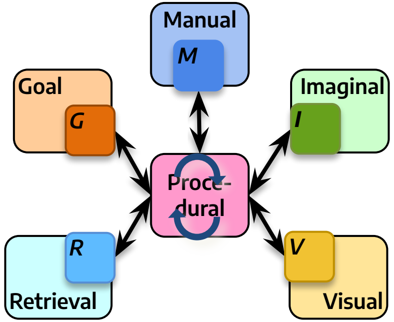
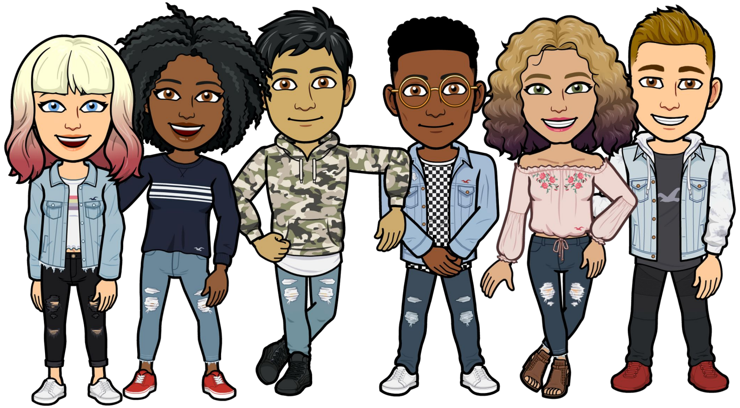
Decoding Idiographic ACT-R Parameters from Brain Data

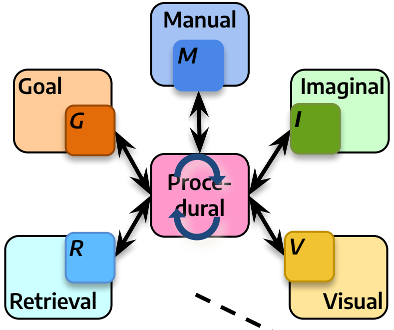
Andrea Stocco[1], Peiyun Zhou[1],
Florian Sense[2], & Hedderik van Rijn[2]
[1] University of Washington, Seattle
[2] University of Groningen, The Netherlands

