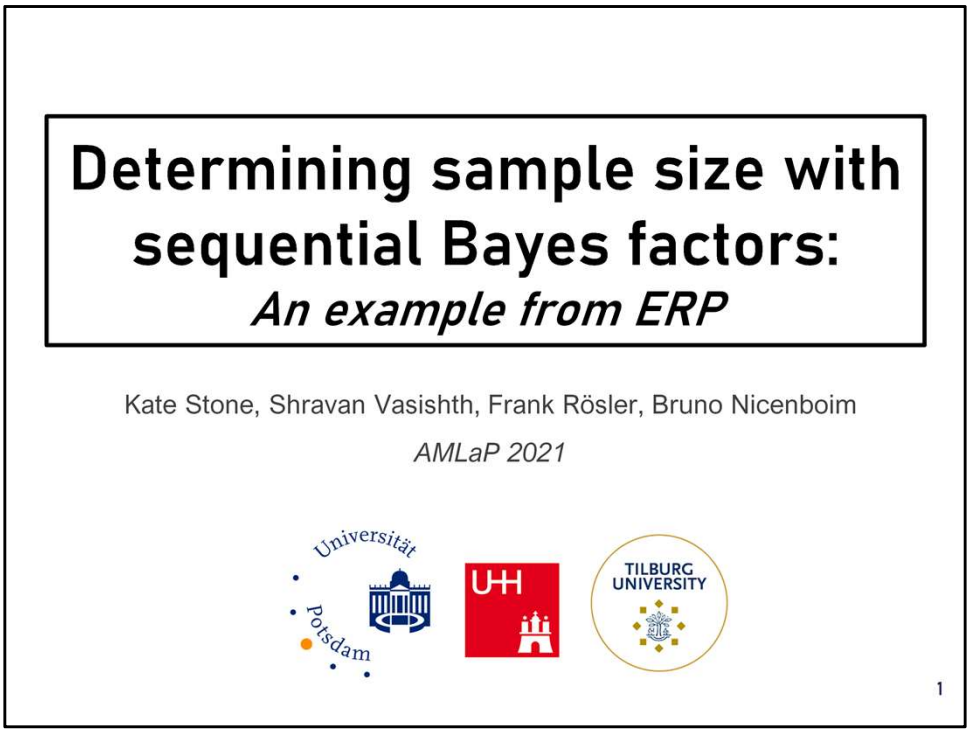
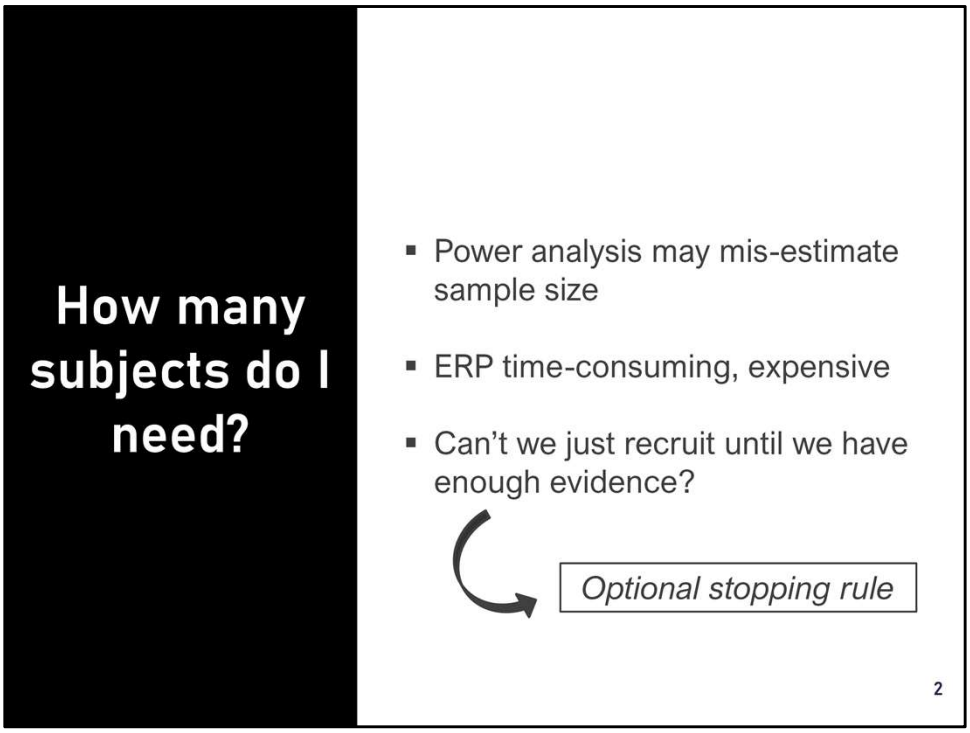


I'm going to talk about our experiences of determining sample size for an ERP experiment using sequential bayes factors



Burning question for every scientist: How many subjects?

