Analysing the timecourse of experimental effects in eyetracking data

Kate Stone
Salzburg 2023

Temporal questions in psycholinguistics

What drives eye fixations?

Oculomotor variables

Cognition

e.g. Yang & McConkie (2001) Vs. Engbert et al. (2002)

When are different information types processed?

Syntax

Semantics

e.g. Frazier (1987) vs. MacDonald (1993)

How do predictions develop over time?

Local constraints

Global constraints

e.g. Kukona et al. (2014), Stone et al (2021)

Why is L2 comprehension slower than L1?

Speed

Capacity

e.g. Clahsen & Felser (2018) vs. Hopp (2013)

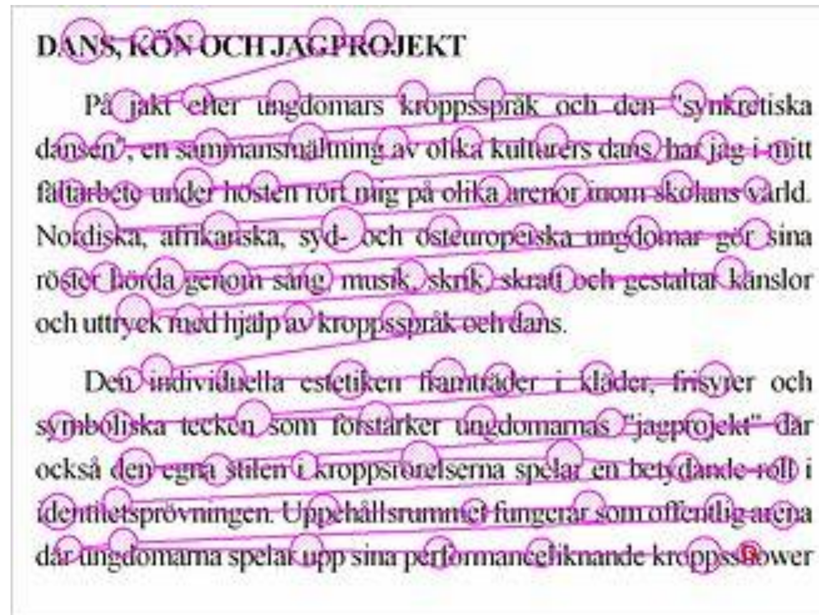
Outline

1. Temporal analysis methods in eye tracking
2. Example: Effect onset analysis

1. Temporal analysis methods in eye tracking

- Eye tracking while reading

Eye tracking while reading



Common reading eye tracking measures

"Early"

- ***First fixation duration:*** Duration of initial fixation on a region
- ***First pass reading time:*** All fixations in region before leaving it to the left or right
- ***Regression path duration:*** All fixations in a region until leaving it to the right
- ***Re-reading time:*** All non-first pass times until the first fixation to the right
- ***Total fixation time:*** Sum of all fixations in a region

"Late"

Liversedge et al. (1998)

Is syntactic processing slower in L2?

*The key to the cabinets **are** on the table

Lim & Christianson (2014)

Is syntactic processing slower in L2?

L1

*The key to the cabinets **are** on the table

L2

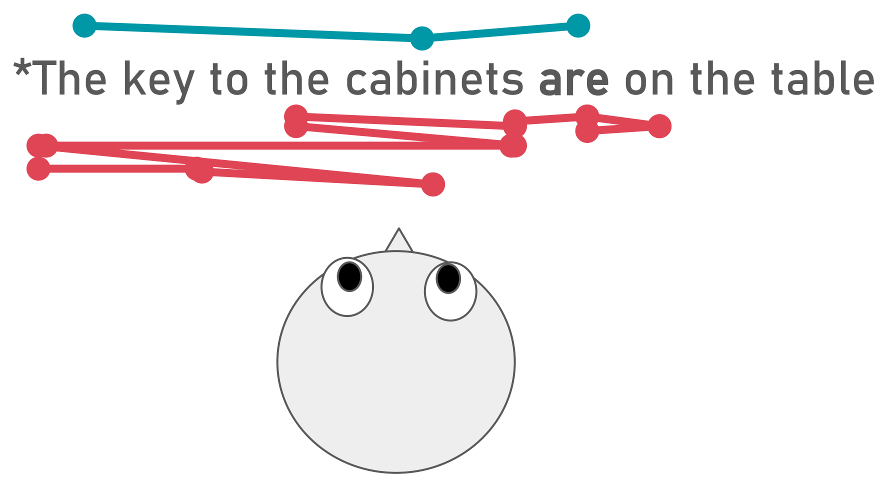
Lim & Christianson (2014)

Is syntactic processing slower in L2?

L1

*The key to the cabinets **are** on the table

L2



Lim & Christianson (2014)

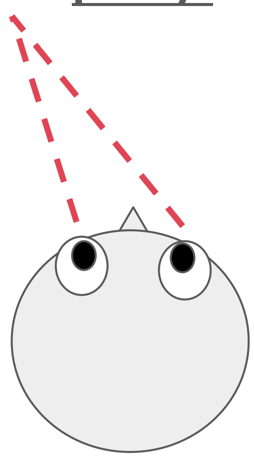
When does word frequency affect processing?

John decided to sell the **table/banjo** in the garage sale.

Reingold et al. (2012)

When does word frequency affect processing?

John decided to sell the **purty** in the garage sale.



Reingold et al. (2012)

When does word frequency affect processing?

John decided to sell the **<u>table</u>** in the garage sale.

Reingold et al. (2012)

When does word frequency affect processing?

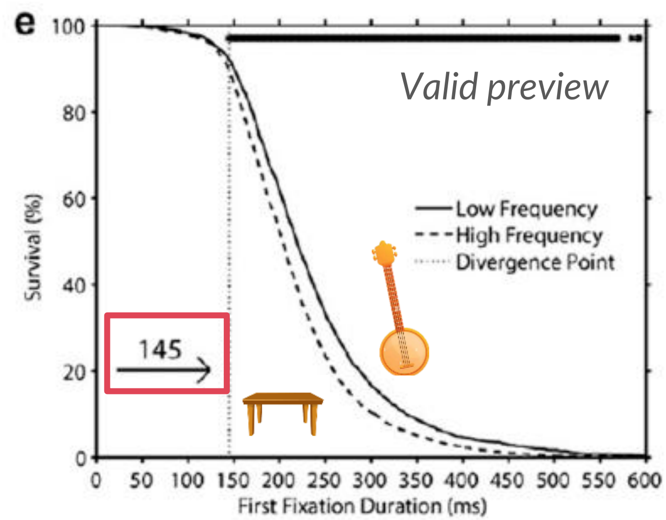
- Survival analysis: % of all fixations longer than time “*t*”

- Divergence point: when do the survival curves diverge significantly? (i.e. what is the shortest fixation duration that was influenced by frequency?)

When does word frequency affect processing?

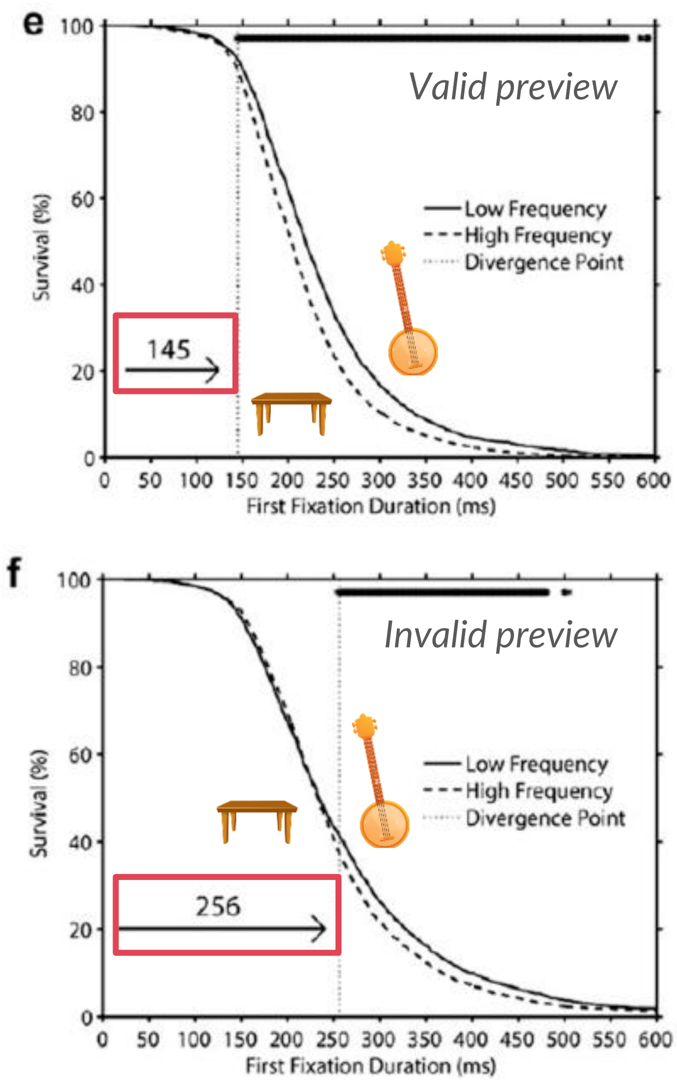
- Survival analysis: % of all fixations longer than time "*t*"

- Divergence point: when do the survival curves diverge significantly? (i.e. what is the shortest fixation duration that was influenced by frequency?)



When does word frequency affect processing?

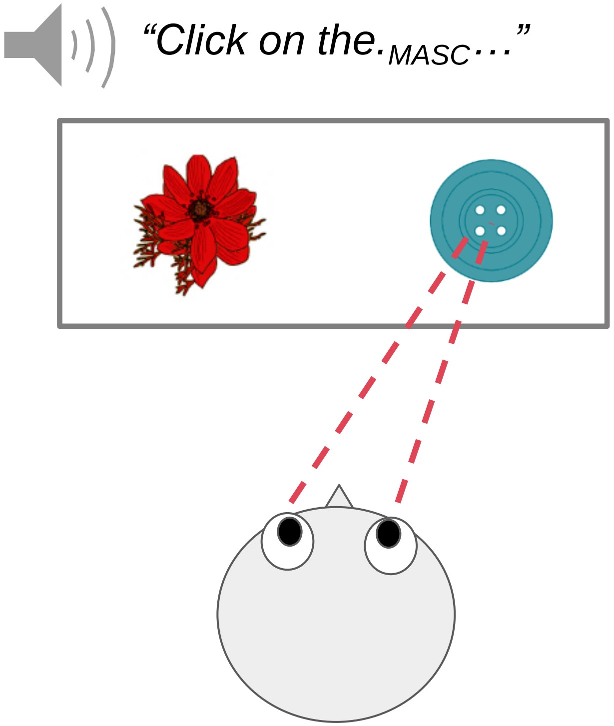
- Survival analysis: % of all fixations longer than time "*t*"
- Divergence point: when do the survival curves diverge significantly? (i.e. what is the shortest fixation duration that was influenced by frequency?)