Idiographic vs. Nomothetic

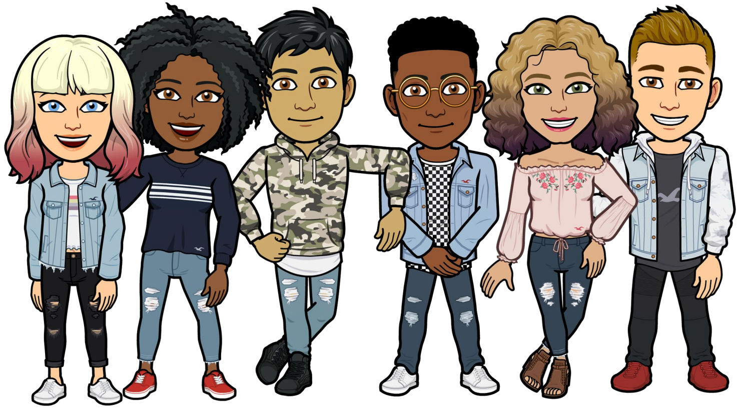




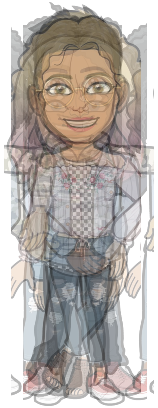
Idiographic vs. Nomothetic



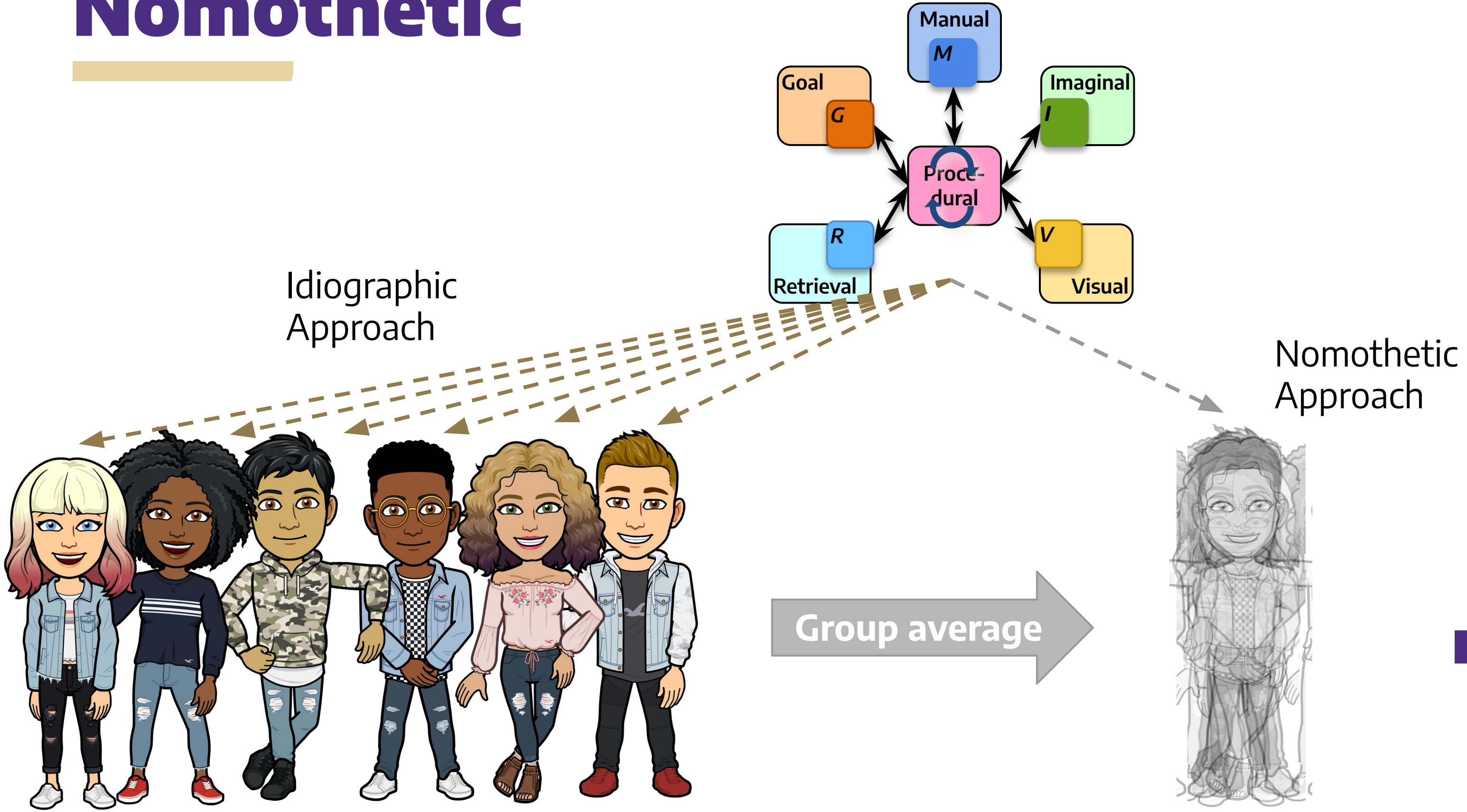
Nomothetic
Approach



Group average



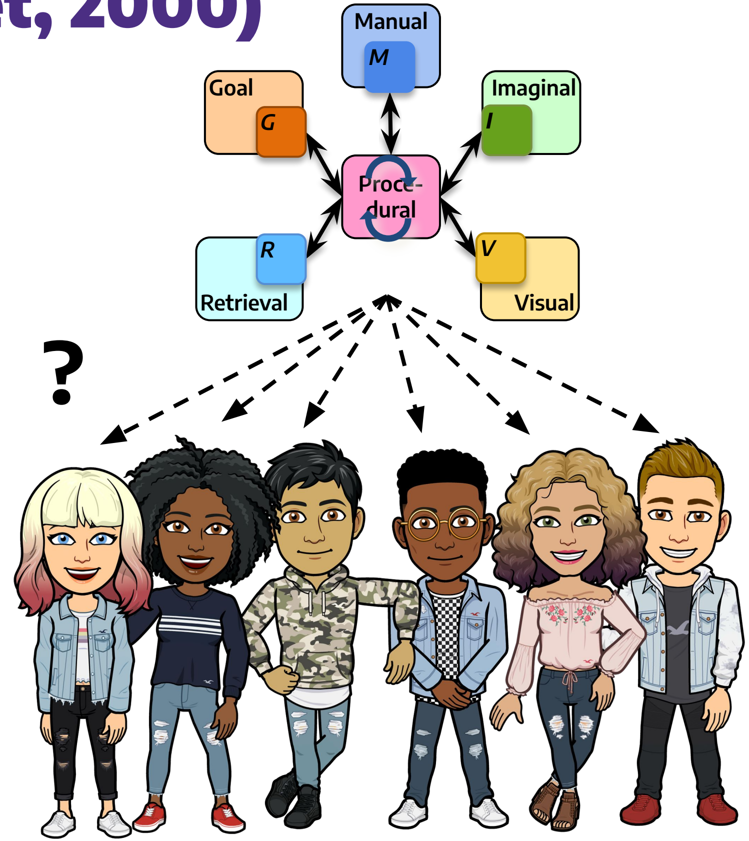
Idiographic vs. Nomothetic



This talk

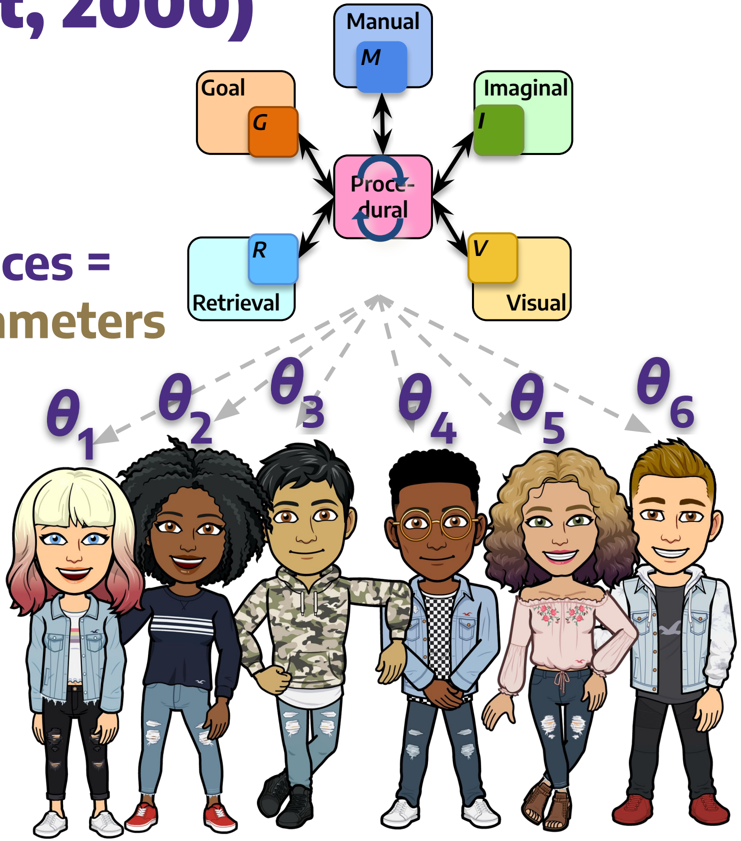
- Pilot data presented at the 2018 ACT-R workshop
- Newer and better results
- Displays the power of this approach (and a peek to the future)

A General Framework
(Ritter & Gobet, 2000)



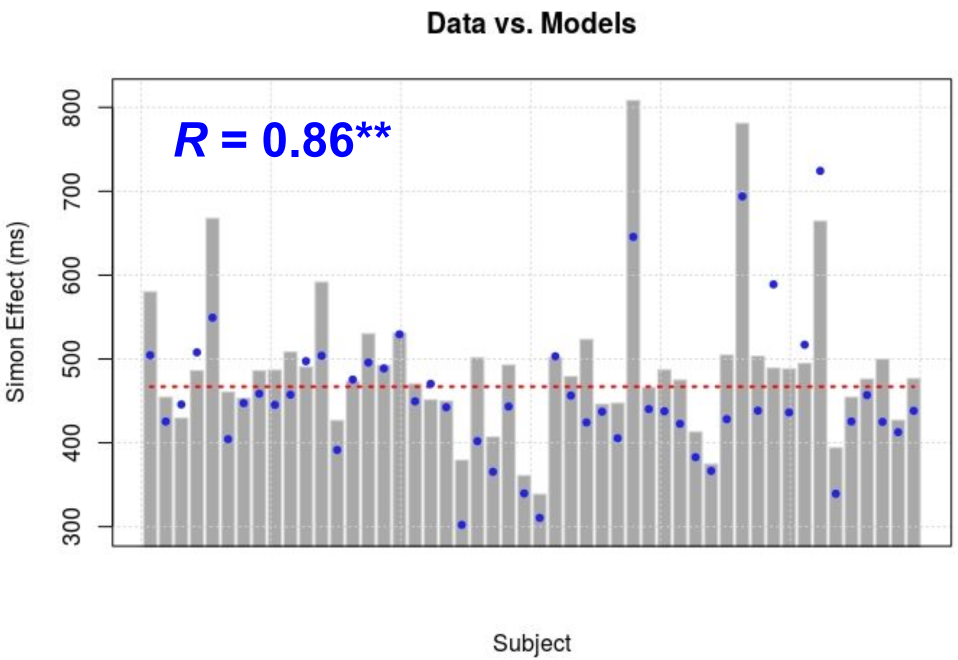
A General Framework
(Ritter & Gobet, 2000)