Power analysis may under-estimate or over-estimate sample size

Especially bad for something like ERP, which is time-consuming and expensive

Nice: only recruit as many subjects as necessary to provide evidence for/against hypotheses

This is called an optional stopping rule

And I'm going to give an example of how we applied it to an ERP experiment

Optional stopping with p-values can be misleading

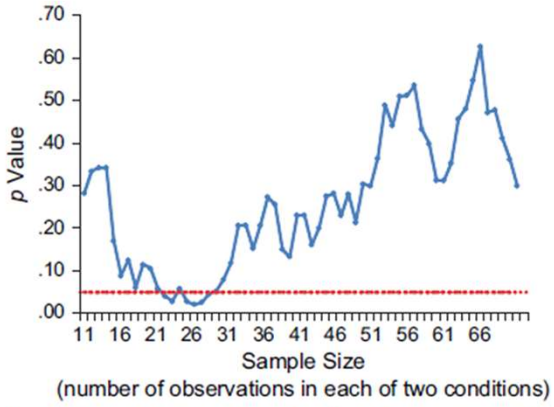


Fig. 2. Illustrative simulation of *p* values obtained by a researcher who continuously adds an observation to each of two conditions, conducting a *t* test after each addition. The dotted line highlights the conventional significance criterion of $p \leq .05$.

Simmons, Nelson & Simonsohn (2011) 3

One way to do optional stopping is to compute a p-value after adding subjects and stop when p-value significant

VERY misleading if don't take measures to control false positive rate

In this simulation by Simmons et al we see how the p-value varies as we add participants

Note that it dips below red line (0.05) at 26 participants.

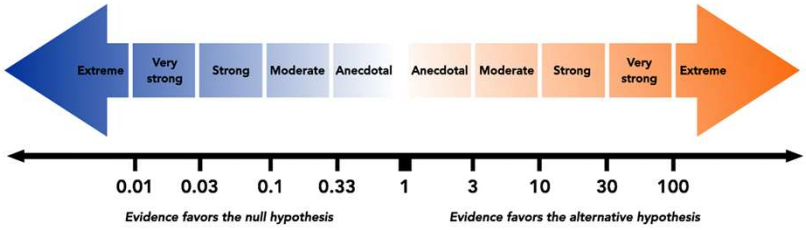
So we would stop recruiting, claim evidence for an effect → false positive

False positive rate inflated because essentially doing an unforeseeable number of multiple comparisons

Instead, I'm going to show you an example using Bayes factors

What is a Bayes factor?

- Fit two statistical models representing competing hypotheses
- Bayes factor is ratio of evidence for one model over the other



Quintana & Williams (2018)

4

To compute a Bayes factor…

Fit models representing competing hypotheses

BF = ratio of evidence for one model over the other

GIVEN THE DATA AND THE PRIORS

Scale: Ratio 1:1 indicates equivalent evidence for each model

Depending on which direction we move away from 1 and how far, evidence increases for one model or the other

Optional stopping with Bayes factors

Bayes factor's advantage is interpretation:

- Belief about relative plausibility of two models given data

5

Why is it ok to compute multiple Bayes factors when doing optional stopping?

Because of the BF's interpretation

We're comparing the "relative plausibility of two models" given data

Relative plausibility is not affected by how many times we compute Bayes factor or how many participants we add

Sequential Bayes factor design

- Decide a priori cut-off Bayes factor
- Start recruitment
- Compute Bayes factor periodically
- Stop recruitment when Bayes factor reaches the cut-off
- Option: Set a maximum

