

# 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

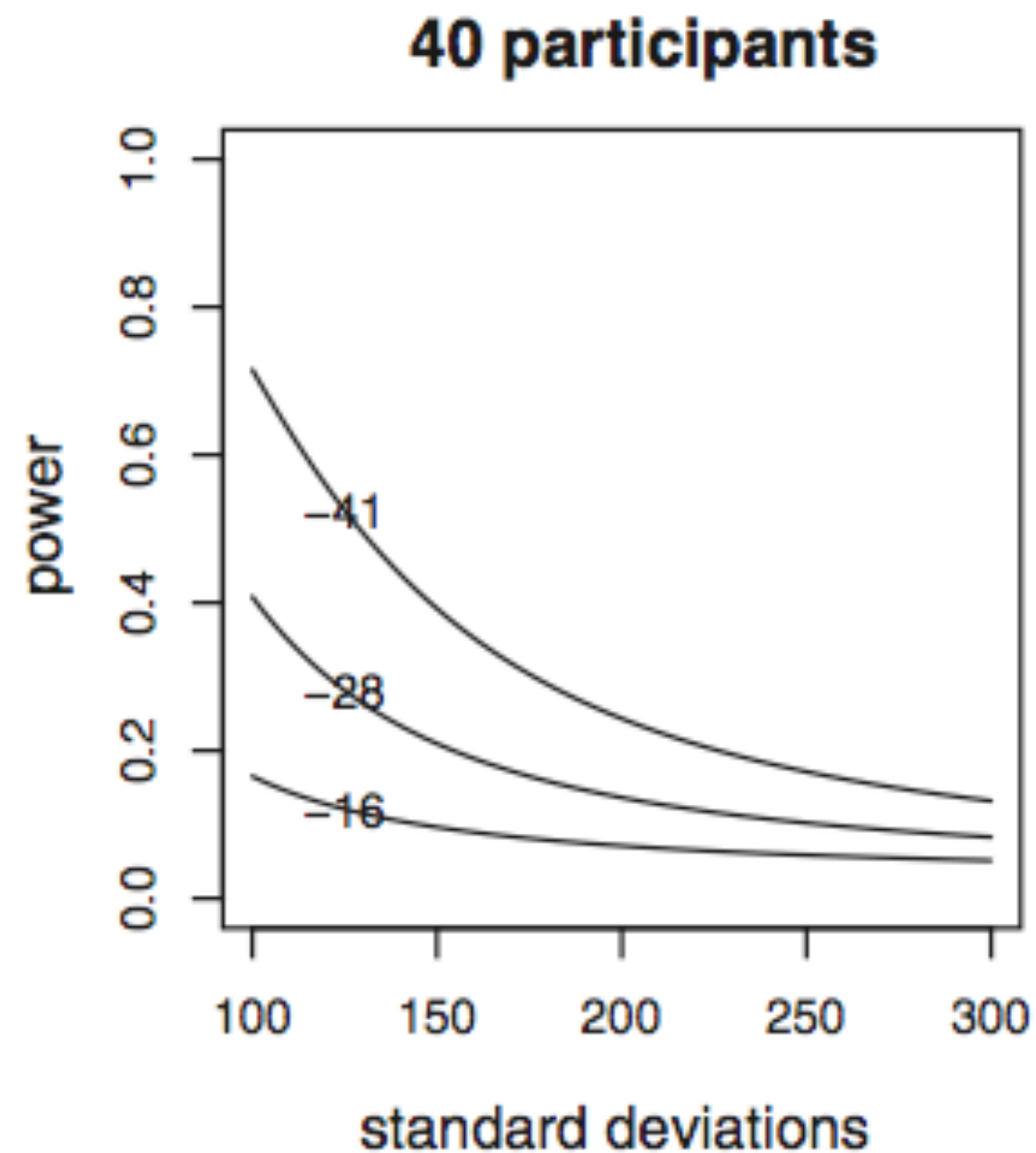