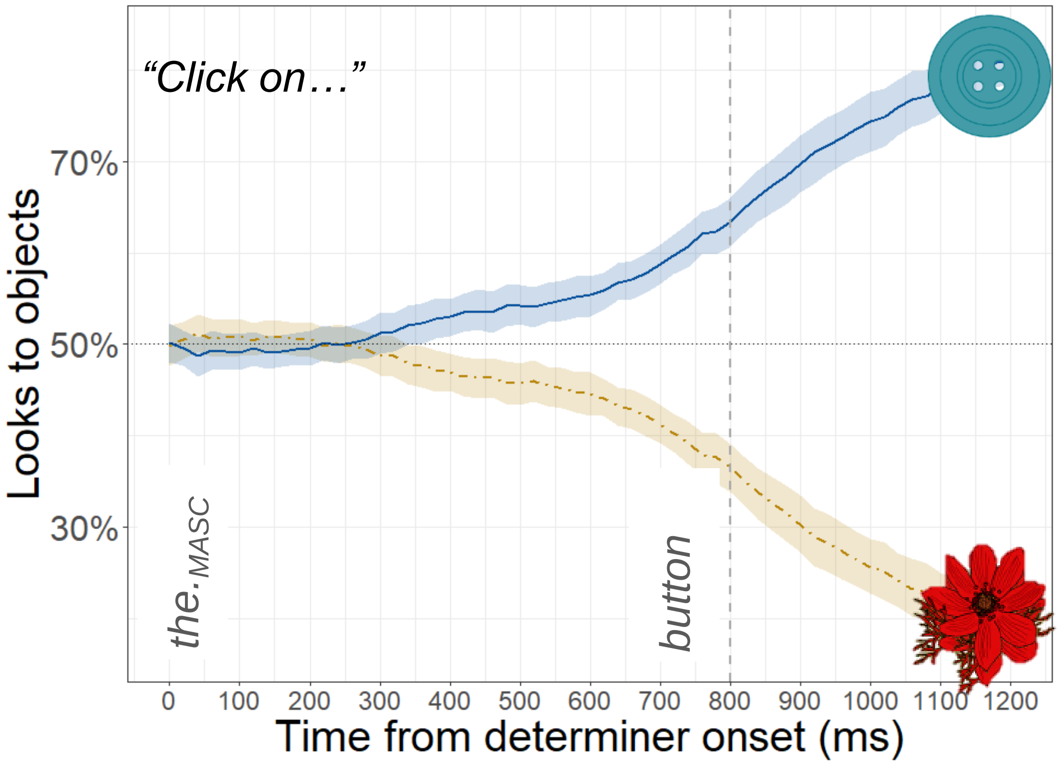
1. Temporal analysis in eye tracking

- Eye tracking while reading
- Eye tracking while listening (visual world paradigm)

The visual world paradigm

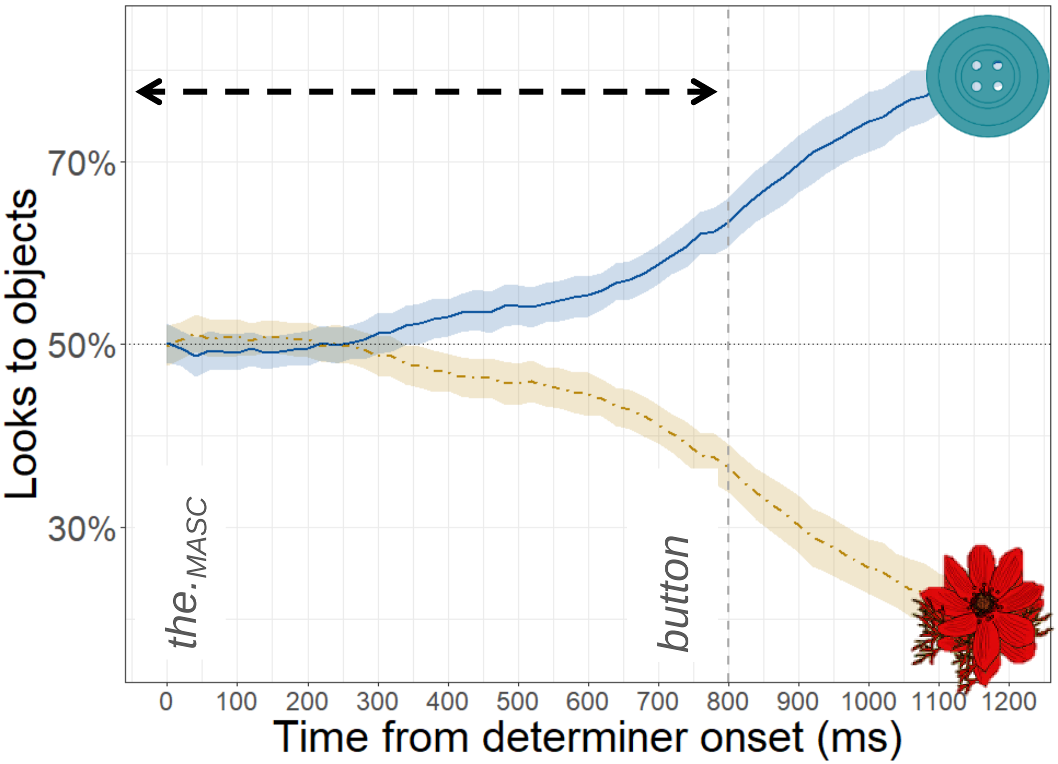


The visual world paradigm



Analysis methods

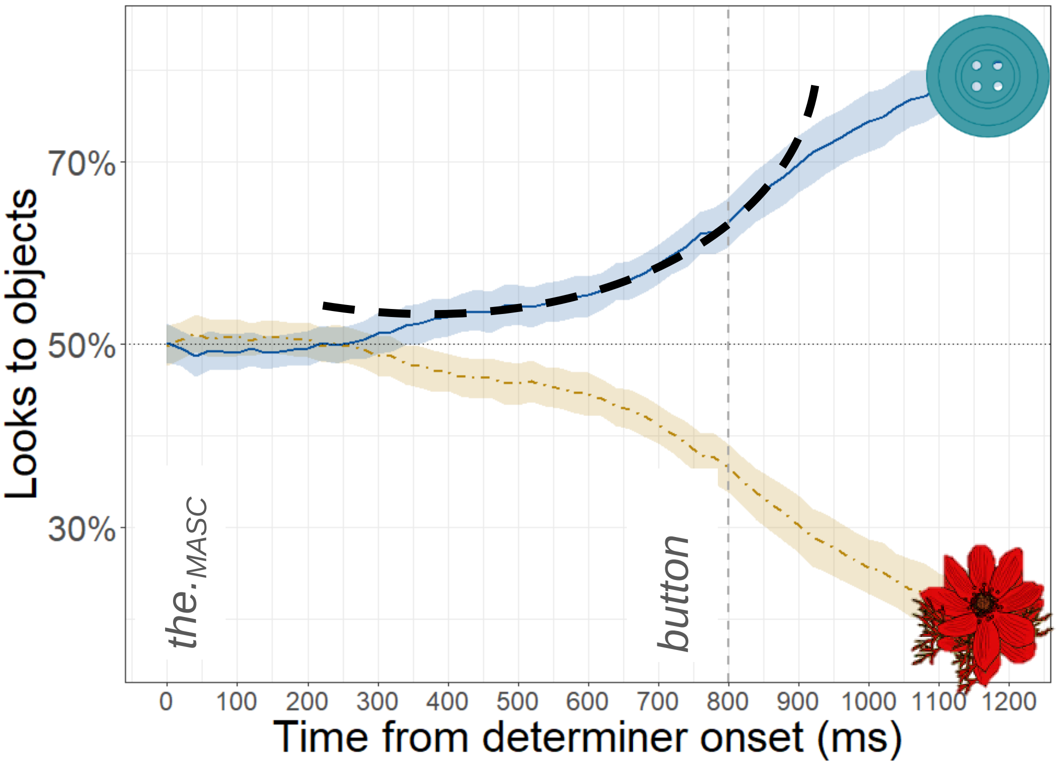
1. Were there more looks to one object during a certain time window?