Individual differences = different sets of parameters



In-House Example

- Max likelihood to fit four parameters in a DM task
  - PSS task
- Plugged parameters in model of different task
  - Simon task
- Parameters predicted response times in incongruent trials

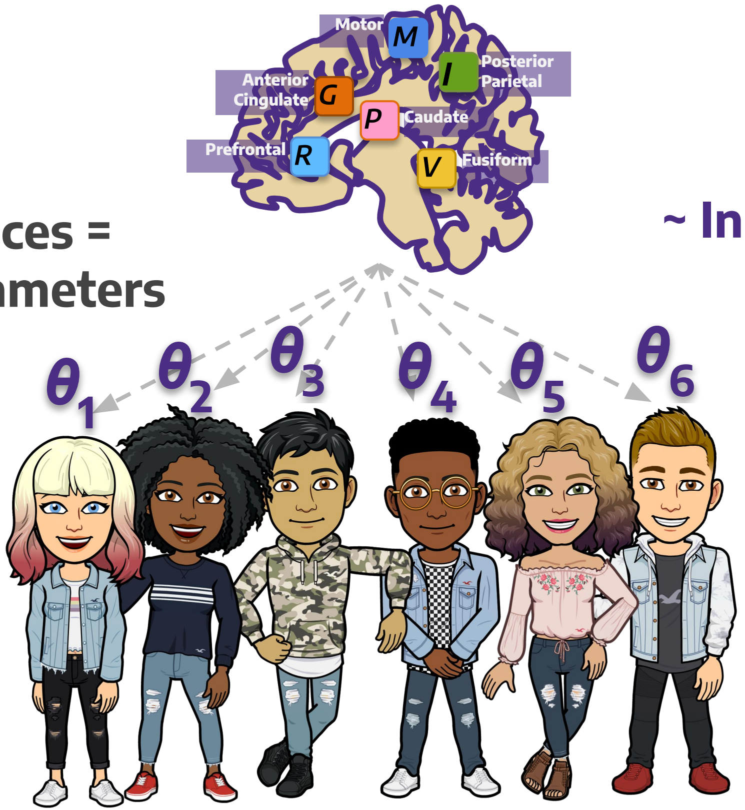


Task-Based Inference and Its limits

- Optimize parameters that fit a set of task data
- Depends on behavioral testing
  - can be long and complicated
  - Many many trials to get reliable measures
- Requires reasonable models of a task
  - Garbage in, garbage out
- Parameters should be the same across tasks
  - "Cognitive supermodels", à la Salvucci

A reductionist approach

Individual differences = different sets of parameters



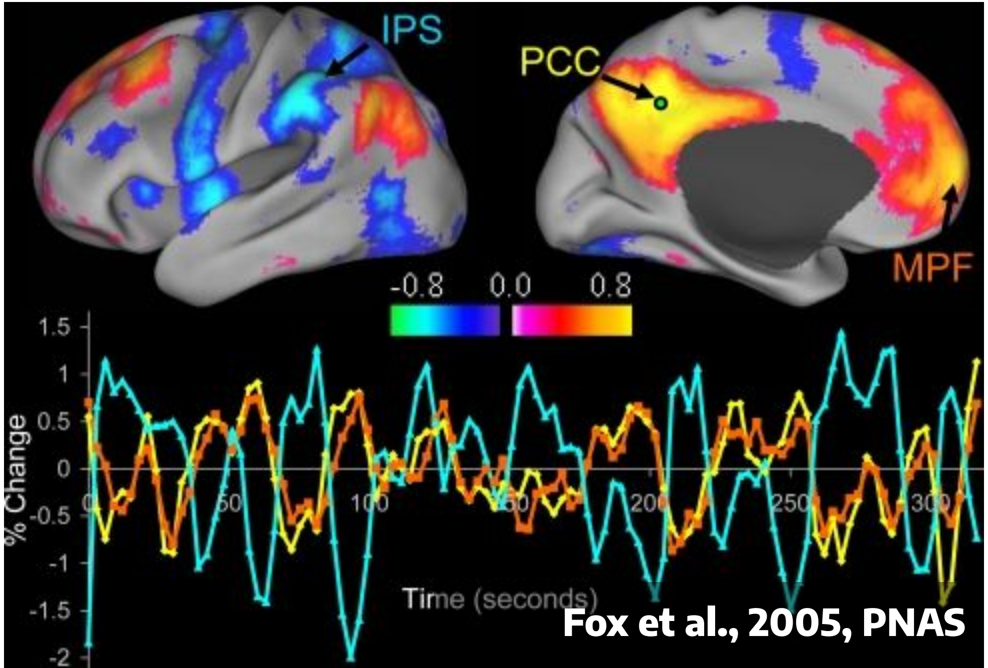
~ Individual differences in biology

What if we Could Bypass Behavior?

- Parameters should reflect basic neural activity
- **Task-free** neural measures exist
  - Anatomical MRI, DTI, SPECT/PET…
  - Most importantly, **resting state functional connectivity**

Resting-State Functional Connectivity

- Participants rest for ~5 mins while brain activity is recorded
- **Spontaneous** fluctuations in activity are highly organized
- Identify networks of stably connected regions
- Connectivity measures predict individual variables (Age, IQ).