Schönbrodt et al., 2015; Schönbrodt & Wagenmakers, 2018

6

Applied to an experiment: Design called SBF

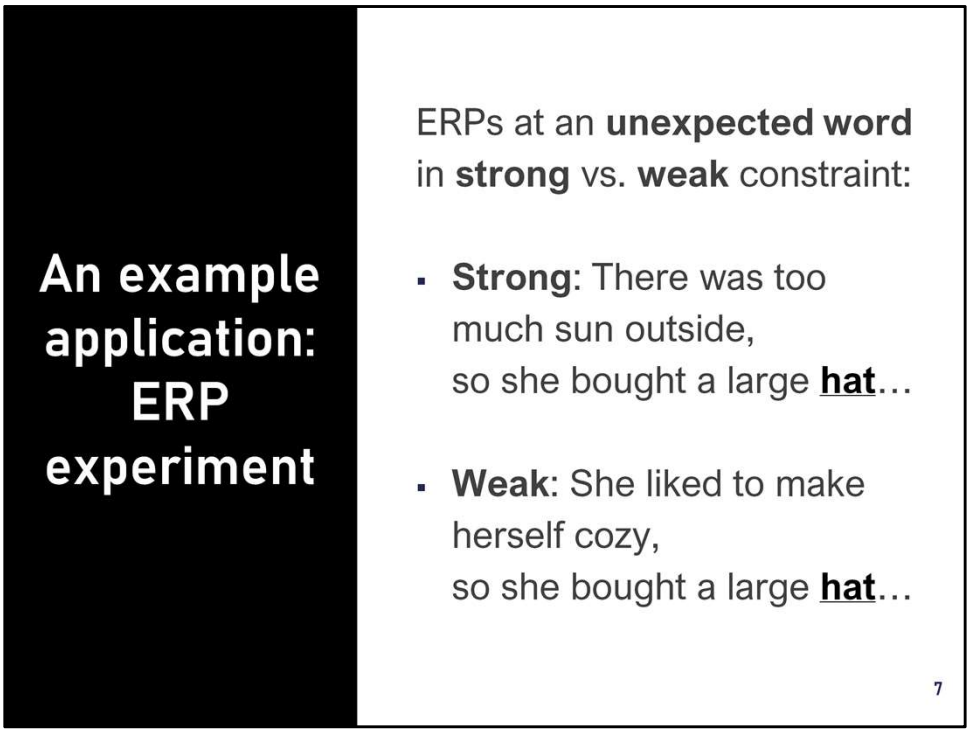
Decide a priori on Bayes factor cut-off

Recruit until cut-off reached

Option: set a maximum

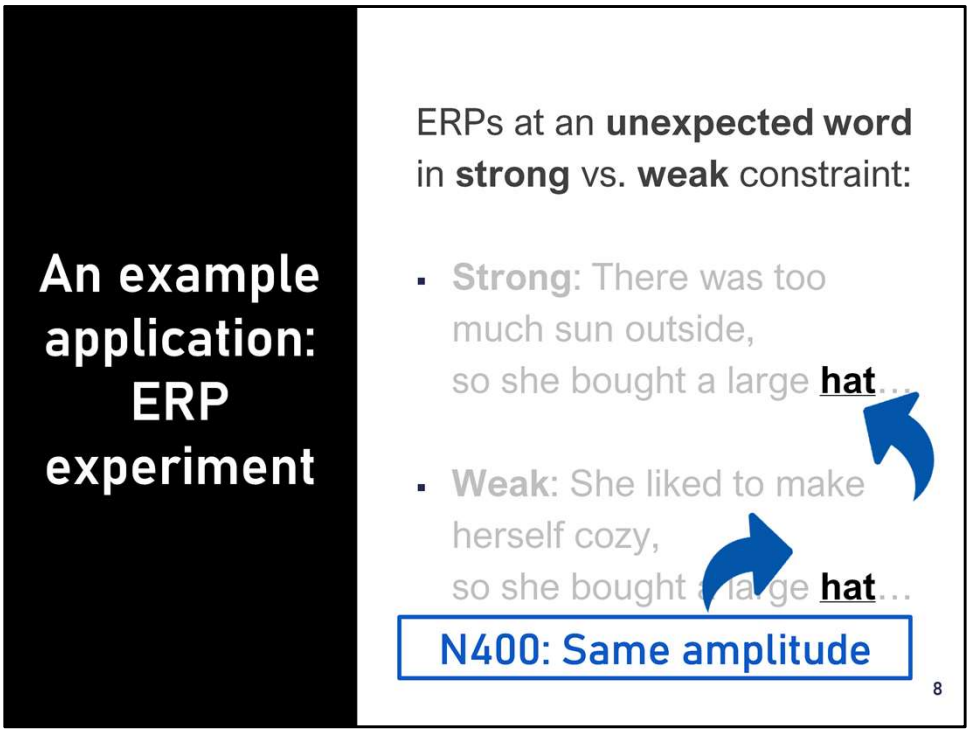
Nothing new, seen a couple of examples in psycholing, but new and exciting to me