Analysis methods

1. Were there more looks to one object during a certain time window?

2. Does the slope/shape of the button fixation pattern differ from the flower?

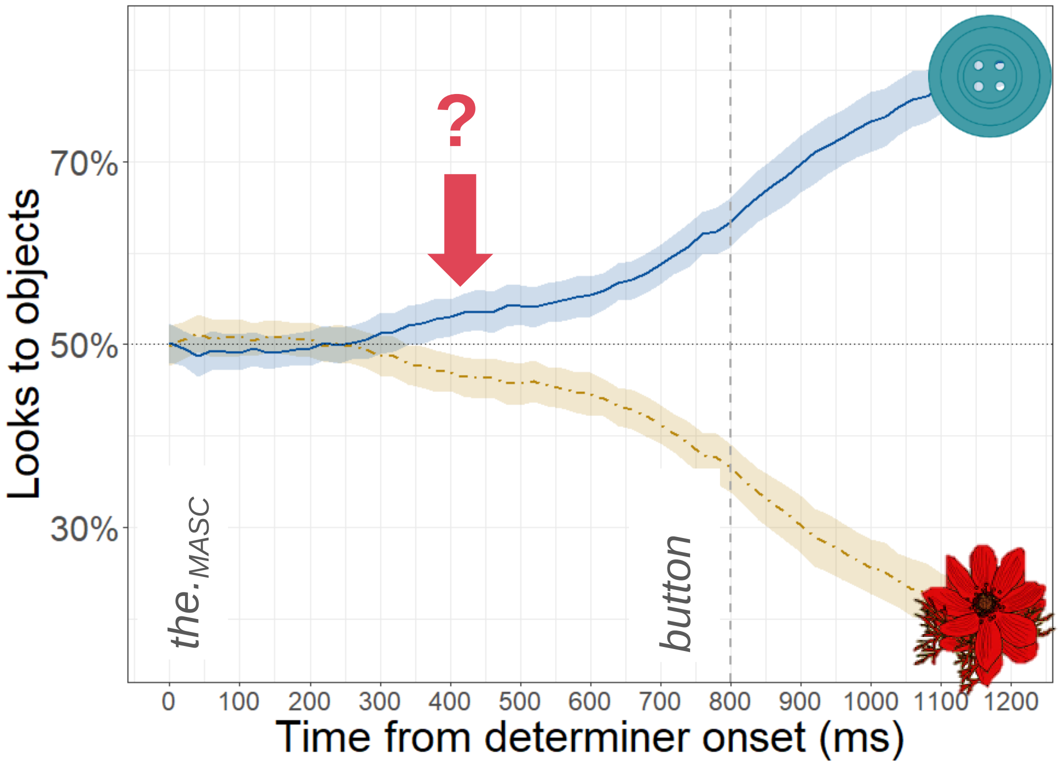


Analysis methods

1. Were there more looks to one object during a certain time window?

2. Does the curve of the button fixation pattern show a significant upswing?

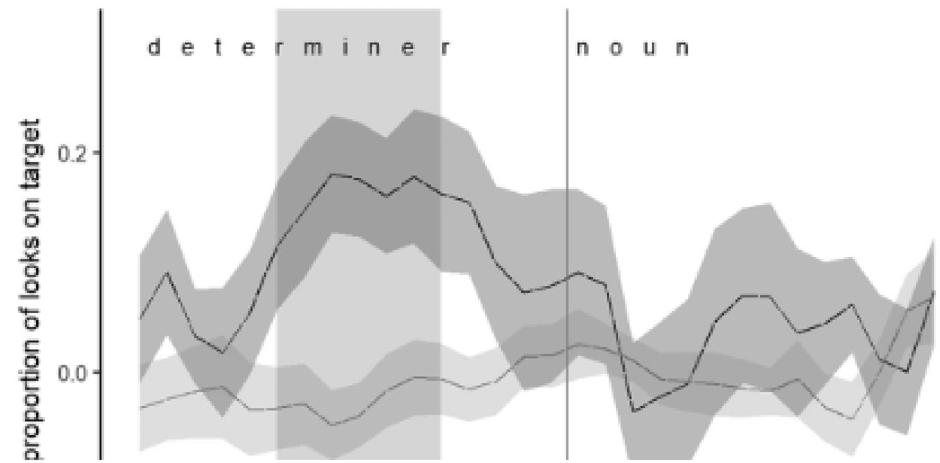
When do people first prefer the button?



Existing temporal analysis methods

In which window was there a significant difference?

Barr et al., 2014; Seedorff et al., 2017



DOI: 10.1111/psyp.13335

WILEY PSYCHOPHYSIOLOGY

ORIGINAL ARTICLE

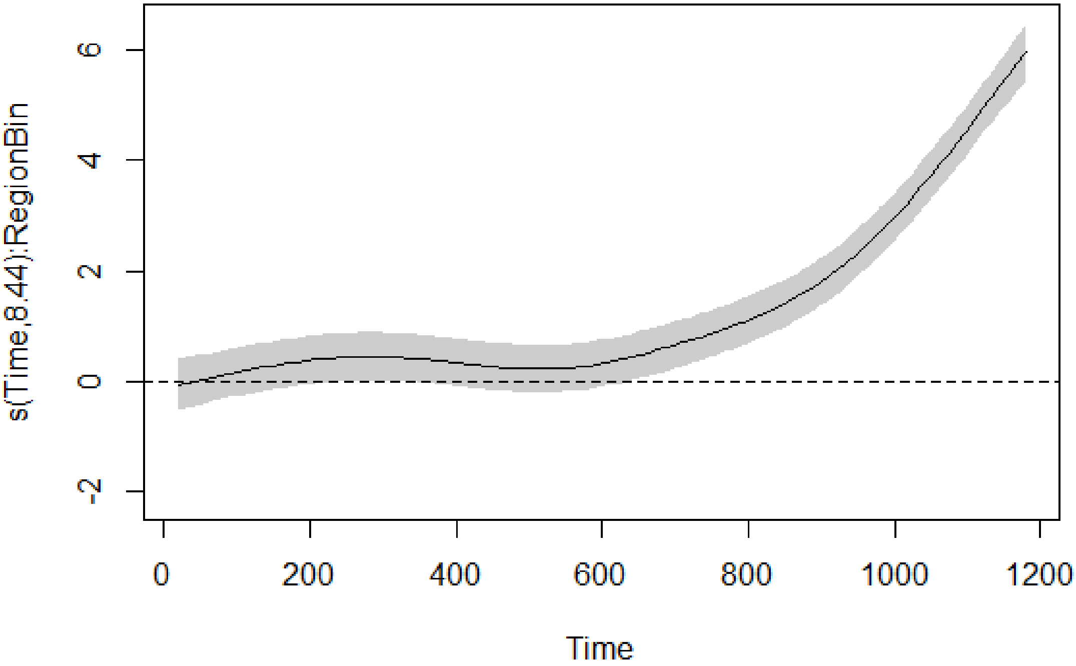
Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location

Jona Sassenhagen | **Dejan Draschkow**

Existing temporal analysis methods

When does the difference curve rise above chance?

GAMMs; van Rij, 2015; van Rij et al., 2020; Miwa & Baayen, 2020



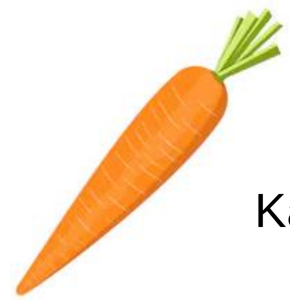
2. (Our) divergence point analysis

- Onset estimate
- Temporal uncertainty
- Compare onsets between groups

Experiment: Syntactic gender predictions

"Der Hase frisst den... "

The rabbit eats the.MASC...



Karrotte.FEM

Kohl.**MASC**

Dussias et al., 2013; Grüter et al., 2012; Hopp, 2013; Hopp & Lemmerth , 2018; Lemmerth & Hopp, 2018

Experiment: Syntactic gender predictions

Non-native (L2) speakers are not as fast as native speakers:

- Even if they're highly proficient *Grüter et al. (2012)*
- Even if only one object matches the gender *Hopp (2013)*

Experiment: Syntactic gender predictions

How much slower are non-native predictions?

Could L2 speakers' native language impact prediction speed?

Three speaker groups

74 German native speakers

48 Spanish intermediate-advanced learners of German (L1 with gender)

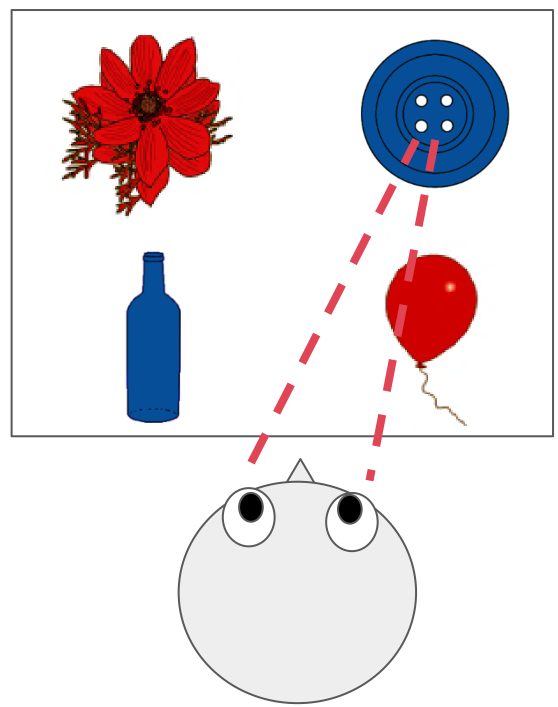
48 English intermediate-advanced learners of German (L1 without gender)

