When less is (sometimes) more.

Evaluating the effect of trial number in classical experimental psychology paradigms

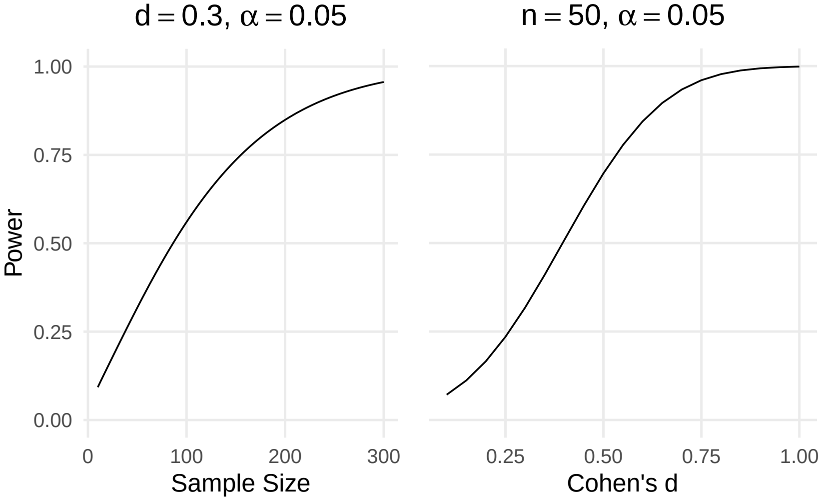
Filippo Gambarota
Giovanni Bruno, Roberta Sellaro & Simone Cutini

University of Padova

@AIP Sperimentale Torino 2025

The usual power analysis workflow

Nowadays, (fortunately), sample size justification using e.g. the power analysis is mandatory or highly suggested in several journals.



Test statistics

With some assumptions, the test statistic is usually:

$$t = \frac{b}{\mathsf{SE}_b}$$

Where b is the effect size (e.g., difference between two conditions) and SE_b is the standard error of the numerator.

Increasing participants

In simple settings, SE_b is:

$$\mathsf{SE}_b = \sqrt{\frac{\sigma_b^2}{n}}$$

Thus our job is reducing SE_b, mainly increasing the number of participants.

Not only participants

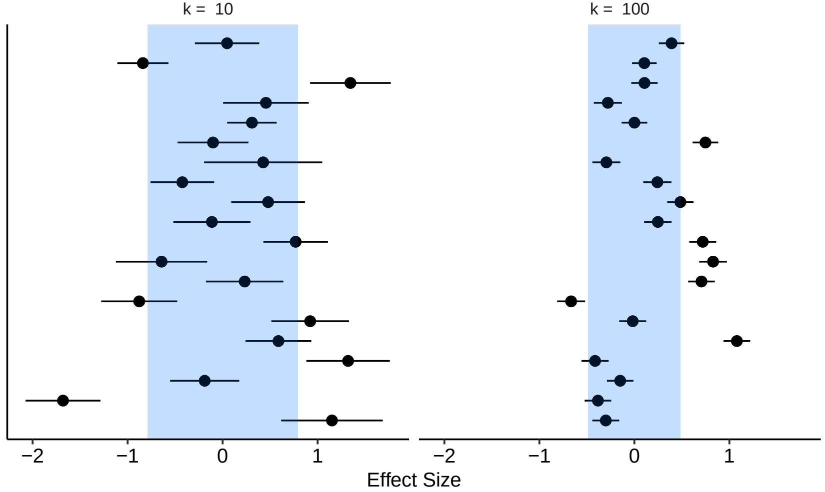
Often, the power can be affected also increasing trials (k), not only participants (n)[1]

$$\mathsf{SE}_b^\star = \sqrt{\frac{\sigma_s^2}{n} + \frac{\sigma_w^2}{kn}}$$

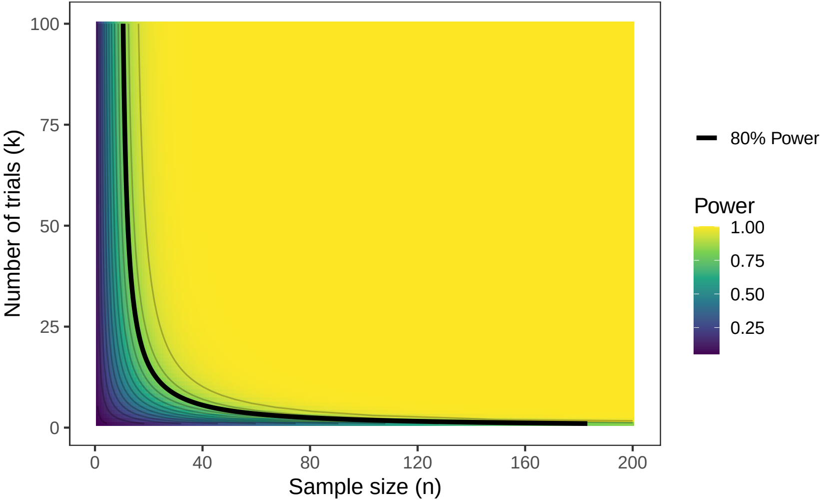
Where σ_s^2 is the variance between participants and σ_w^2 is the variance within participants. When σ_w^2 is close to zero, there is no advantage in adding trials.