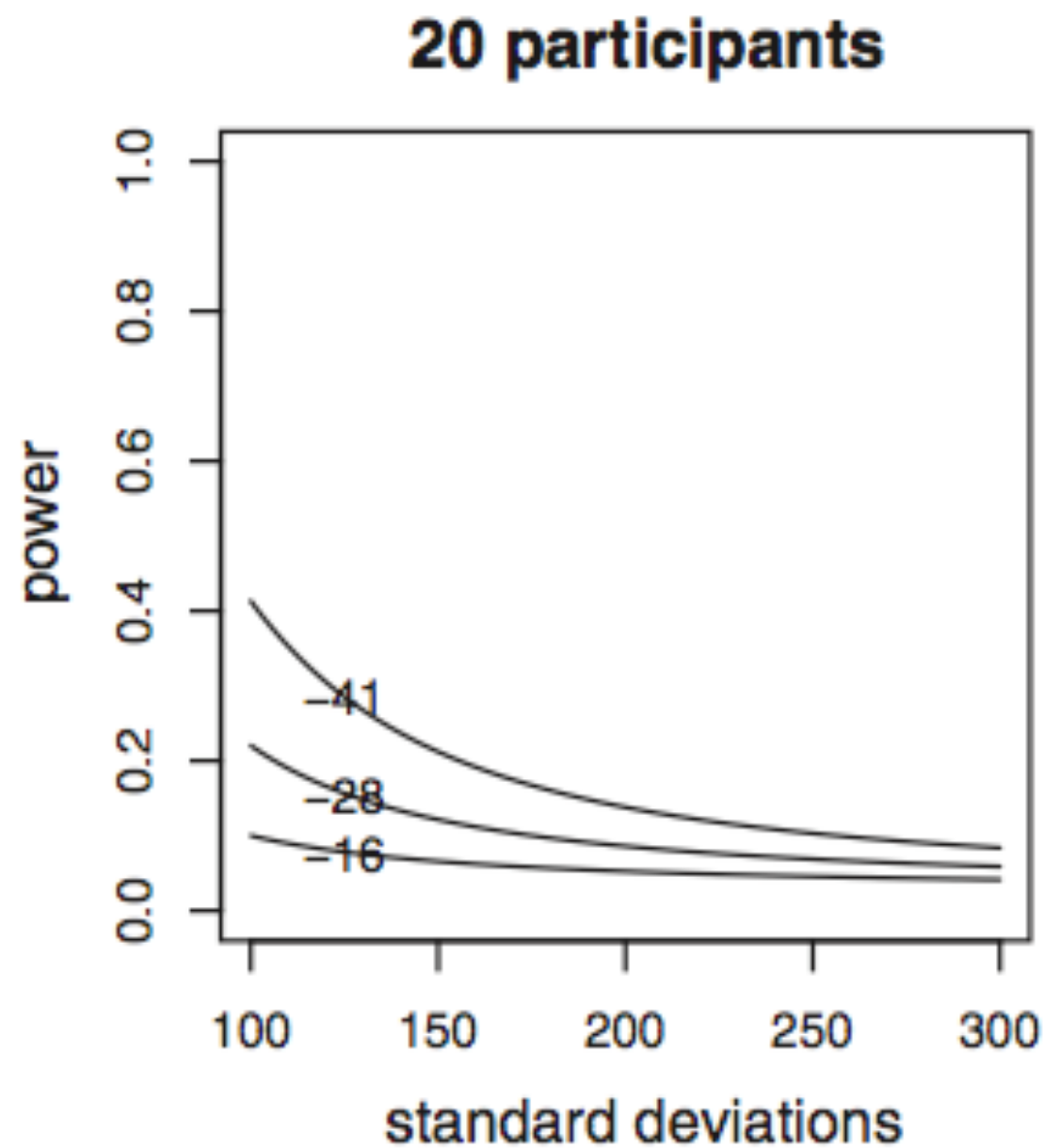
Research area: Reading processes in cognitive psychology