Martin and Sarah have to clean up the house before their parents get home

Martin and Sarah have to clean up the house before their parents get home

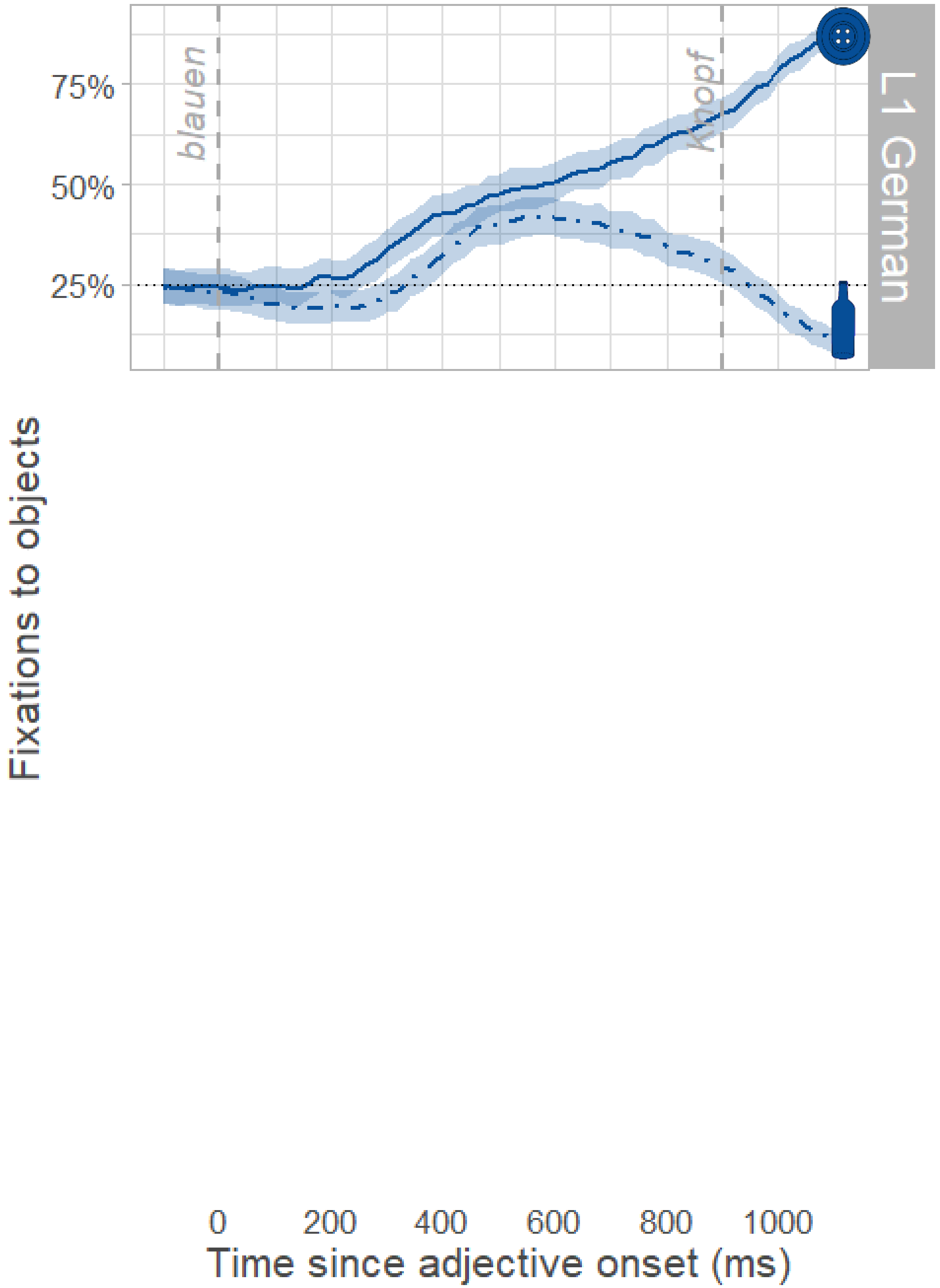
Critical window

“Klicke auf den	blauen	Knopf”
Click on the.$_{MASC}$	*blue.$_{MASC}$*	*button*



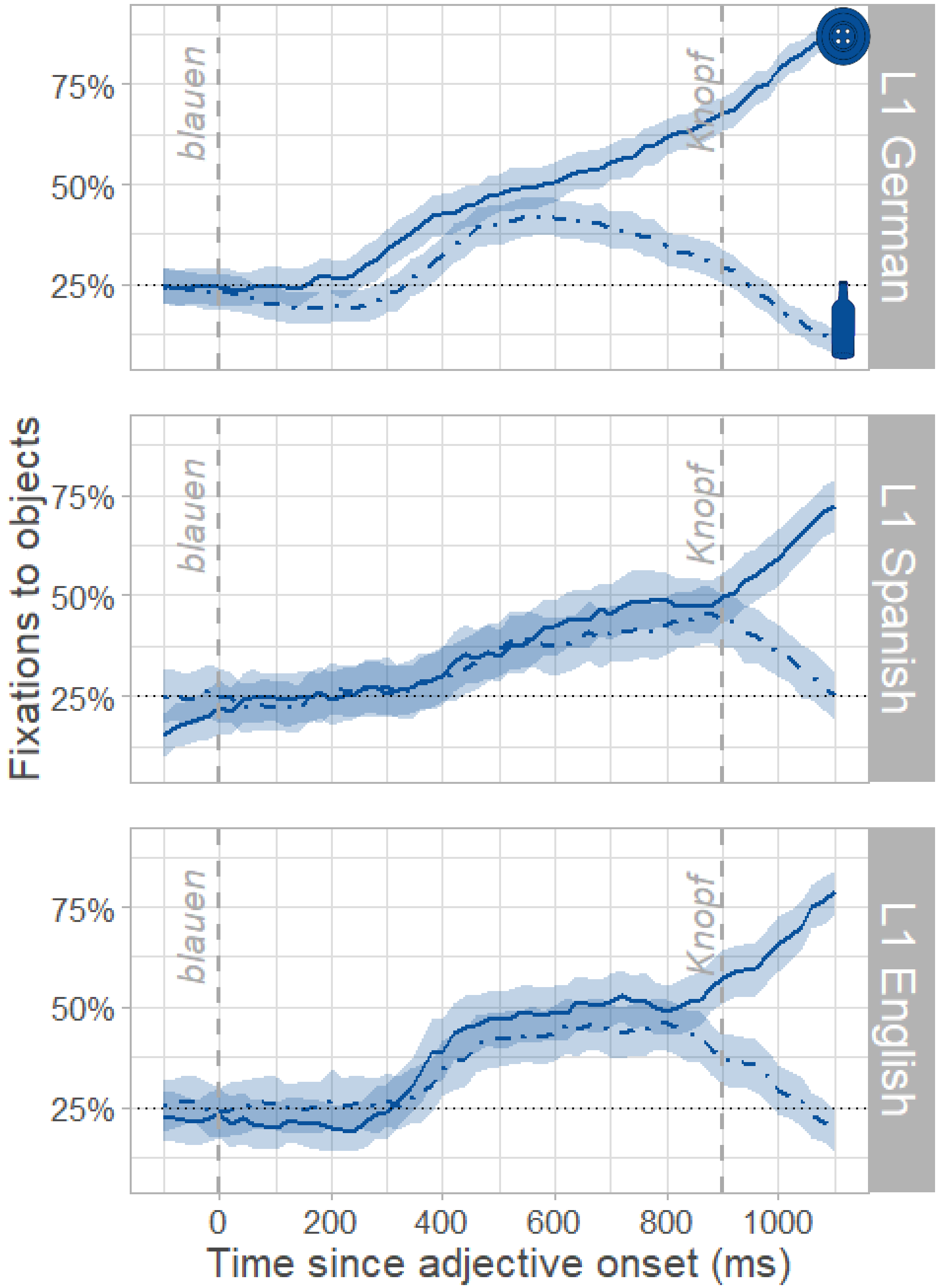
der Knopf.masc
TARGET

die Flasche.fem
COMPETITOR



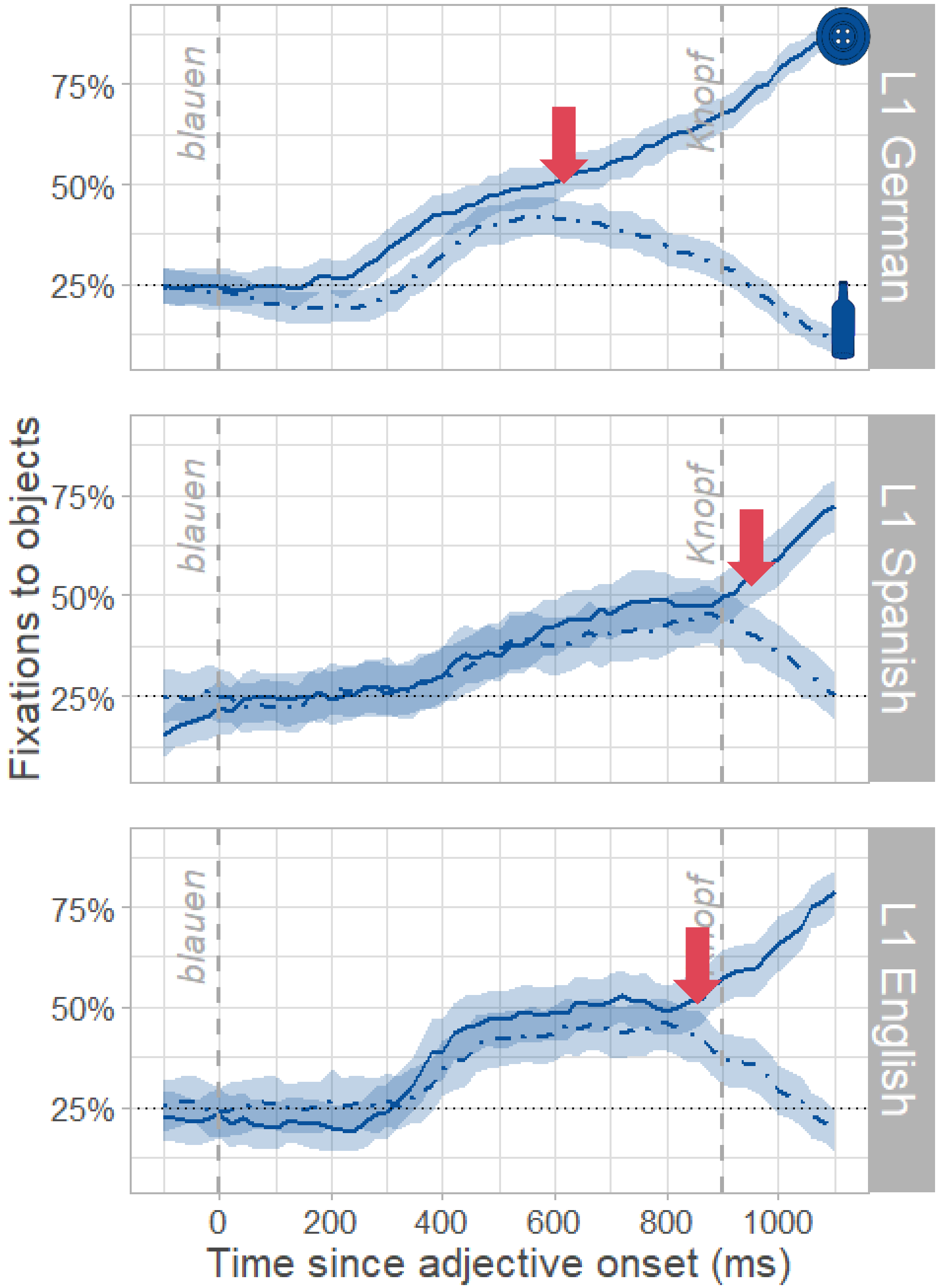
Results

- Native speakers predicted the noun



Results

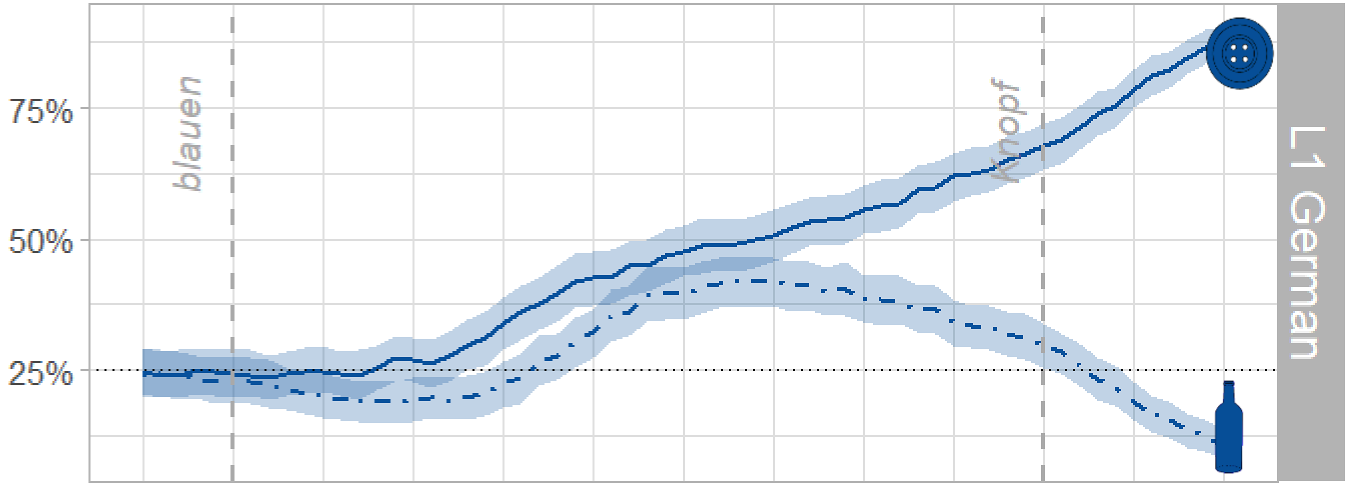
- Native speakers predicted the noun
- Non-native speakers maybe predicted
- Slower predictions in non-native speakers