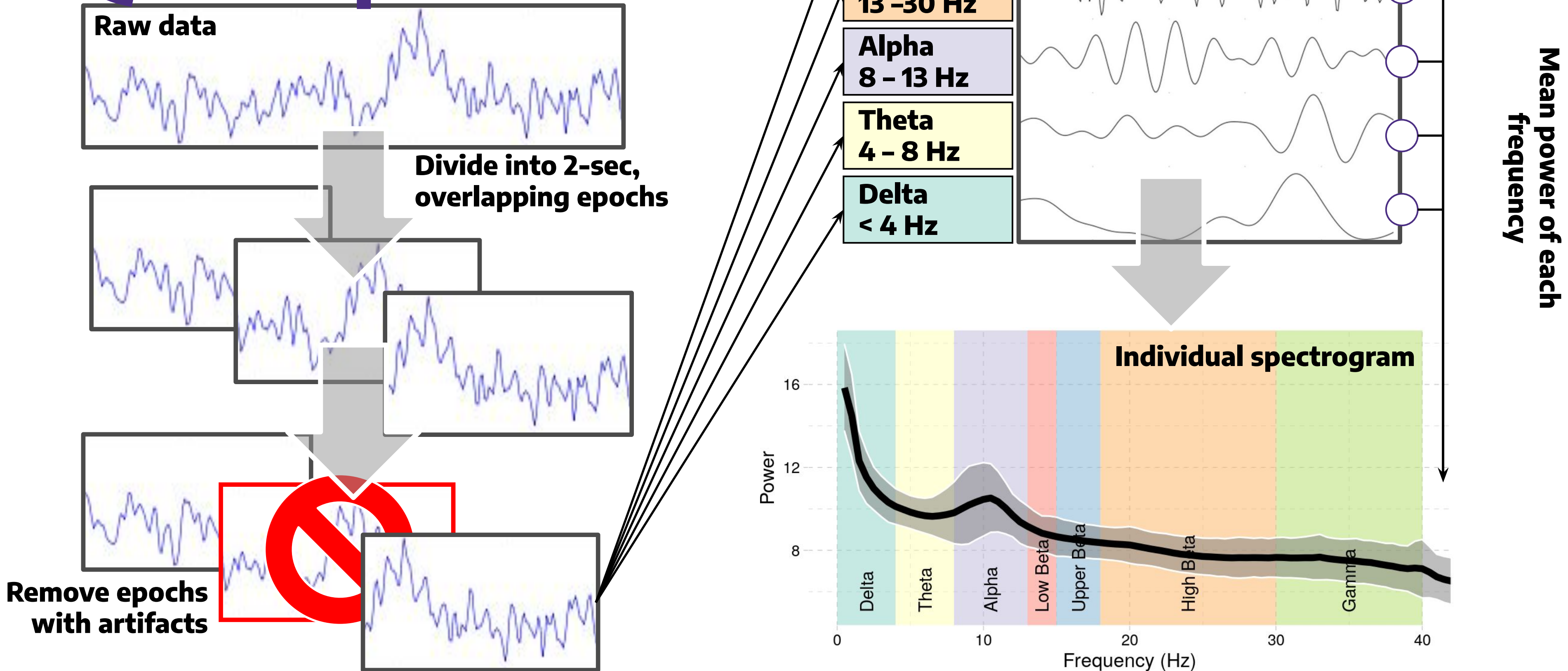


Fox et al., 2005, PNAS

EEG: Emotiv EPOC Headsets

- Reasonable price (< 1K)
- Decent characteristics
  - 14 channels @ 128 Hz
  - Frequently used for BCIs
- Easy:
  - Portable, wireless systems
  - Saline-based electrodes
  - ~15 mins for correct application
  - Minimal training required
  - Great for individual difference studies

QEEG Pipeline



Which Parameter?
Long-Term Memory Decay

- Perhaps the **cornerstone** of ACT-R
- The most fundamental **learning** mechanism
- Activation *A* is controlled by decay parameter *d*

$$A = \sum_j t_j^{-d}$$

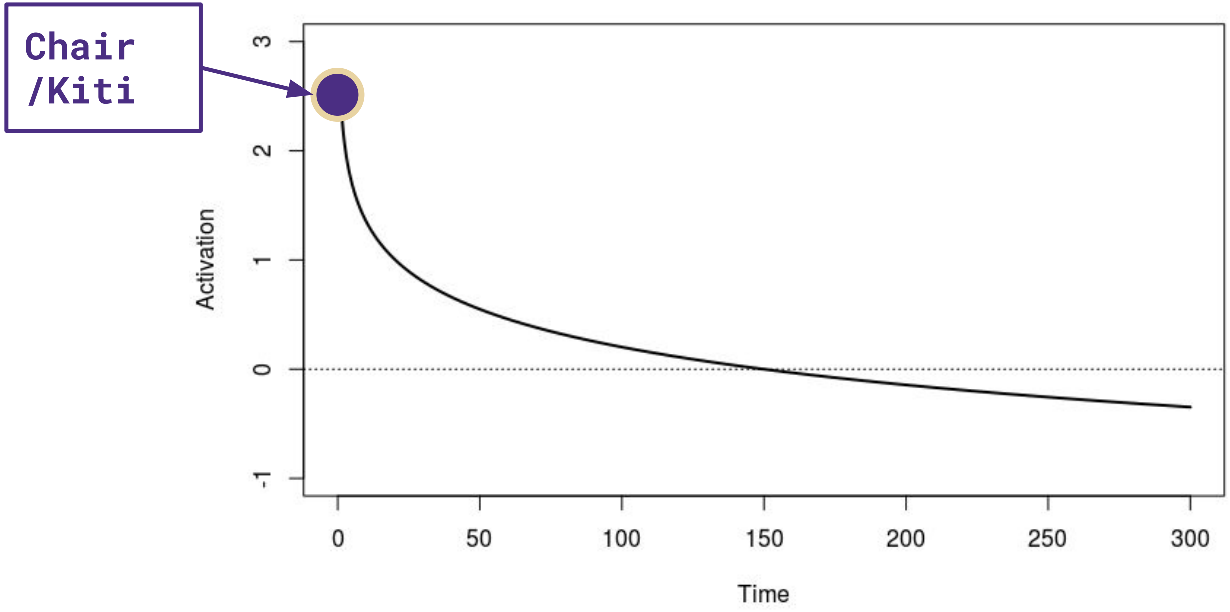
> Used Pavlik & Anderson's (2005) equation

$$A = \sum_j t_j^{-d}$$

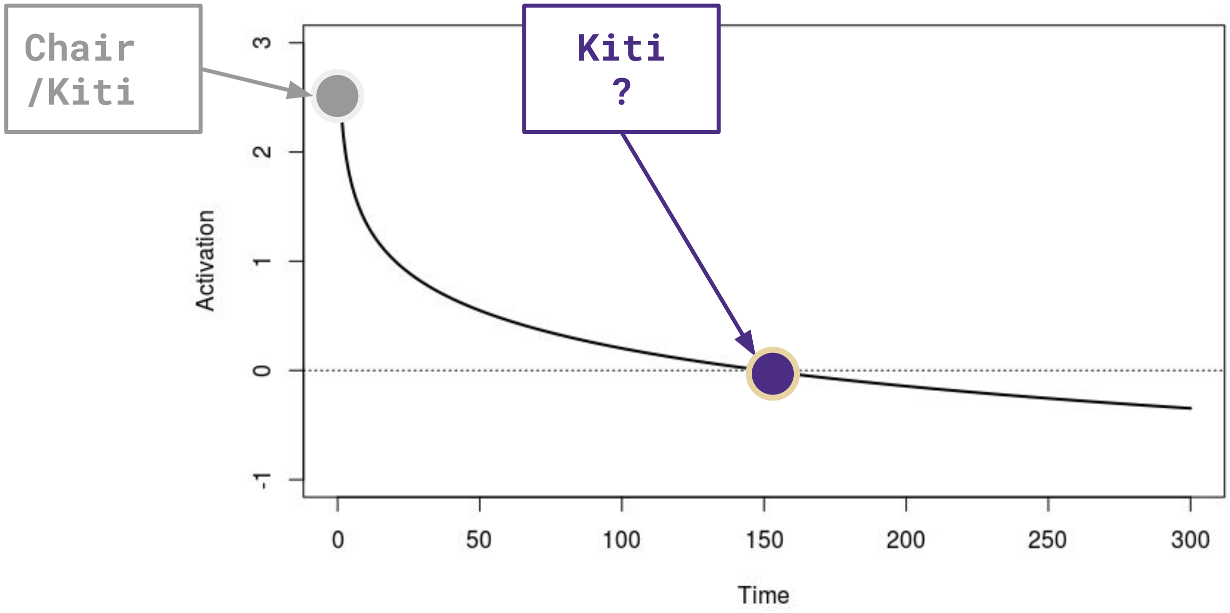
$$\boxed{d = ce^{A} + \alpha}$$

- Consistent across very short and very long intervals
- Accounts for spacing effects