## 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 Culliton'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

# 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](http://www.elsevier.com/locate/jml)



## Expectation and locality effects in German verb-final structures

Roger P. Levy<sup>a,\*</sup>, Frank Keller<sup>b,1</sup>

<sup>a</sup> Department of Linguistics, UC San Diego, 9500 Gilman Drive #0108, La Jolla, CA 92093-0108, USA

<sup>b</sup> School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, UK

# 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](https://bit.ly/TypeMError)]



# 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:

$$\beta_0 \sim \text{Normal}(0, 10)$$

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

$$\sigma \sim \text{Normal}_+(0, 1)$$

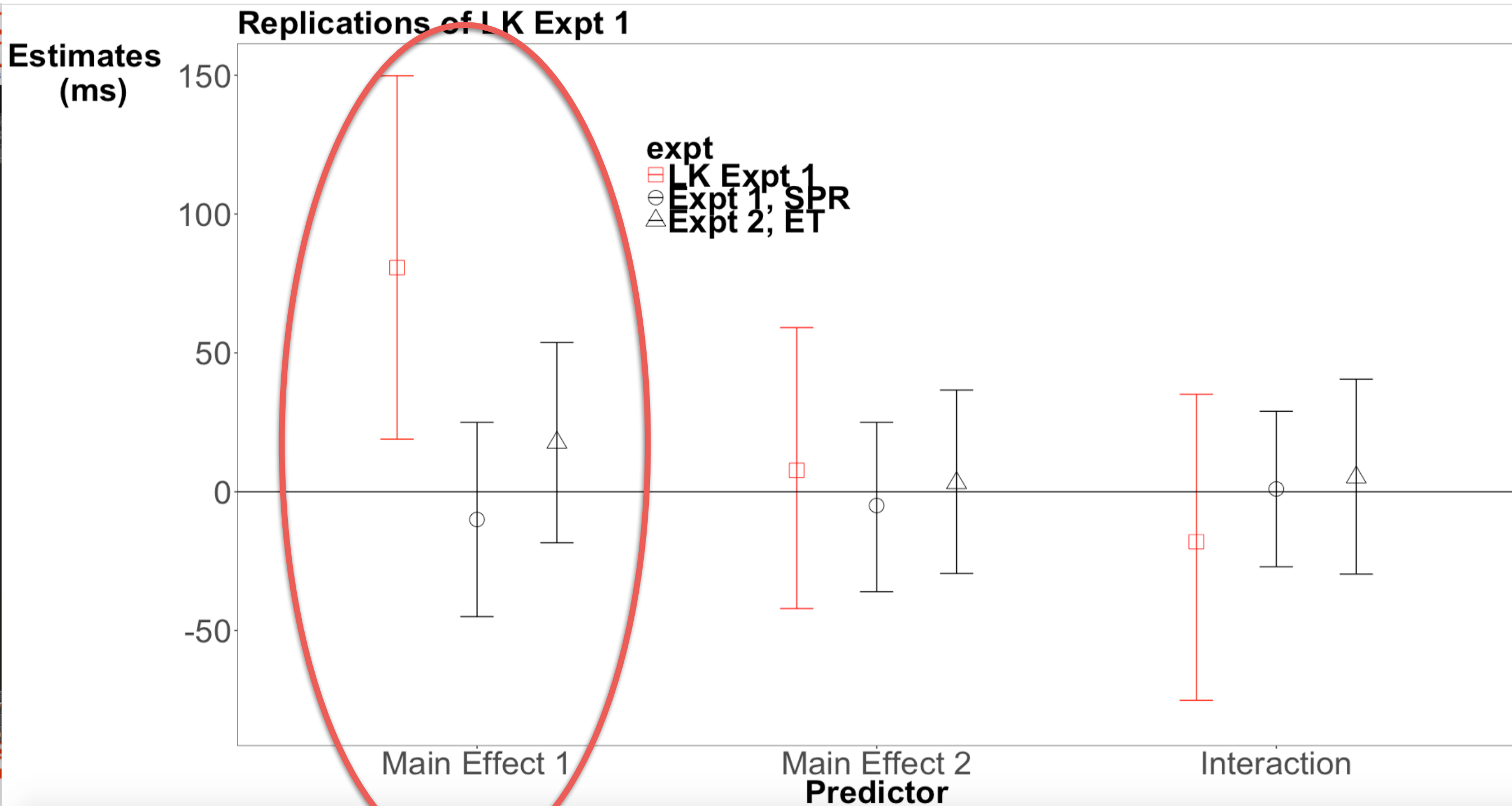
$$\rho \sim \text{LKJ}(\nu = 2)$$

$$b_u \sim \text{MVN}_4(\mathbf{0}, \Sigma_u)$$

$$b_w \sim \text{MVN}_4(\mathbf{0}, \Sigma_w)$$

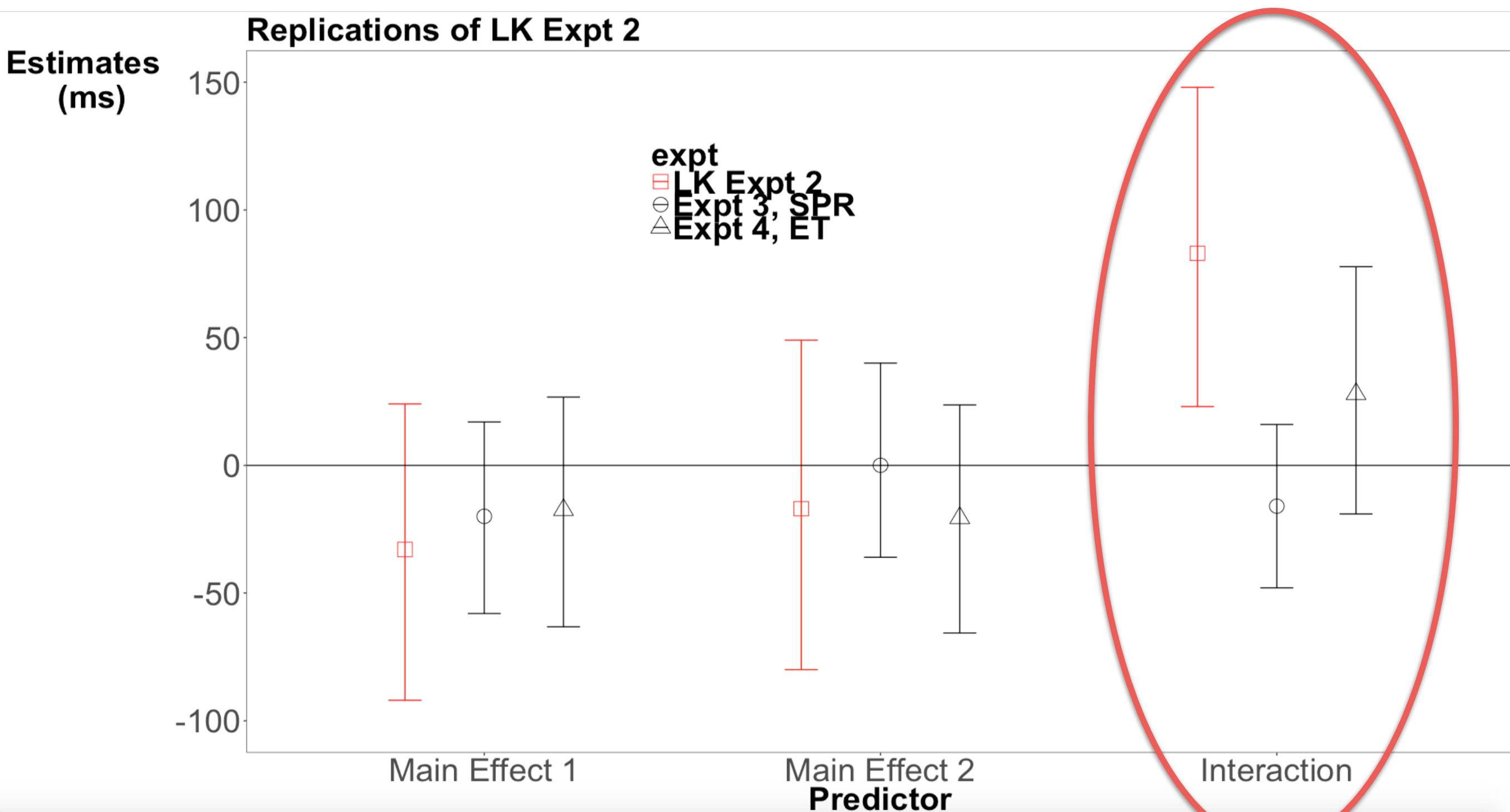
$$\varepsilon \sim \text{Normal}(0, \sigma)$$