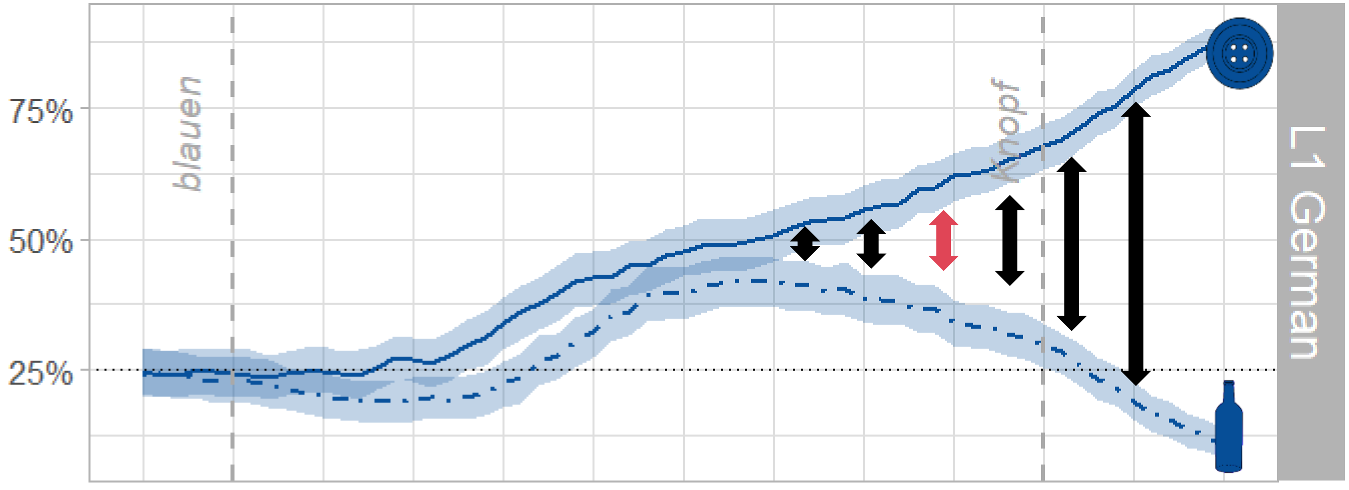
Results

- Native speakers predicted the noun
- Non-native speakers maybe predicted
- Slower predictions in non-native speakers