[1] Miller, J. (2024). How many participants? How many trials? Maximizing the power of reaction time studies. *Behavior Research Methods*, *56*, 2398–2421. https://doi.org/10.3758/s13428-023-02155-9

Same participants, more trials



Power ~~curves~~ contours[2]



[2] Baker, D. H. … Andrews, T. J. (2021). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*, *26*, 295–314. https://doi.org/10.1037/met0000337

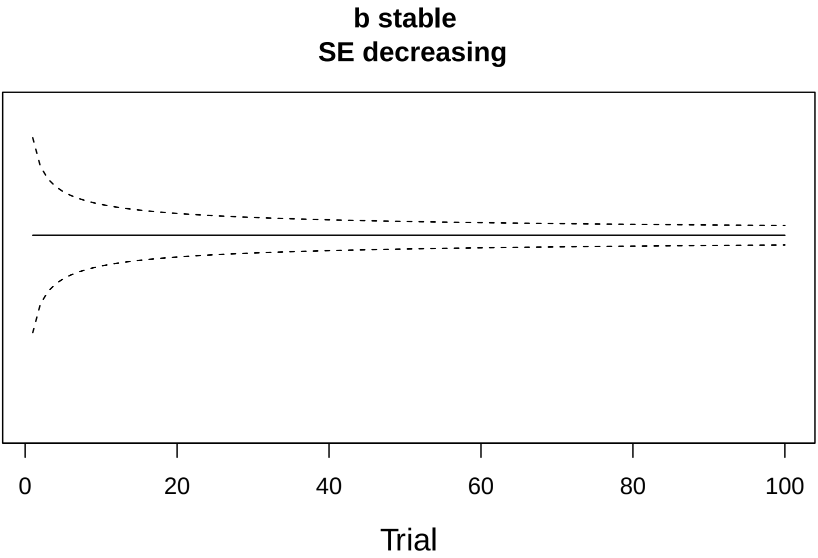
Are all trials the same?

The main problem...

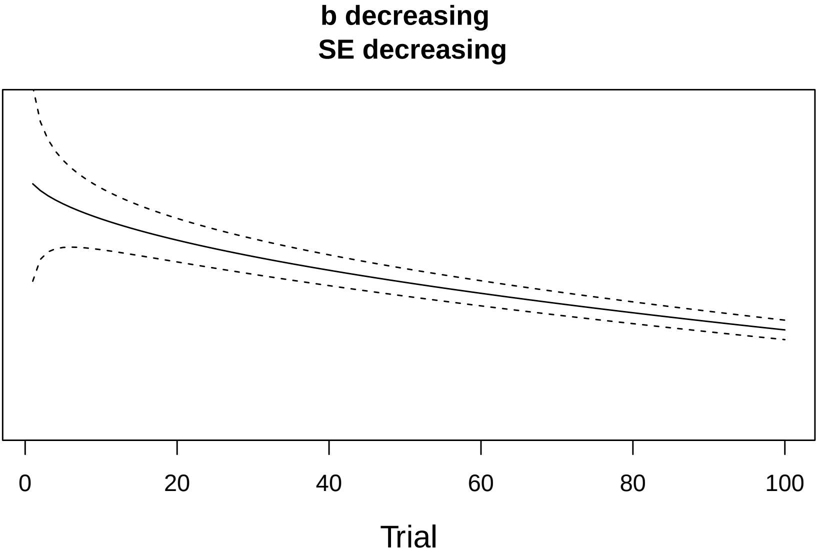
When doing simulations taking into account the trials k we are (usually) assuming that each trial is the same, regardless of:

- fatigue
- learning effects
- attention
- ...

The usual assumption



What about this?



Application to real data

Classic experiments

We collected 214 university students performing ~ 330 trials on three classical experimental paradigms:

- Simon Effect
- Snarc Effect
- Task Switching

In all paradigms there is a comparison between congruent and incongruent trials where incongruent trials are expected to elict slower reaction times.