# Levy & Keller's Expt 1 replication attempts



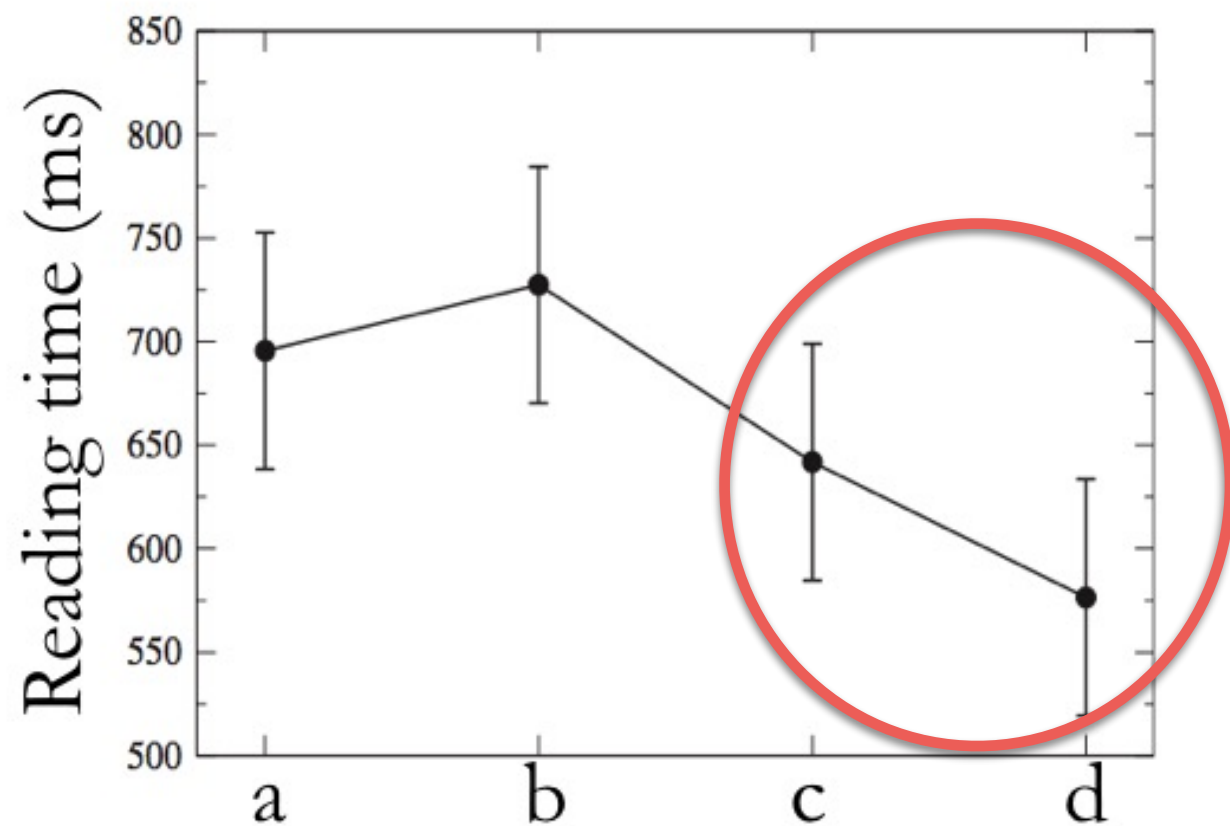


# Levy & Keller's Expt 2 replication attempts