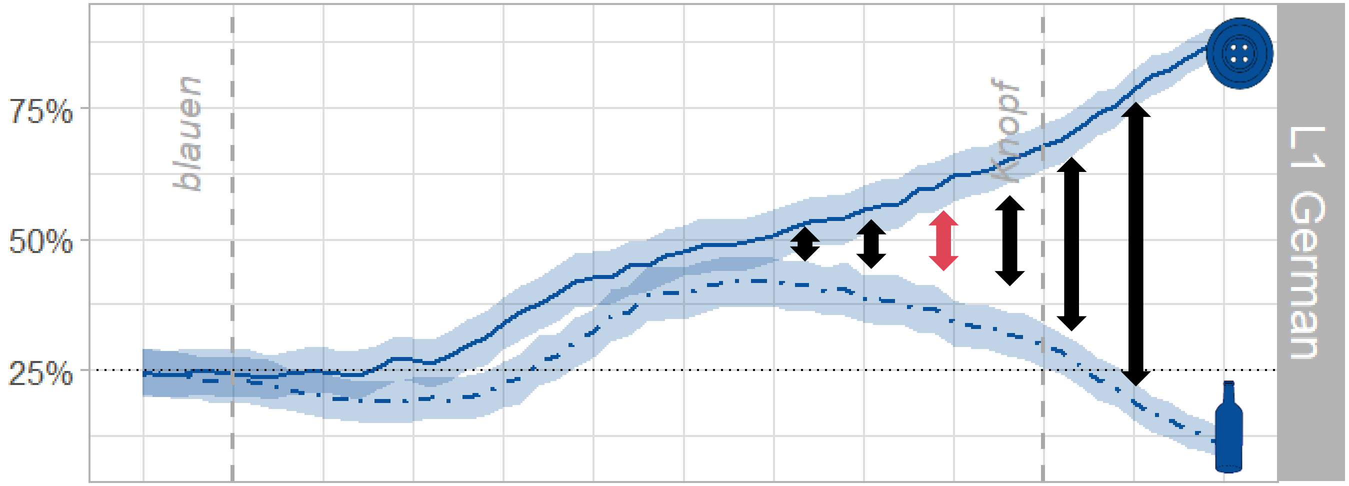
Finding a divergence point



Finding a divergence point



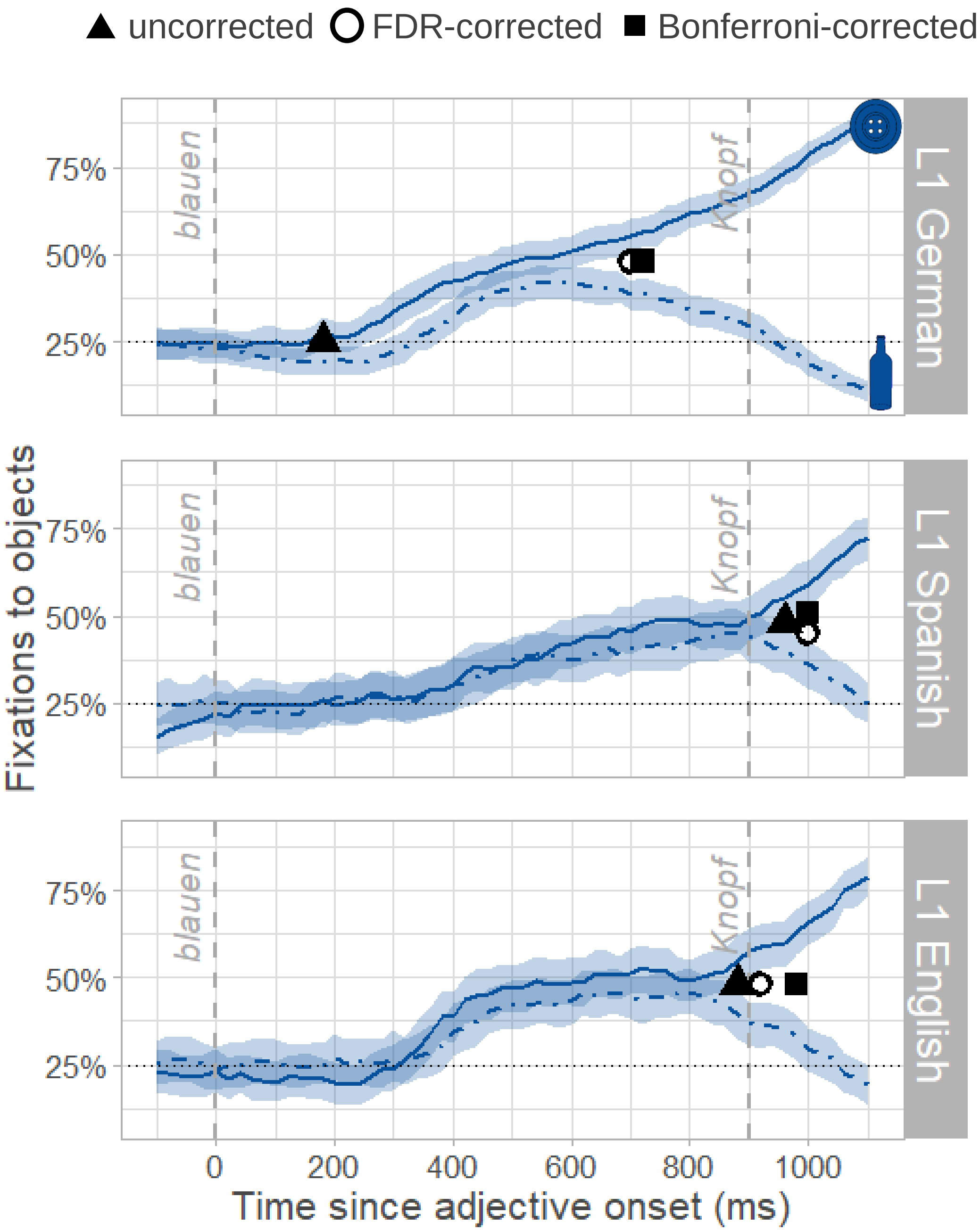
Problem: Multiple comparisons



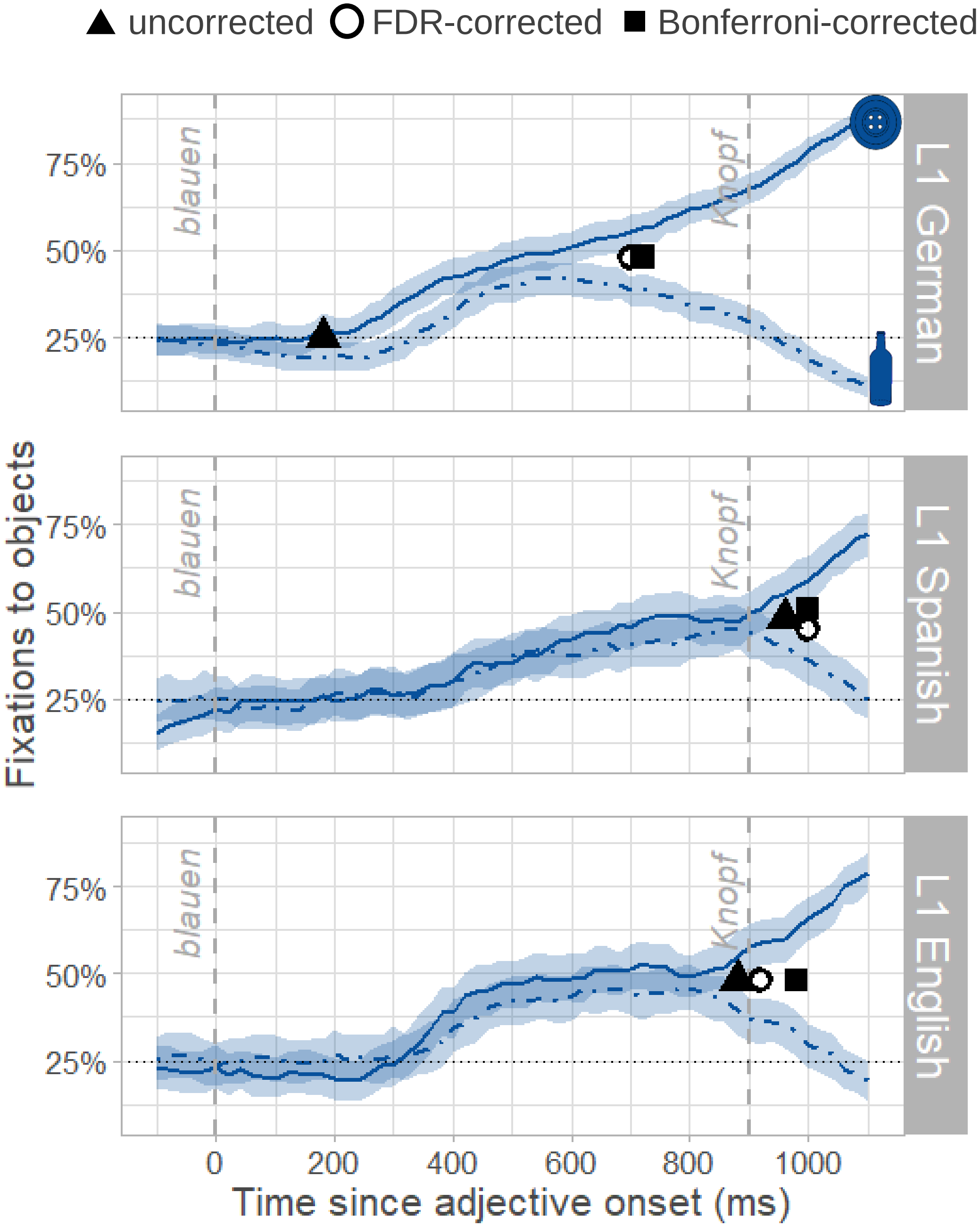
- Number of statistical comparisons = number of timepoints
- Type II error rate for 1 test = 5%
- 45 tests: 90% ($1 - 0.95^{45}$)
- Familywise error rate (FWER)

Bonferroni correction

False discovery rate (FDR) control



uncorrected
FDR-corrected
Bonferroni-corrected
Fixations to objects
75%
50%
25%
blauen
Knopf
L1 German
L1 Spanish
L1 English
0
200
400
600
800
1000
Time since adjective onset (ms)

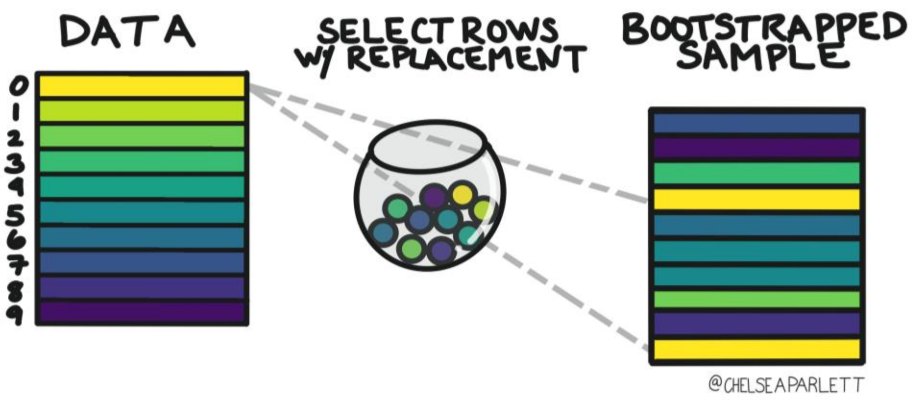


How to compare the groups?

Bootstrapping

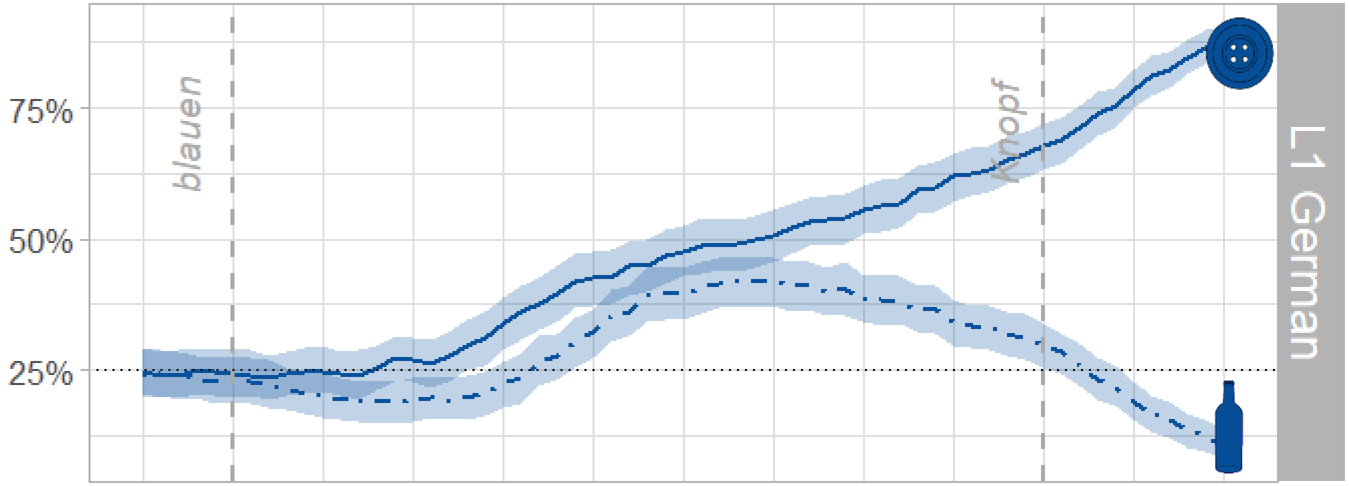
- Estimate the sampling distribution of our divergence point estimate.
- What would the divergence point estimate be if we did the experiment again?
- What about 2000 times?

Bootstrapping



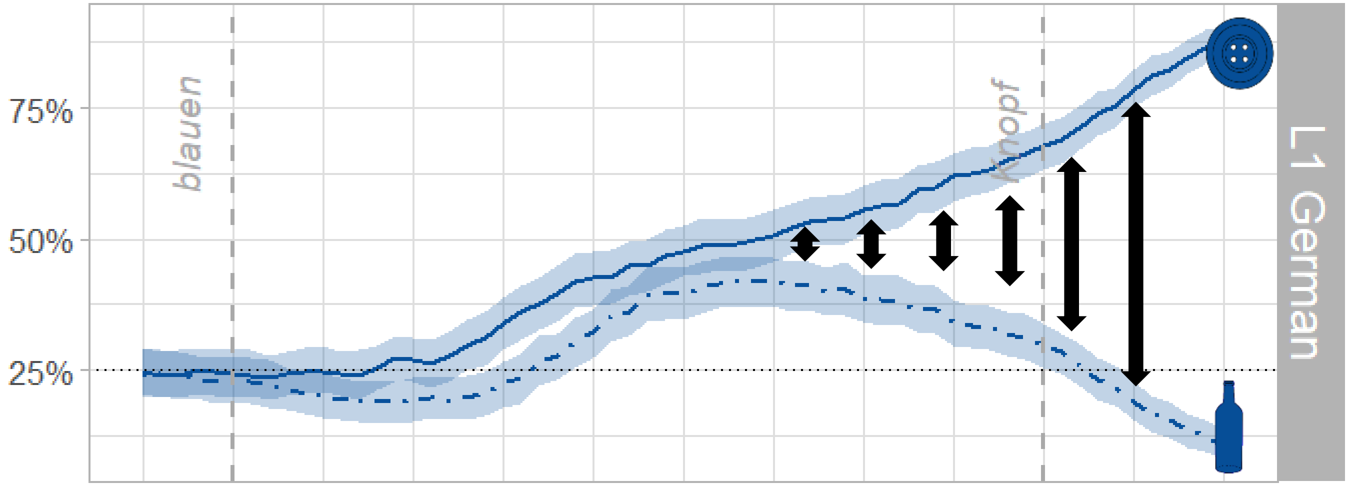
- Estimate the sampling distribution of our divergence point estimate.
- What would the divergence point estimate be if we did the experiment again?
- What about 2000 times?

