): 28 participants, 24 items
- Expt 7 (ET): 100 participants, 24 items

Expt 7: Stopping rule determined by region of practical equivalence



Freedman, Lowe, Macaskill, 1984

Three replication attempts of the claimed interaction



Type M error in practice: A case study

Concluding remarks

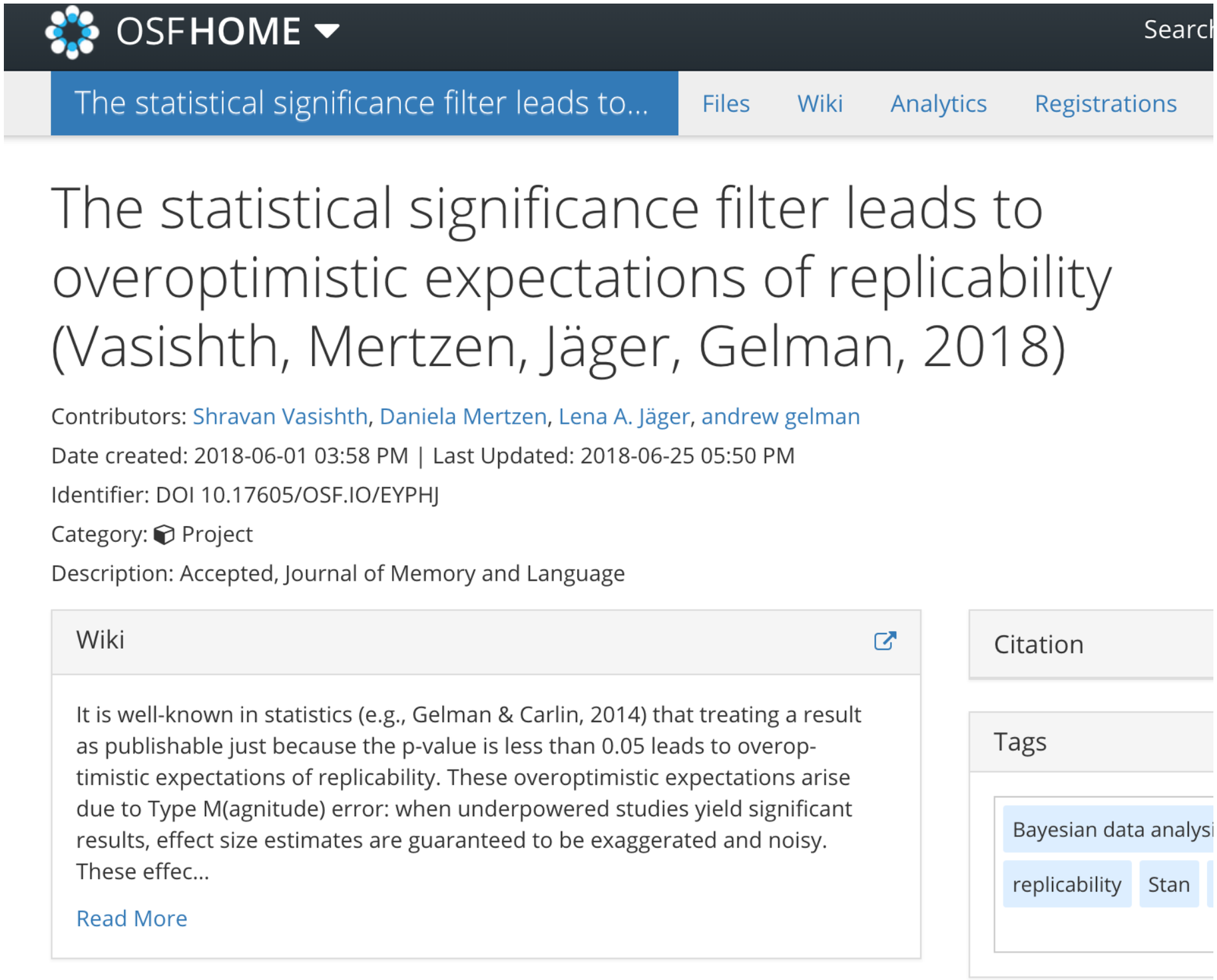
1. Expts with 268 subjects show not a single effect
2. The published effects are Type M errors
3. Many researchers still don't understand this point

Type M error in practice: A case study

Concluding remarks

1. Move focus away from significance
2. Focus instead on estimation
3. Run higher-precision studies
4. Pre-register experiments
5. Conduct direct replications

Type M error in practice: A case study



bit.ly/TypeMError