The mixed-effects model

In R-like notation the model is:

```
rt ~ congruence + (congruence|participant)
```

```
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 2687.8   51.84
          congruencei  111.6   10.57    -0.02
 Residual             9144.5   95.63
Number of obs: 65601, groups:  id, 207

Fixed effects:
            Estimate Std. Error t value
(Intercept)  423.643      3.642  116.32
congruencei   24.449      1.048   23.34
```

The mixed-effects model

In R-like notation the model is:

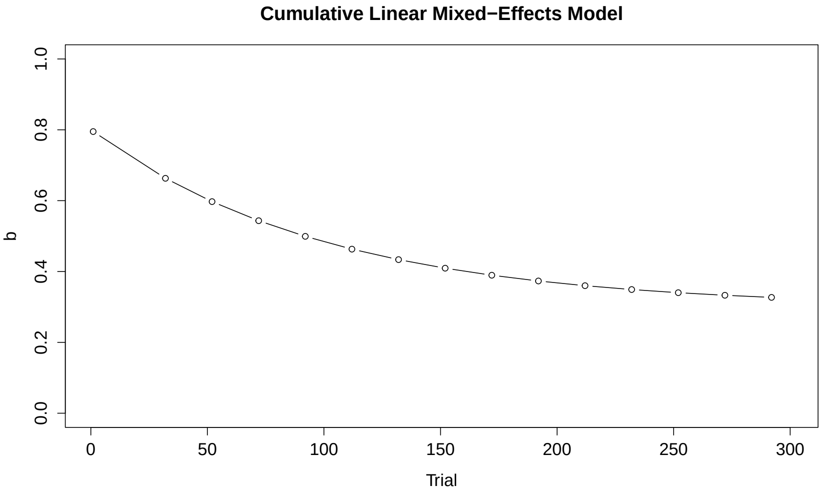
```
rt ~ congruence + (congruence|participant)
```

```
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 2687.8   51.84
          congruencei  111.6   10.57    -0.02
 Residual             9144.5   95.63
Number of obs: 65601, groups:  id, 207

Fixed effects:
            Estimate Std. Error t value
(Intercept)  423.643      3.642  116.32
congruencei   24.449      1.048   23.34
```

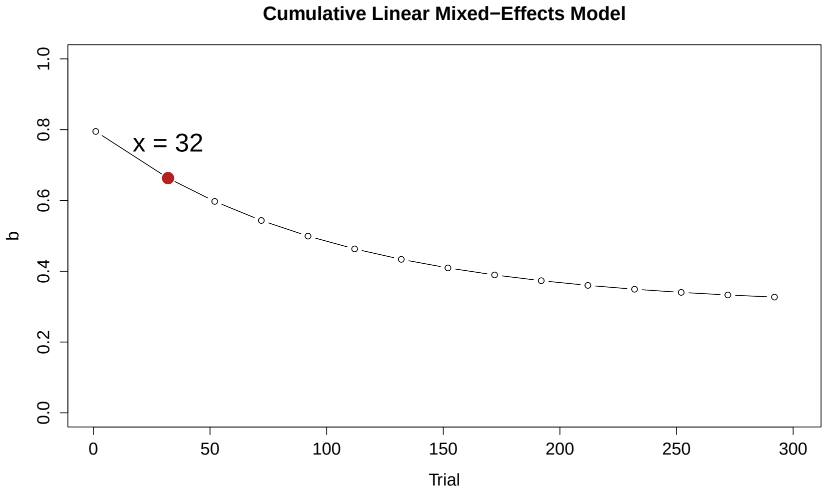
Cumulative mixed-effects model

We fitted the previous model starting with 32 trials and then adding k trials.



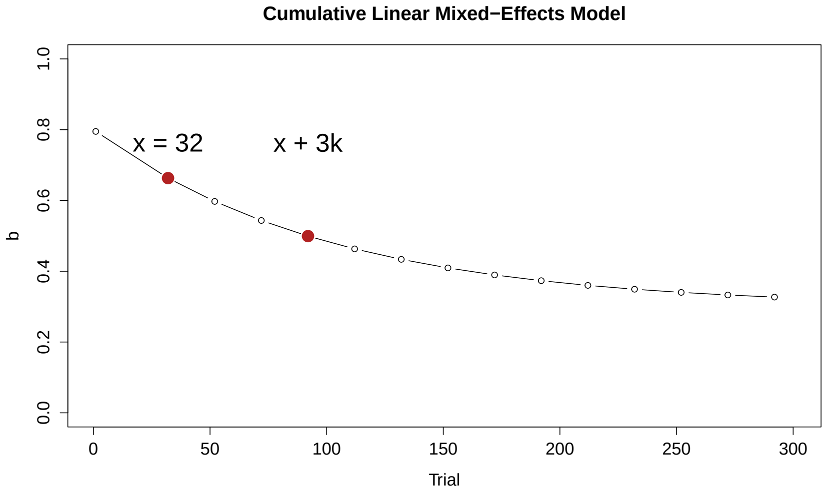
Cumulative mixed-effects model

We fitted the previous model starting with 32 trials and then adding k trials.