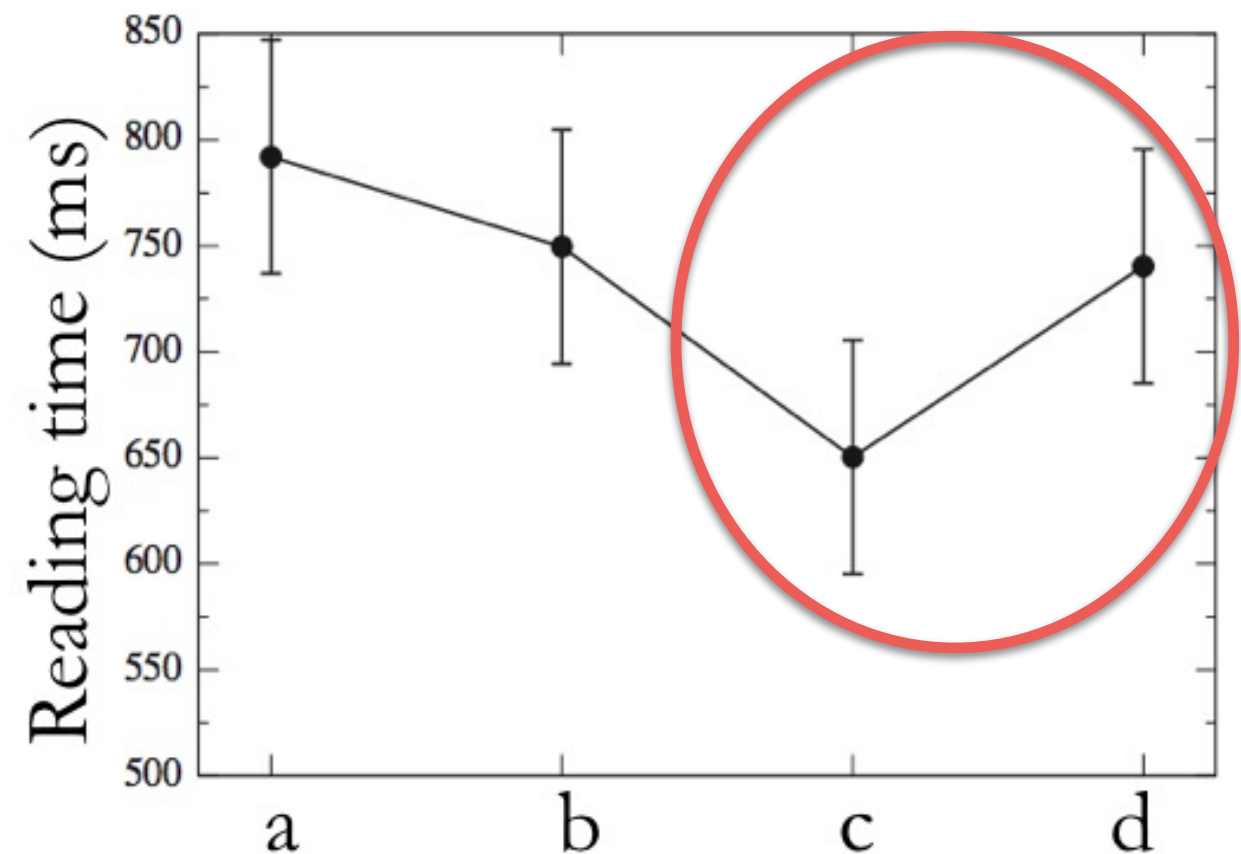


Levy & Keller 2013 claimed an interaction across the two experiments but never checked it statistically