Now I'll describe our example experiment



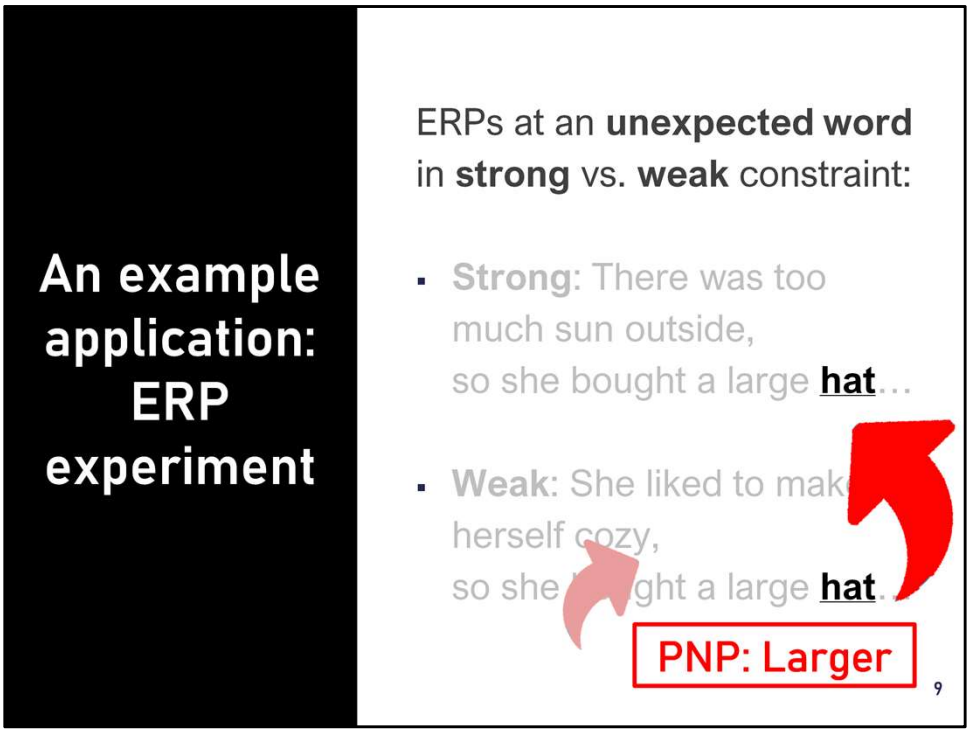
Compared ERPs at unexpected words in strong vs. weakly constraining contexts

We looked at two ERP components:



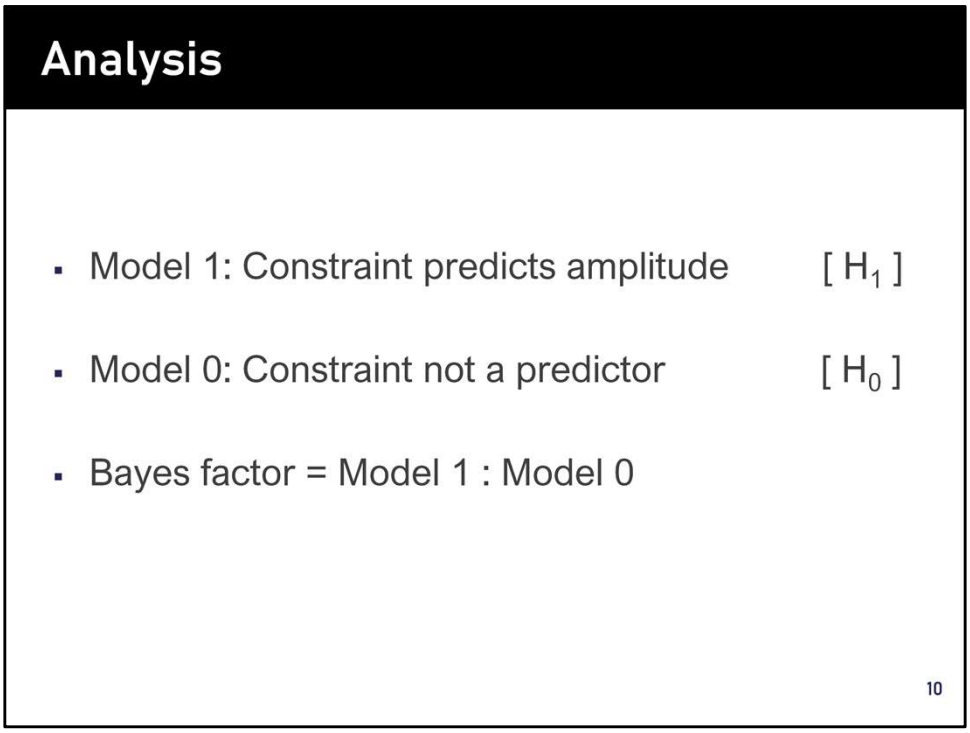
The first was the N400

Based on literature…



The second was the post-N400 positivity

Based on the literature…



Here is how our models mapped to our hypotheses:

Model 1 assumed constraint would predict amplitude – mapped to H1

Model 0 assumed constraint not a predictor – mapped to H0

BF = ratio M1:M0

Sample size

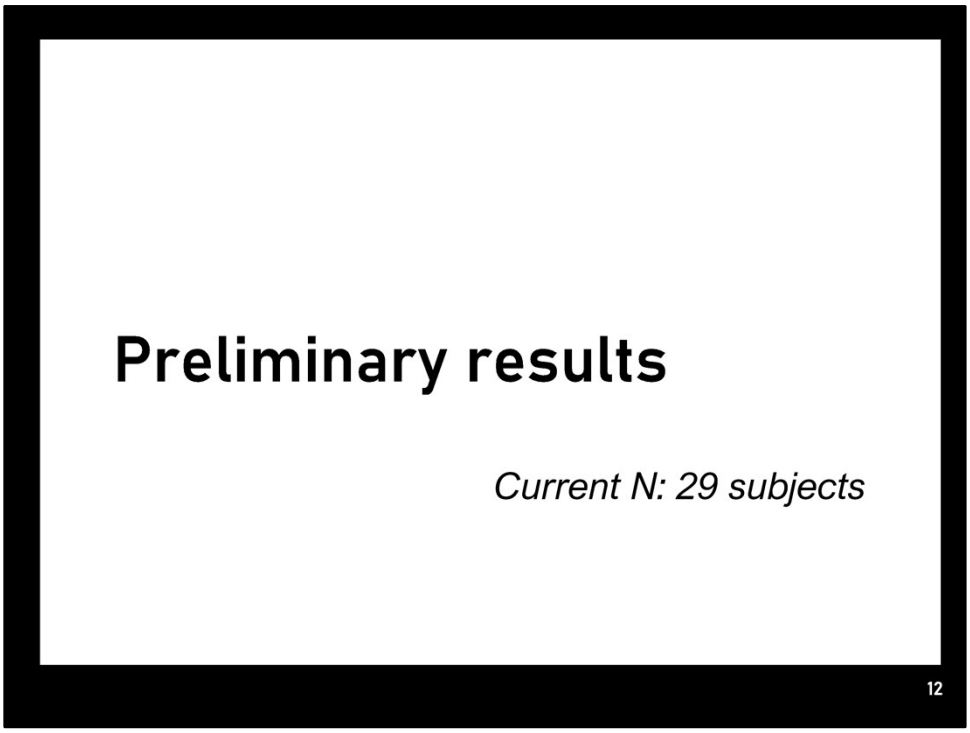
- Recruit until Bayes factor of at least 10 (strong evidence)
- Max. 150 participants

11

We determined our sample size like so

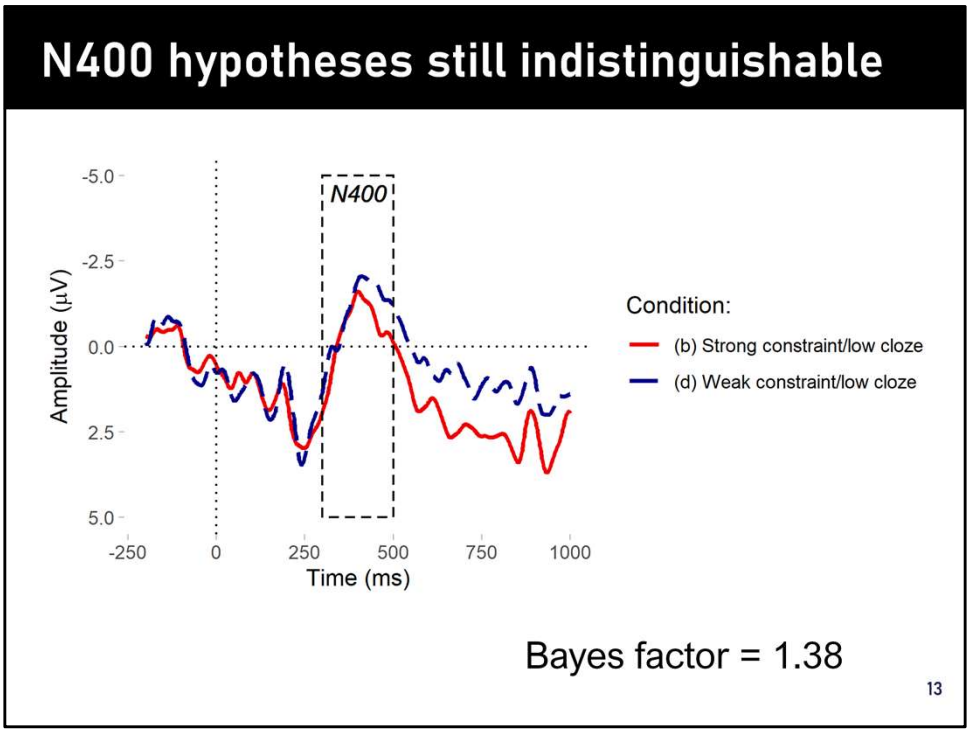
We decided on a cut-off of 10 based on priors / models