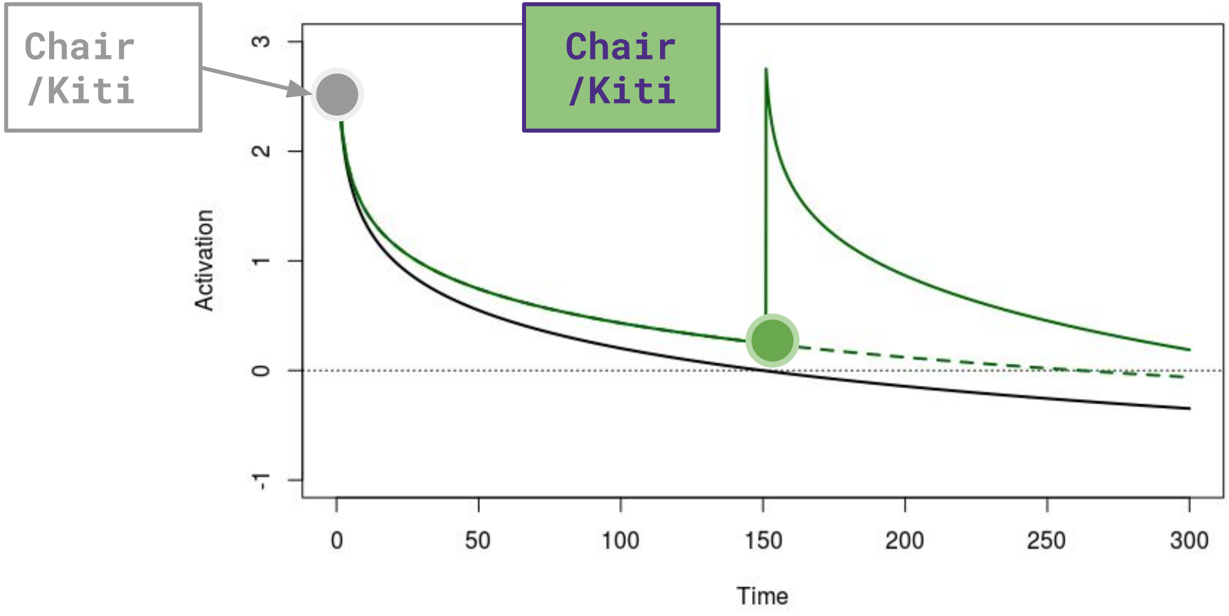
How is α Measured?