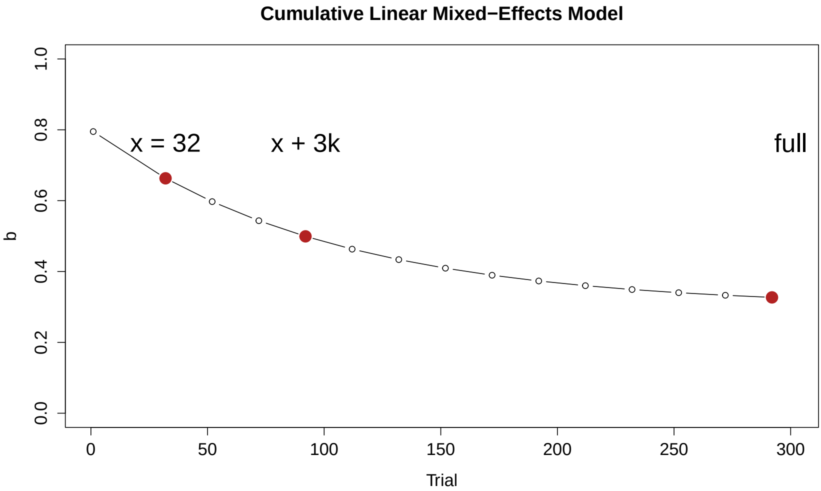
Cumulative mixed-effects model

We fitted the previous model starting with 32 trials and then adding k trials.



Cumulative mixed-effects model

We fitted the previous model starting with 32 trials and then adding k trials.



Results

We present the results according to the from the test statistics:

$$t = \frac{b}{\mathsf{SE}_b}$$

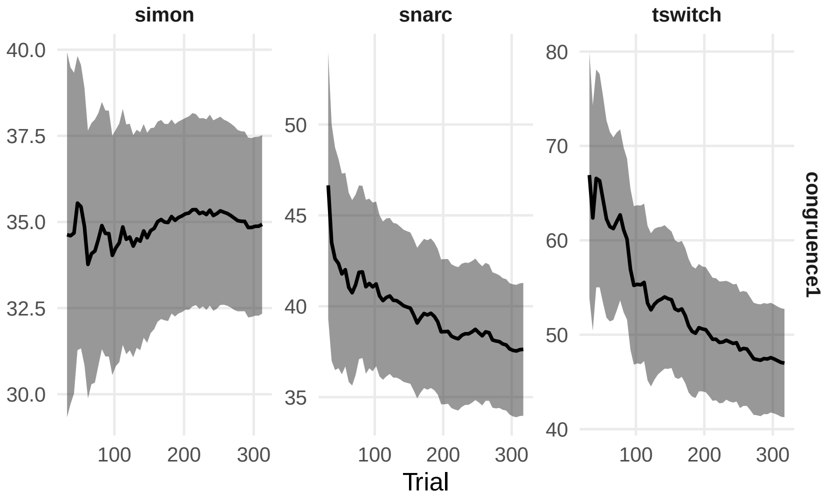
Results, b

We present the results according to the test statistics:

$$t = \frac{b}{\mathrm{SE}_b}$$

Results, b

Only the Simon effect is stable, the other effects decrease over time.



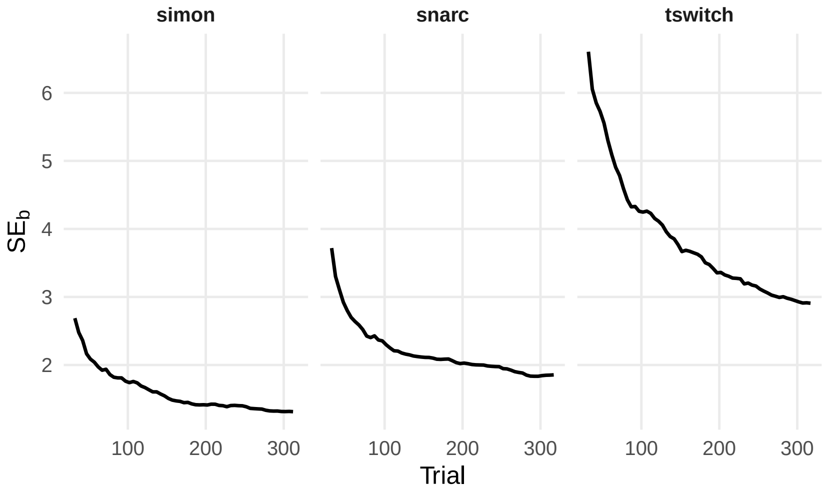
Results, SE

We present the results according to the test statistics:

$$t = \frac{b}{\mathsf{SE}_b}$$

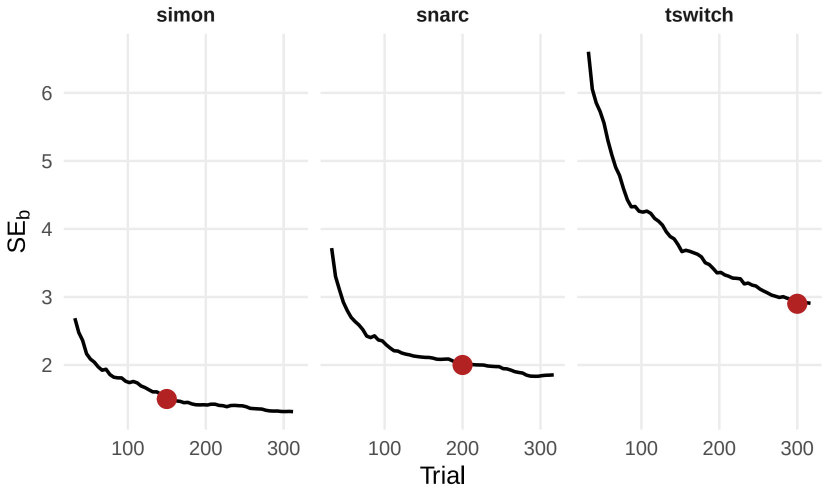
Results, SE

The reduction of the SE reaches a plateau by the midpoint of the experiment (except for Task Switching).



Results, SE

The reduction of the SE reaches a plateau by the midpoint of the experiment (except for Task Switching).



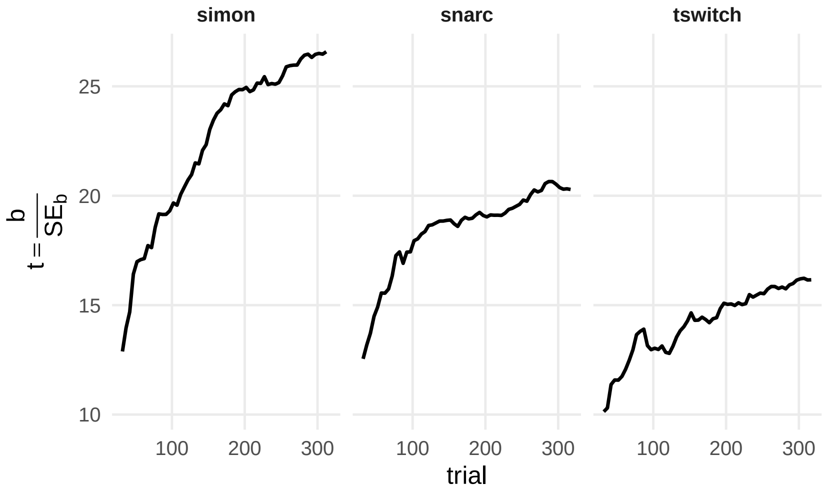
Results, t

We present the results according to the test statistics:

$$t = \frac{b}{\mathsf{SE}_b}$$

Results, t

The Simon effect is the only one that seems to benefit, whereas the others reach a plateau by the midpoint of the experiment.



Conclusions

- these results are not exaustive but related to these specific experiments

Conclusions

- these results are not exaustive but related to these specific experiments
- the crucial point is considering how the effect evolves over time, improving our power analysis and experimental planning

Conclusions

- these results are not exaustive but related to these specific experiments
- the crucial point is considering how the effect evolves over time, improving our power analysis and experimental planning
- interactions or more complex effects could require a large number of trials

References

Miller, J. (2024). How many participants? How many trials? Maximizing the power of reaction time studies. *Behavior Research Methods*, *56*, 2398–2421. https://doi.org/10.3758/s13428-023-02155-9

Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2021). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*, *26*, 295–314. https://doi.org/10.1037/met0000337

filippo.gambarota@unipd.it

filippogambarota.github.io

Slides

(*Last update: 2025-09-13*)