Stage 1 RR approved



Unfortunately only 29 recruited so far

So I will present preliminary results

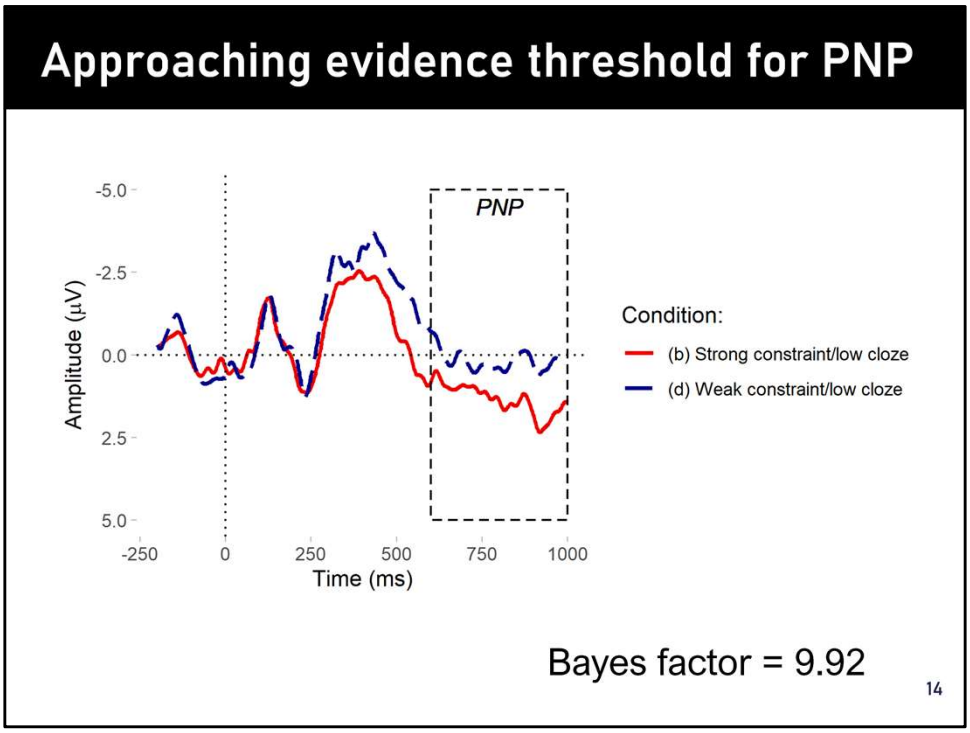


Interesting distinction in pre- / post-PEAK N400 amplitude

Maybe because of this, Bayes factor of close to 1

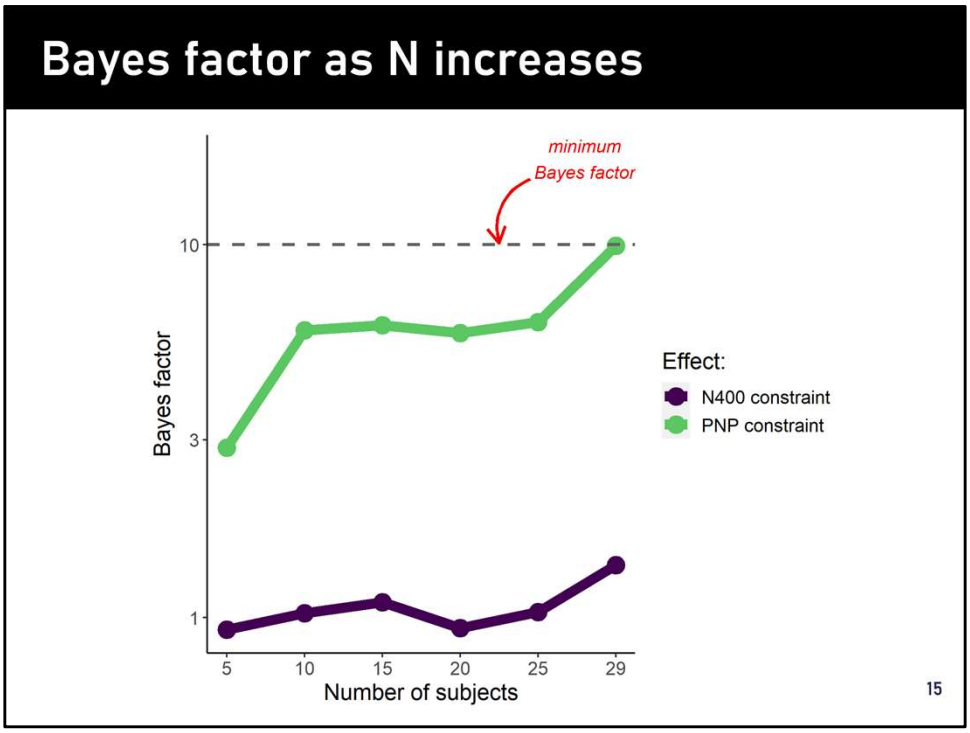
Can talk more about this in question time

Summary: At current sample size, unable to distinguish between N400 hypotheses



There appears to be a PNP constraint effect in expected direction

Bayes factor already very close to evidence threshold!



Just quickly, evolution of BF as sample size increases

1. Imagine we had reached cut-off with 29 subjects! Saved a lot of resources, provided strong evidence for hypotheses

Would this mean that all ERP experiments only need 29 subjects? NO! Specific to this design, these models, these priors, these data.

2. Imagine we recruit to 150 and don't reach cut-off, results will still be interpretable:

- Posterior estimates effect size
- BFs < 10 still interpretable
- BF < 3 (inconclusive) = tells us something about the adequacy of our design for answering the question

Thank you!

Q & A

Q: *Have you re-analysed previous ERP studies to see whether they would have reached the BF=10 cut-off? What would be your intuition about whether more/less participants were needed?*
A: No, no re-analysis. It would depend heavily on what priors had been assumed and what a reasonable cut-off for those studies was (we only chose 10 because of our specific priors/models). But my intuition would be they needed more participants rather than less.

Q: *To determine the priors, could you first test a small number of participants and then use the posterior estimate as a prior?*