Expt 1



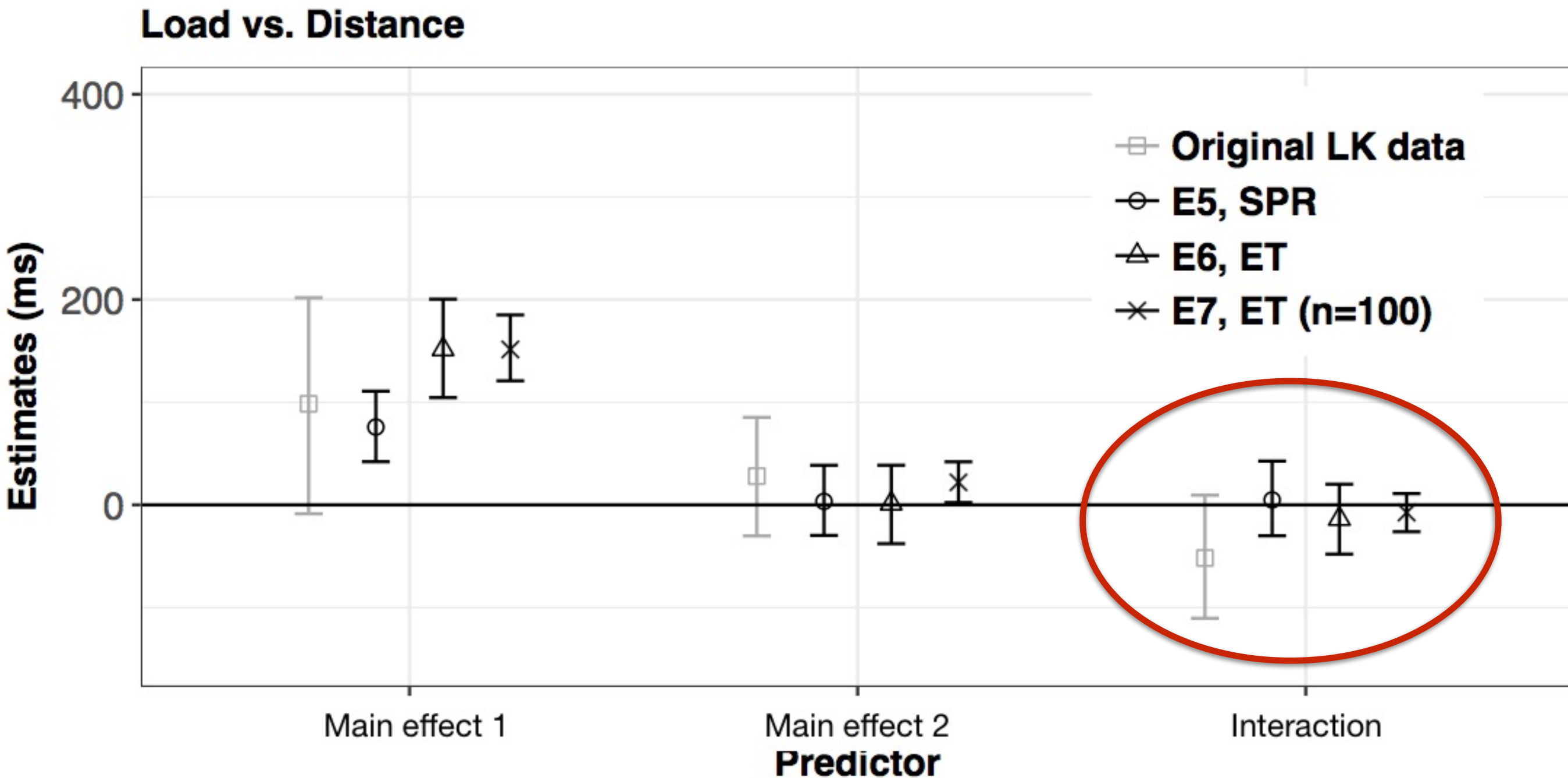
Expt 2



# 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

# Three replication attempts of the claimed interaction





# Type M error in practice: A case study

## Concluding remarks

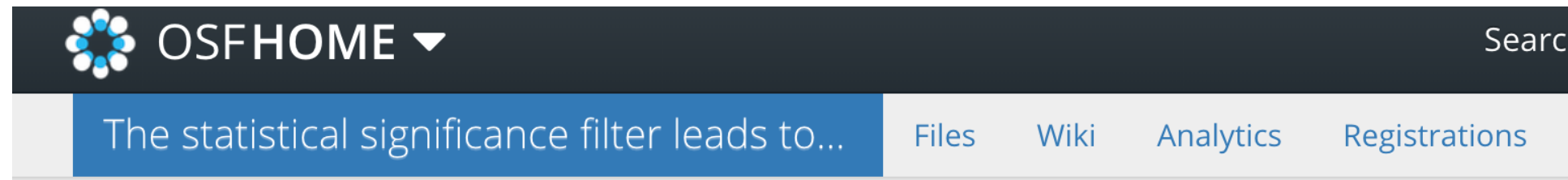
1. Expts with 268 subjects show not a single effect
2. The published effects are Type M errors
3. Many researchers still don't understand this point

# 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



## The statistical significance filter leads to overoptimistic expectations of replicability (Vasishth, Mertzen, Jäger, Gelman, 2018)

Contributors: [Shravan Vasishth](#), [Daniela Mertzen](#), [Lena A. Jäger](#), [andrew gelman](#)

Date created: 2018-06-01 03:58 PM | Last Updated: 2018-06-25 05:50 PM

Identifier: DOI 10.17605/OSF.IO/EYPHJ

Category:  Project

Description: Accepted, Journal of Memory and Language

### Wiki



It is well-known in statistics (e.g., Gelman & Carlin, 2014) that treating a result as publishable just because the p-value is less than 0.05 leads to overoptimistic expectations of replicability. These overoptimistic expectations arise due to Type M(agnitude) error: when underpowered studies yield significant results, effect size estimates are guaranteed to be exaggerated and noisy. These effec...

[Read More](#)

### Citation

### Tags

Bayesian data analysis

replicability

Stan

[bit.ly/TypeMError](https://bit.ly/TypeMError)