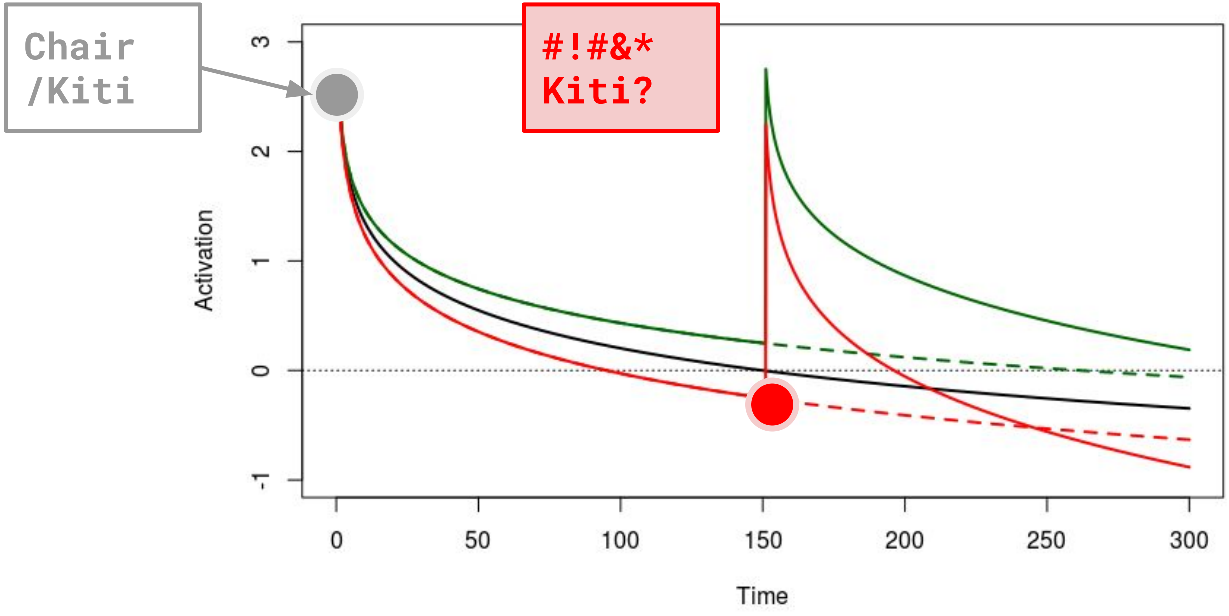
Predict when the chunk is forgotten



If the chunk is remembered, reduce α



If the chunk was forgotten, increase α



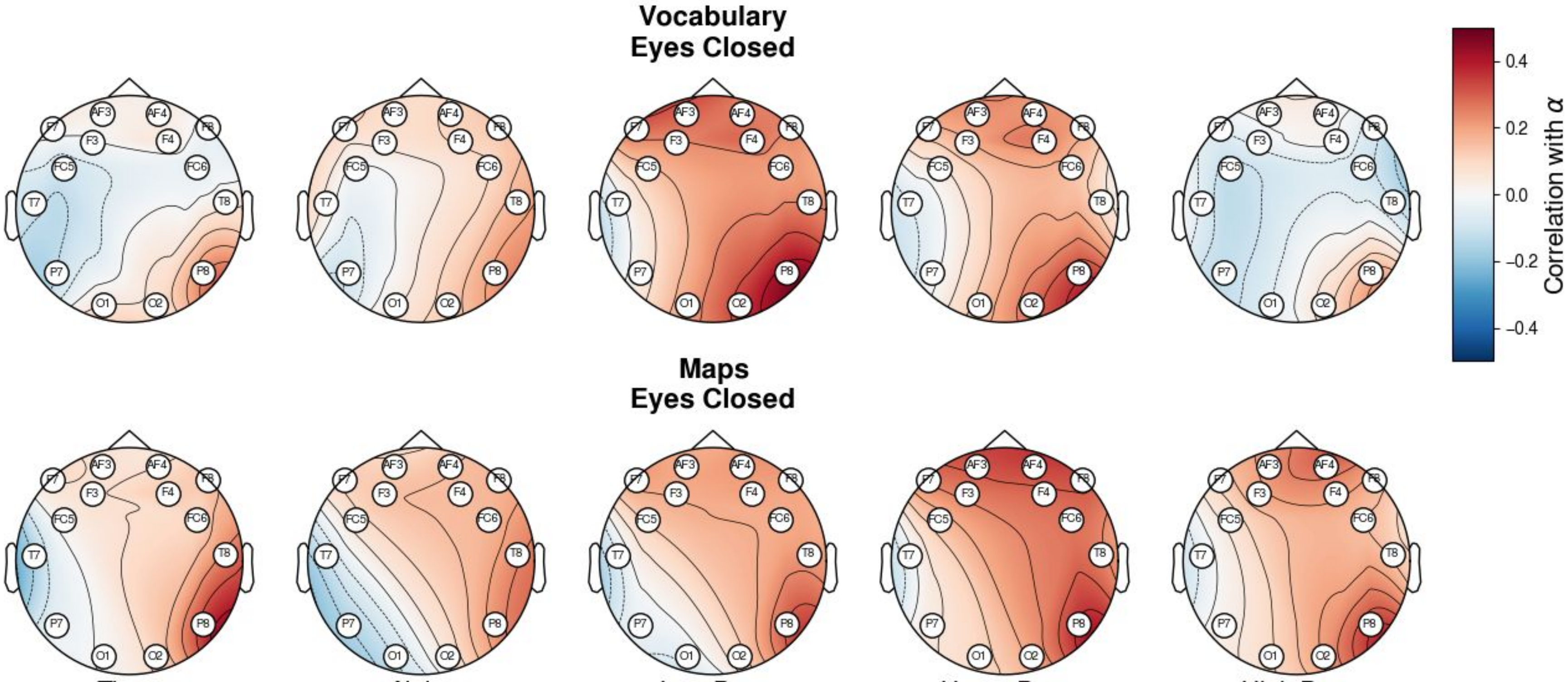
Reliability of estimates

- Hedderik and Florian have done **massive** work showing that ***α*** **…**
  - Can be estimated fast (~ 12 mins, even when participants do not make mistakes)
  - Is consistent across materials (visual, verbal, complex facts)
  - Is highly stable and reliable over time (test-retest reliability between 0.5 and 0.8)
  - (Sense et al., 2016, *TopiCS*)
- In essence, ***α*** is psychological "trait".

Can we decode α from Resting-State QEEG?

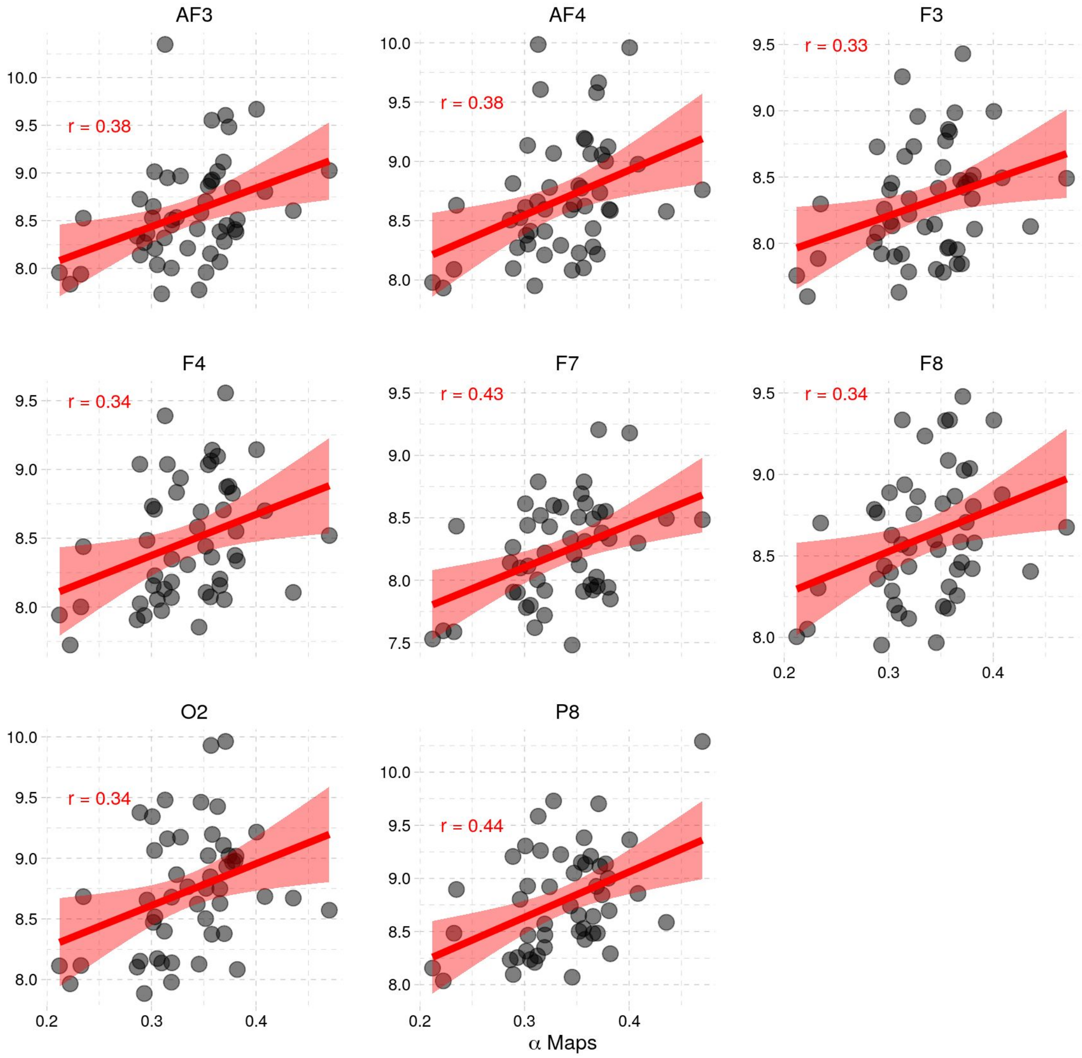
- *N* = 50 UW undergraduates
- All native English speakers
  - This is important!
- Collected 5 minutes of resting state, eyes closed EEG
- Learned 25 pairs of English-Swahili words
  - Same paradigm as Sense et al., 2016
- Learned 25 pairs of USA City-Maps associations
  - Same paradigm as Sense et al., 2016