A: As long as these participants weren't included in the final analysis, yes – a small pilot study can be useful. Although with a small number of pilot participants, the posterior estimates could be quite noisy, so you'd also want to consider the size and direction of effect estimates from the literature.

Q: *Is it possible to do the equivalent of p-hacking with Bayes factors?*
A: Absolutely! You can definitely set up your priors so that you'll find strong evidence with a small sample size. It's important therefore to do (and look out for) sensitivity analyses, where Bayes factors for a range of priors are shown. For example, we pre-registered truncated priors with a narrow standard deviation to test our hypotheses, but to see how different priors might affect our conclusions, we also pre-registered sensitivity analyses with increasingly large standard deviations, as well as with non-truncated priors.

Q: *Would this approach be suitable for new students? What resources can you recommend?*
A: The student would need an understanding of Bayesian analysis, so perhaps it's not suitable for complete beginners. I can highly recommend the Uni Potsdam SMLP Summer School: https://vasishth.github.io/smlp2021/ Materials for all the tracks are posted online, e.g. the Introduction to Bayesian analysis is here: https://vasishth.github.io/IntroBayesSMLP2021/ A textbook is available online here: https://vasishth.github.io/bayescogsci/

References

Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry*, *18*(1), 178. https://doi.org/10.1186/s12888-018-1761-4

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*(2), 301–308. https://doi.org/10.3758/s13423-014-0595-4

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, *25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