Bootstrapping the divergence point



Stone, Lago & Schad (2020)

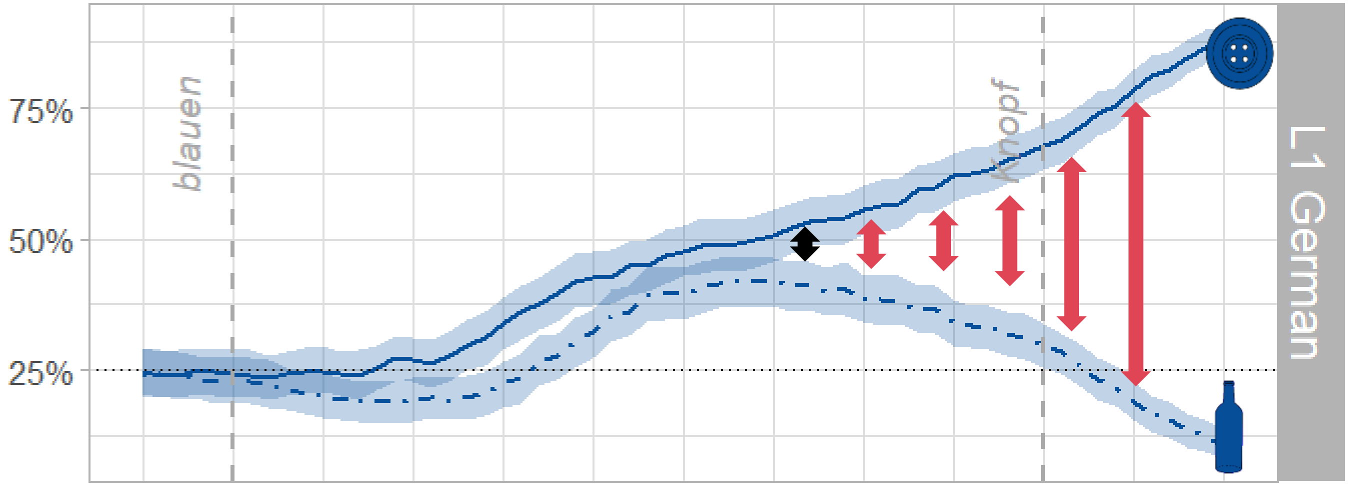
Bootstrapping the divergence point



Steps:

1. Test between curves at each timepoint

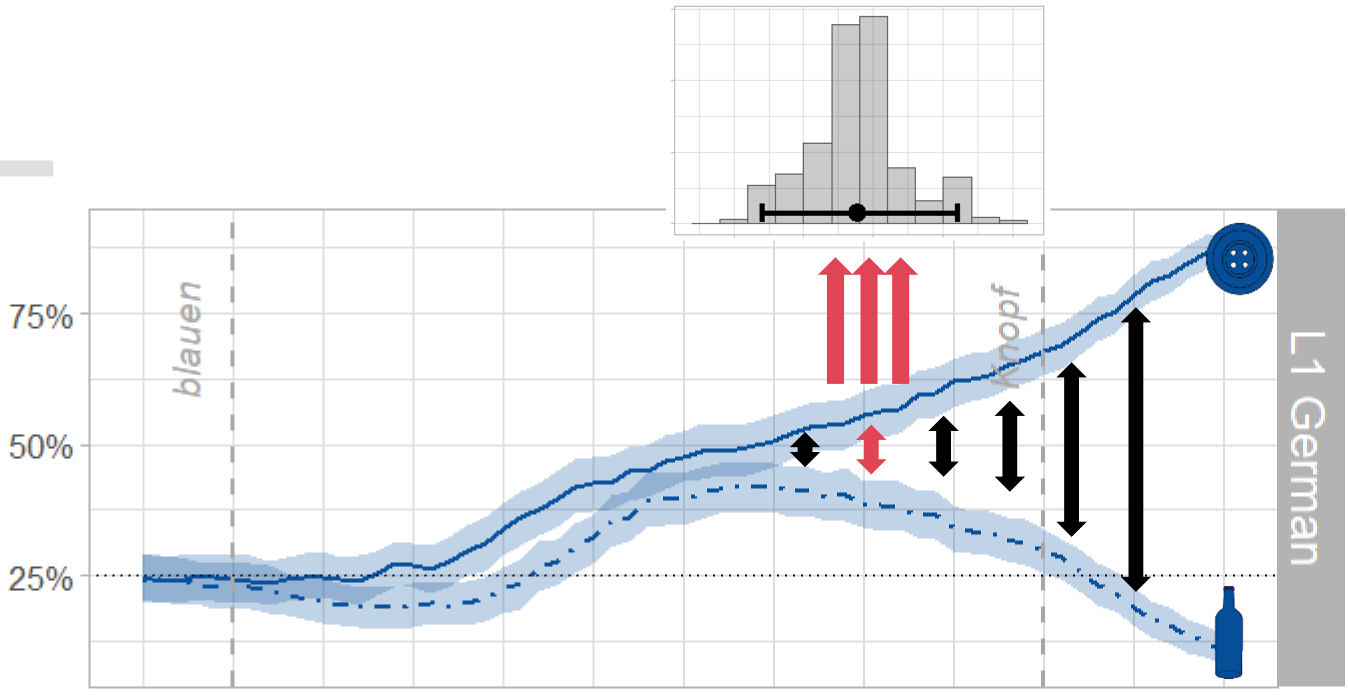
Bootstrapping the divergence point



Steps:

1. Test between curves at each timepoint
2. Find the **first** significant test statistic in a run of five

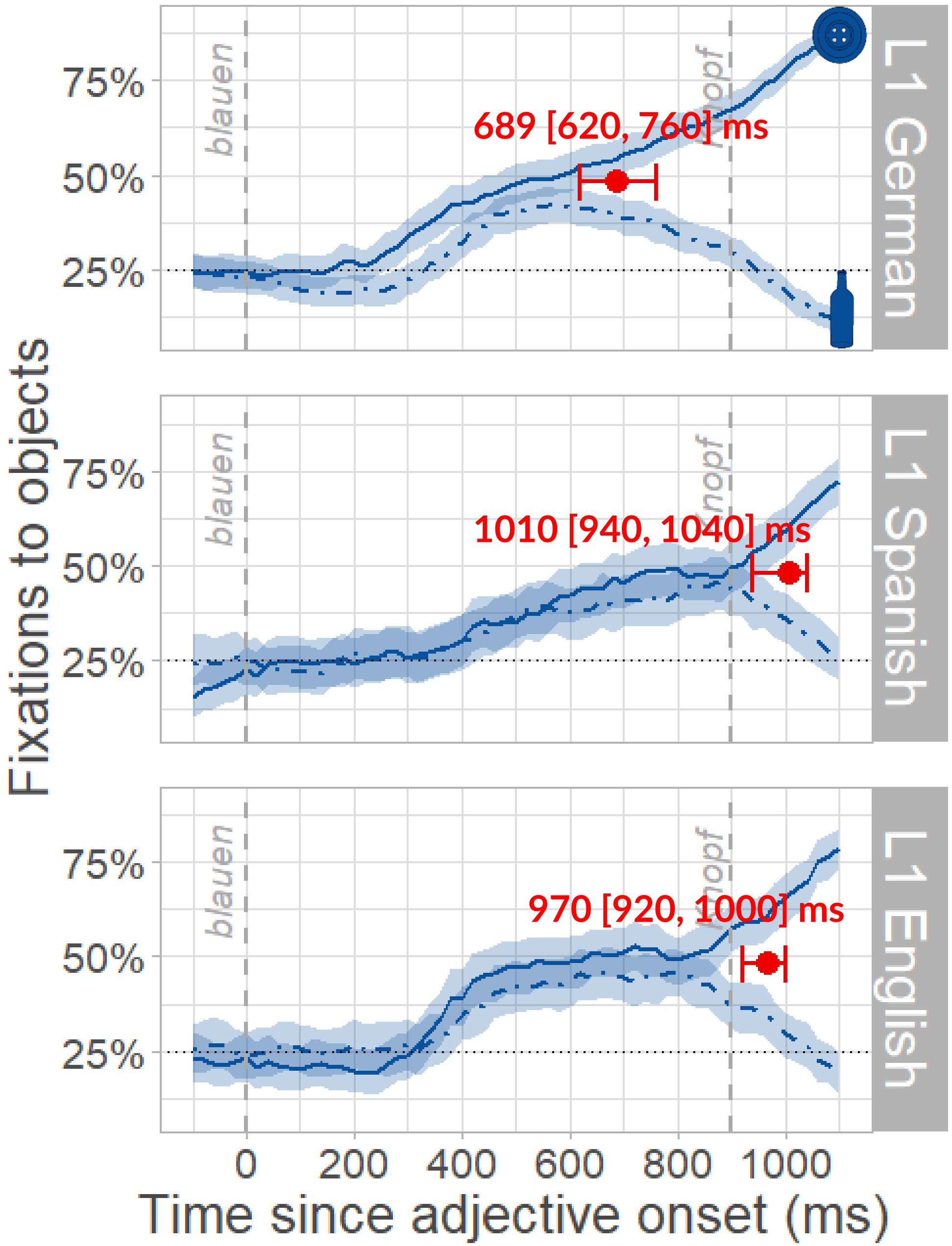
Stone, Lago & Schad (2020)



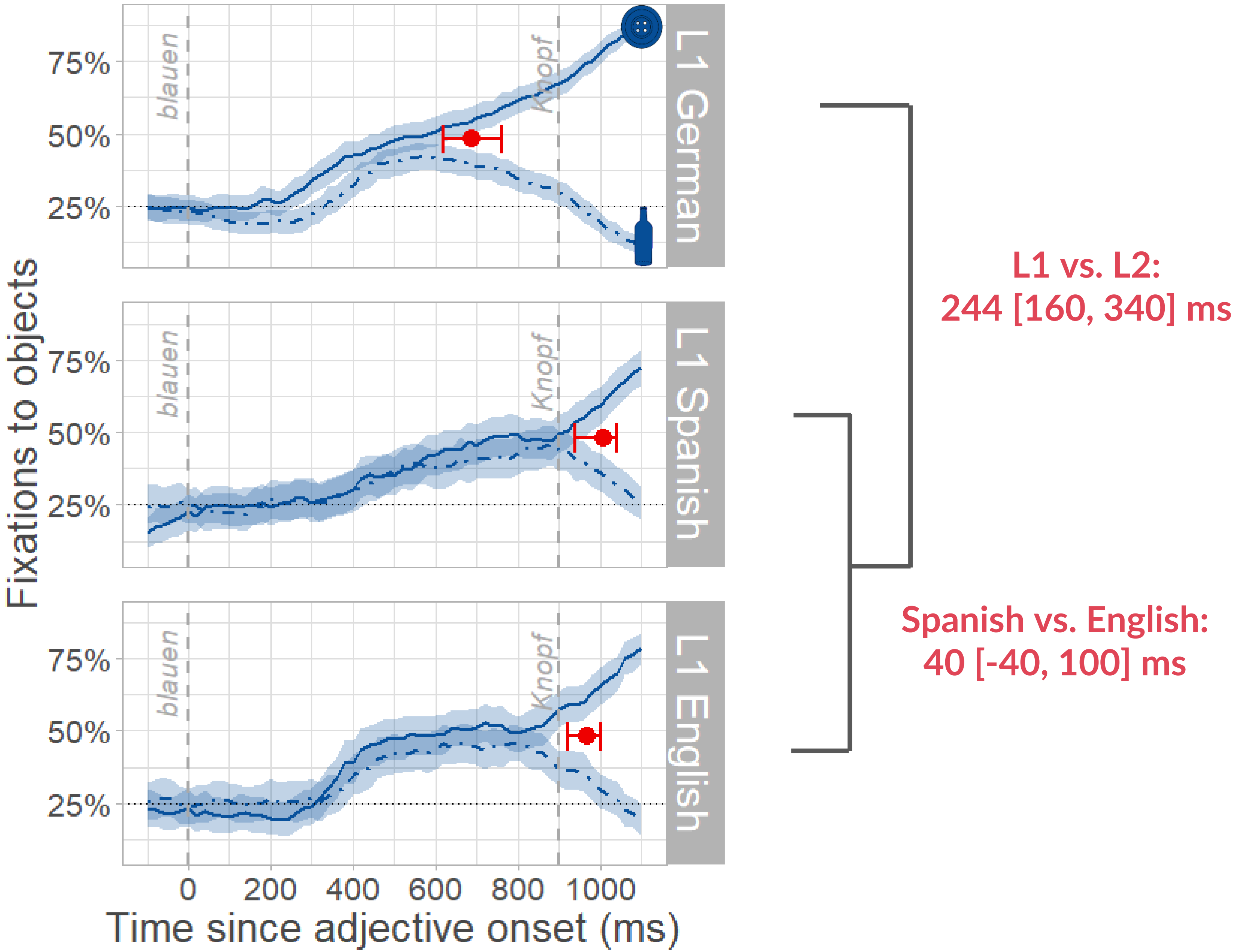
Steps:

1. Test between curves at each timepoint
2. Find the **first** significant test statistic in a run of five
3. Resample the data, repeat 2000 times

Stone, Lago & Schad (2020)



Fixations to objects
75%
50%
25%
blauen
Knopf
689 [620, 760] ms
L1 German
1010 [940, 1040] ms
L1 Spanish
970 [920, 1000] ms
L1 English
0
200
400
600
800
1000
Time since adjective onset (ms)

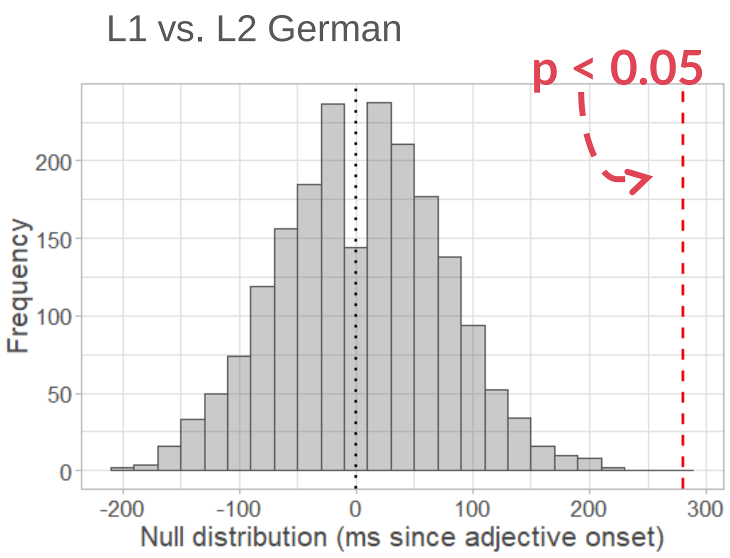


Fixations to objects
75%
50%
25%
blauen
Knopf
L1 German
L1 Spanish
L1 English
0
200
400
600
800
1000
Time since adjective onset (ms)
L1 vs. L2:
244 [160, 340] ms
Spanish vs. English:
40 [-40, 100] ms

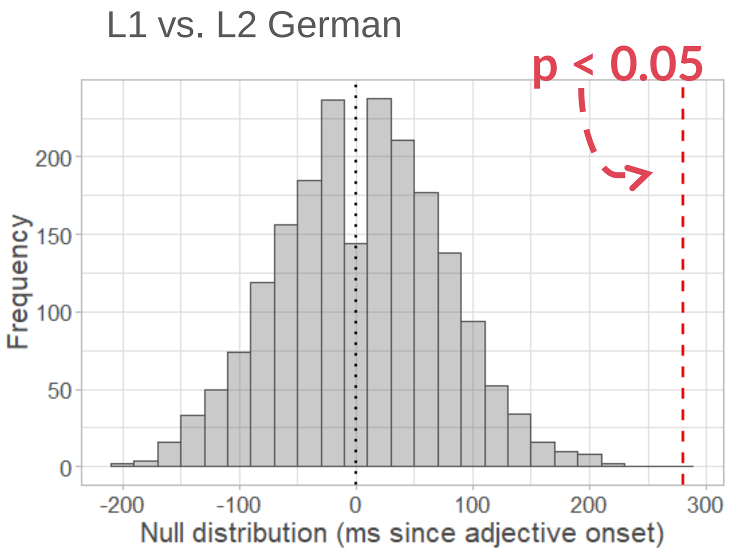
Computing a p-value

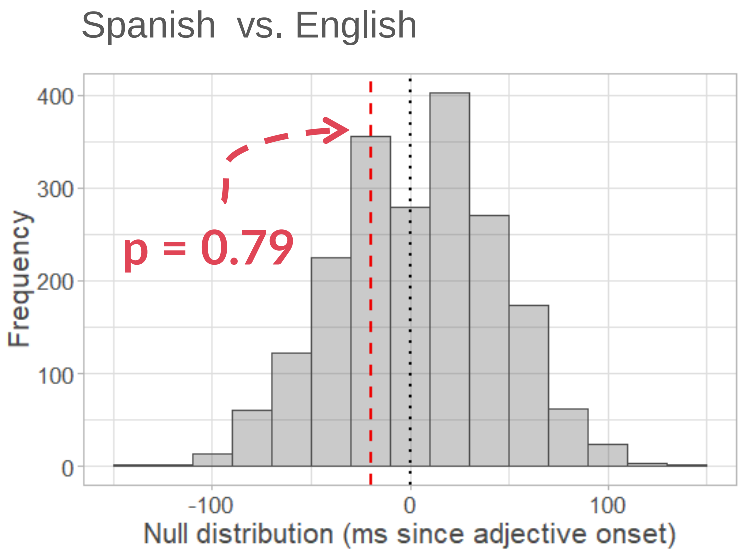
- Randomly reshuffle the group labels (L1 German, L1 Spanish, L1 English)
- Apply the test procedure
- Find the distribution of onsets if “L1 group” was random (null distribution)
- Is the position of the onset difference > 2 SDs from the mean of the null?

Computing a p-value



Computing a p-value





Conclusions

Onset estimate + temporal uncertainty

- FWER control
- Autocorrelation control

Paper:
https://doi.org/10.1017/S1366728920000607

Data and code:
https://osf.io/exbmk/

Statistically compare onsets between groups

- L1 Germans predict 244 [160, 340] ms faster than L2 Germans, $p < 0.05$
- The 40 [-40, 100] ms difference in onset between L2 groups was not significant, $p > 0.05$
- L2 are slower to use syntactic constraints and native language doesn’t help!

What the method can and can't do

Method	Provides divergence point estimates?	Generative model?	Detects (✓) vs. assumes (x) effect?	Estimates uncertainty around a divergence point?	Can divergence points be statistically compared?
Bootstrapping	✓	x	x	✓	✓
Cluster permutation	x	x	✓	x	x
BDOTS	✓	✓	✓	x	x
GAMMs	✓	✓	✓	x	x

Table 1; Stone et al (2020)

Bootstrapping: Stone, Lago & Schad (2020)
Cluster permutation: Barr et al. (2014)
BDOTS: Seedorff et al. (2017)
GAMMs: van Rij, (2015); van Rij et al., (2020); Miwa & Baayen (2020)

Summary

Eye movements give insight into the timing of cognitive processing

Different temporal analyses suitable for different research questions

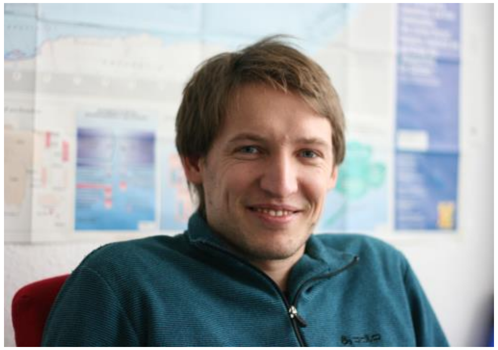
Allows direct test of theories predicting cognitive speed differences

Thank you!

stone@uni-potsdam.de

Sol Lago

Goethe University Frankfurt



Daniel Schad

Health and Medical University, Potsdam

João Veríssimo

Lisbon University

References

Barr et al. (2014) https://doi.org/10.1037/a0031813
Clahsen & Felser (2018) https://doi.org/10.1017/S0272263117000250
Dussias et al. (2013) https://doi.org/10.1017/S0272263112000915
Engbert et al. (2002) https://doi.org/10.1016/S0042-6989(01)00301-7
Frazier (1987) Theories of Sentence Processing, In: Modularity in Knowledge Representation and Natural-Language Understanding. Ed: Garfield, 291-308
Grüter et al. (2012) https://doi.org/10.1177/0267658312437990
Hopp (2013) https://doi.org/10.1177/0267658312461803
Hopp & Lemmerth (2018) https://doi.org/10.1017/S0272263116000437
Kukona et al. (2014) https://doi.org/10.1037/a0034903
Lim & Christianson (2014) https://doi.org/10.1017/S0142716414000290
Liversedge et al. (1998) Eye Movements and Measures of Reading Time, In: Eye Guidarice in Reading and Scene Perception. Ed: Underwood.
MacDonald et al. (1993) https://doi.org/10.1037/0033-295X.101.4.676
Miwa & Baayen (2020) https://doi.org/10.1017/S1366728921000079
Reingold et al. (2012) https://doi.org/10.1016/j.cogpsych.2012.03.001
Sassenhagen & Draschkow (2019) https://doi.org/10.1111/psyp.13335
Seedorf et al. (2017) https://doi.org/10.1016/j.jml.2018.05.004
Staub et al. (2007) https://doi.org/10.1037/0278-7393.33.6.1162
Stone, Lago & Schad (2020) https://doi.org/10.1017/S1366728920000607
Stone et al (2021) https://doi.org/10.1080/23273798.2021.1921816
van Rij (2015) https://jacolienvanrij.com/Tutorials/GAMM.html
van Rij et al. (2020) https://doi.org/10.1515/9783110440577-003
Yang & McConkie (2001) https://doi.org/10.1016/S0042-6989(01)00025-6