Correlations Between α and QEEG Power

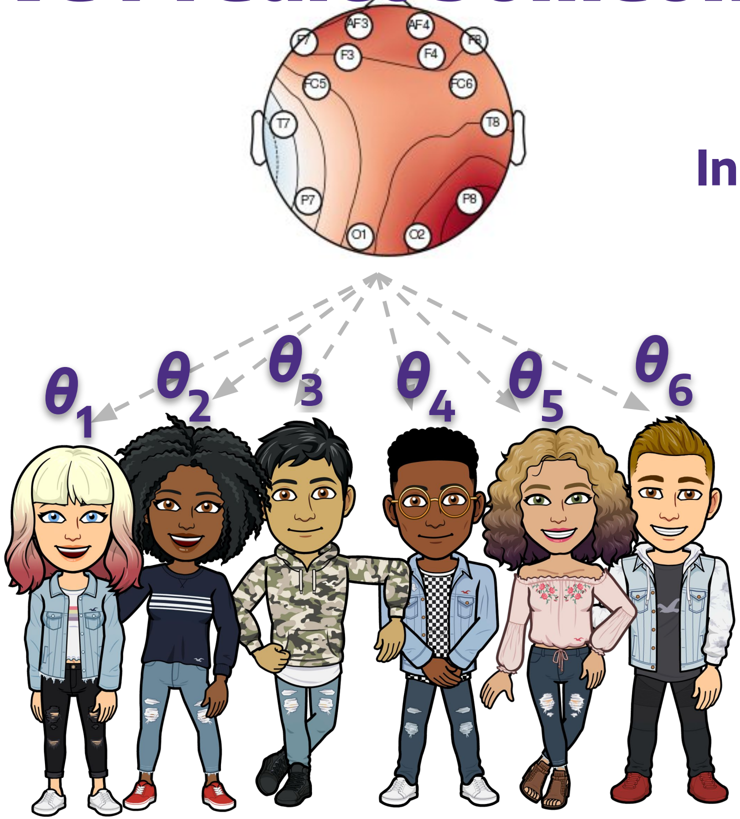


Correlations (FDR-corrected)

- **Many** tests (14 channels x 7 bands), so FDR was applied.
- Nine channels (Eight in Maps) still showed significant correlations after FDR
- All correlations in Low and Upper Beta band (13-18 Hz)



… But Can We Predict Someone's α?



Individual differences in biology

Different sets of parameters

Lasso

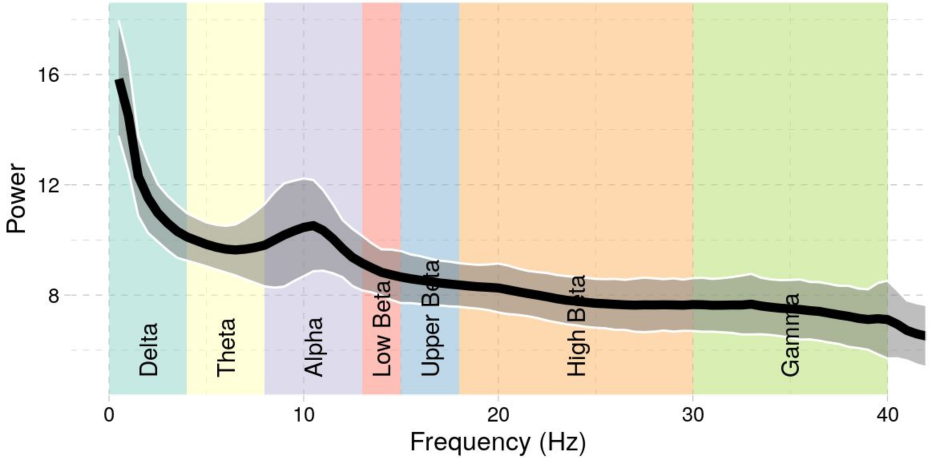
- Common and simple machine learning technique
- Linear regression built-in penalty to reduce the number of explanatory variables/overfitting:

$$\min(\|y - X\beta\|_2^2 + \lambda\|\beta\|_1)$$

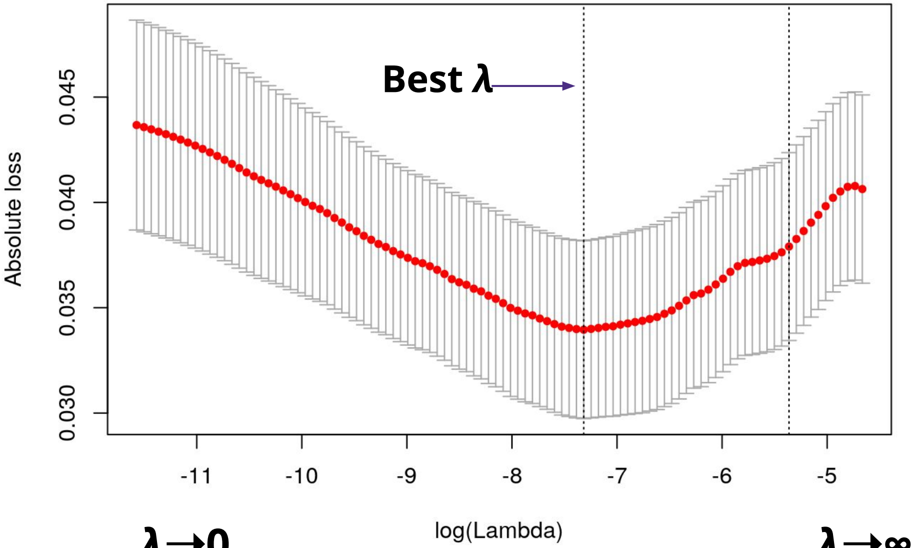
- Can be seen as a GLM with constrain $\|\beta\|_1 < k$ $(\propto 1/\lambda)$
- The term trades off λ accuracy and simplicity

Group Lasso

- When multiple variables are actually grouped together, it makes sense to constrain them in **groups**
- In the case of EEG, groups are naturally formed by frequency **bands**.



Model fitting and validation



λ→0 (simple regression)

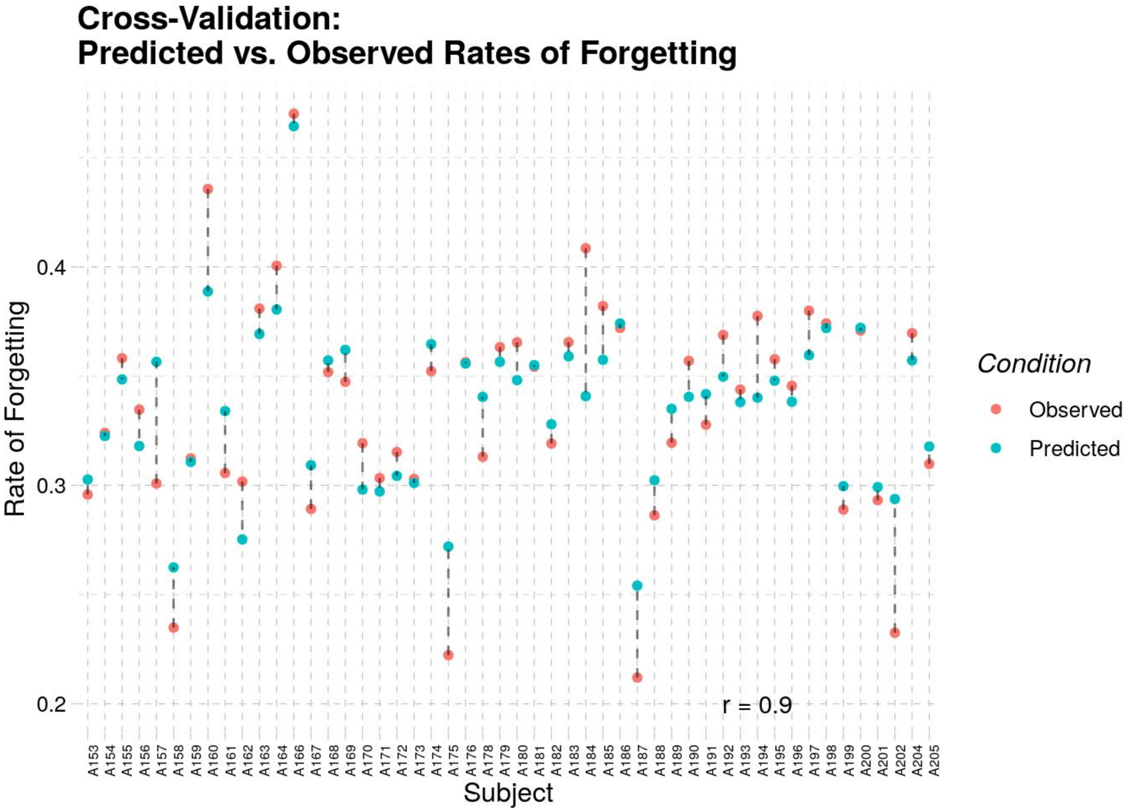
λ→∞ (only intercept)