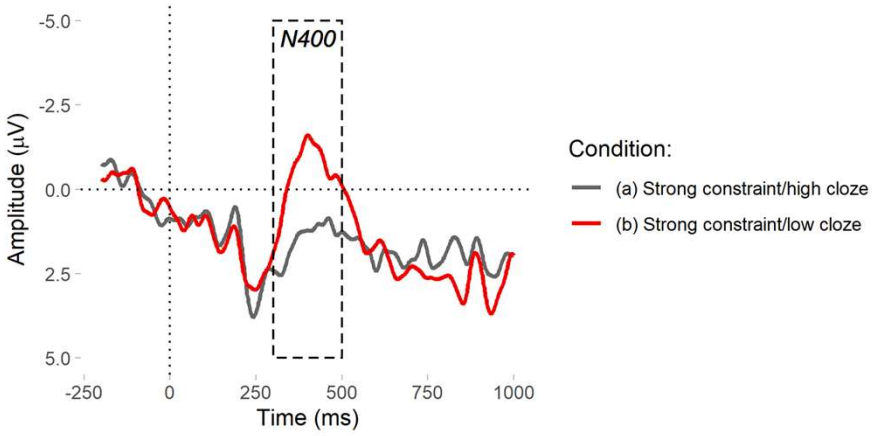
Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2015). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322–339. https://doi.org/10.1037/met0000061

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Bonus slides

19

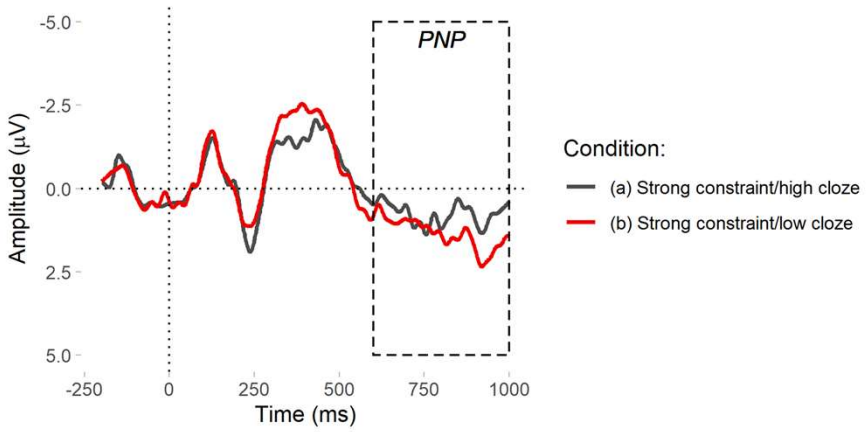
Evidence threshold exceeded for N400 predictability



Bayes factor = 402

20

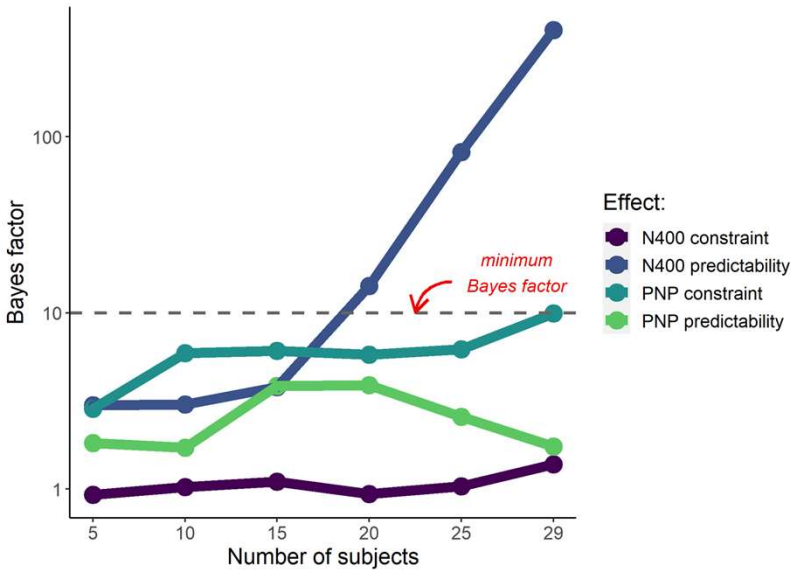
PNP predictability hypotheses indistinguishable



Bayes factor = 1.74

21

Bayes factor as N increases



22

N400 statistical model

```
N400 ~ constraint + predictability +
  (1 | item) + (1 + constraint + predictability | subj)
```

Where:
- Constraint = entropy
- Predictability = log2 smoothed cloze probability

Priors:
- Constraint: N(0, 0.2)
- Predictability: $N_+(0, 0.2)$

23

PNP statistical model

```
PNP ~ constraint + predictability +
  (1 | item) + (1 + constraint + predictability | subj)
```

Where:
- Constraint = entropy
- Predictability = log2 smoothed cloze probability

Priors:
- Constraint: $N(0, 0.2)$
- Predictability: $N(0, 0.2)$

24