> Used **cross-validation**
  – fit vs. generalizability

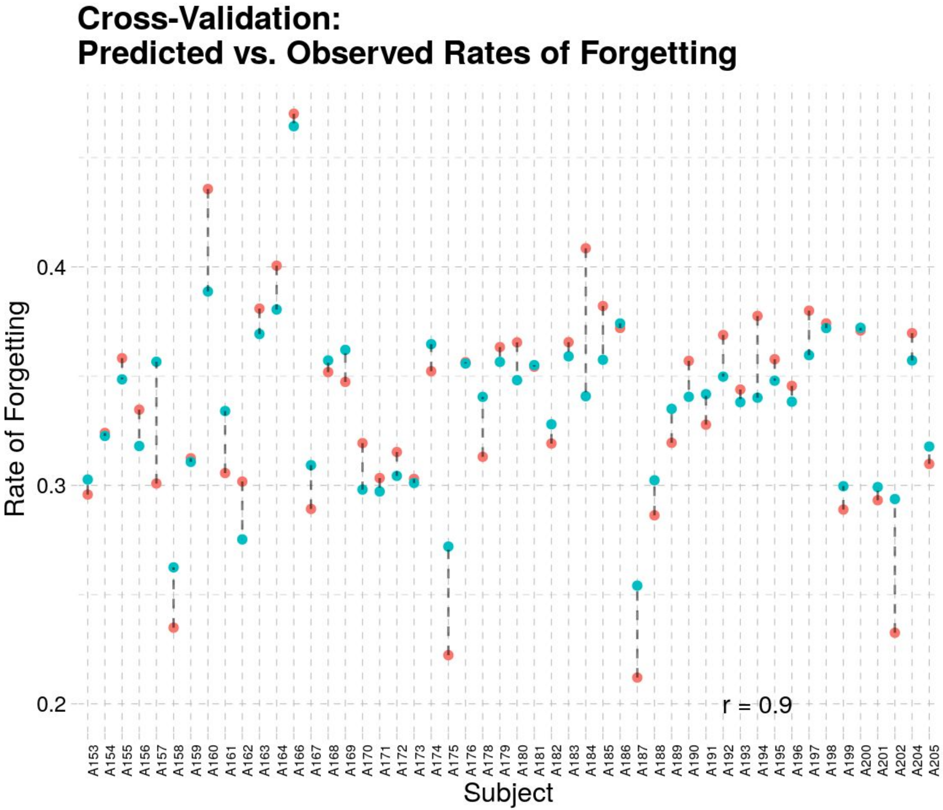
> Used Leave-One-Out (LOOV)

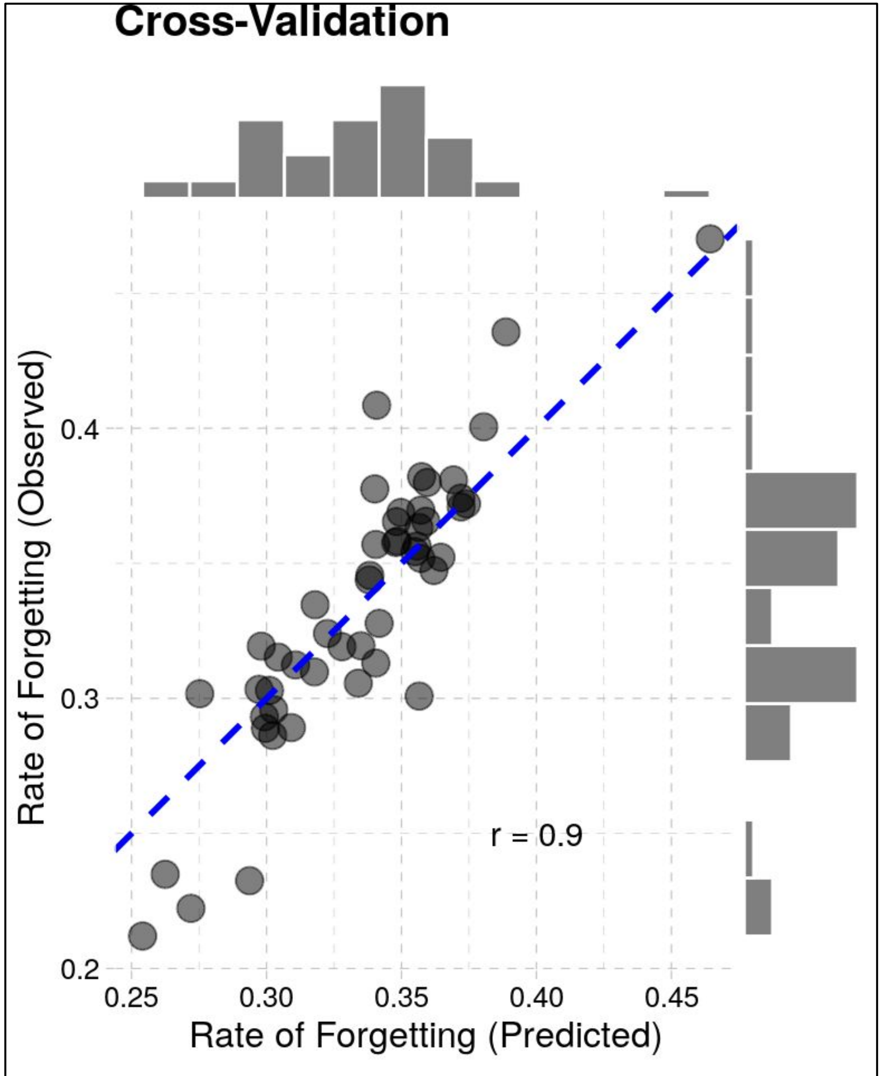
> Guarantees ideal lambda is actually best estimate (minimum absolute **loss**)

Predicted vs. Observed

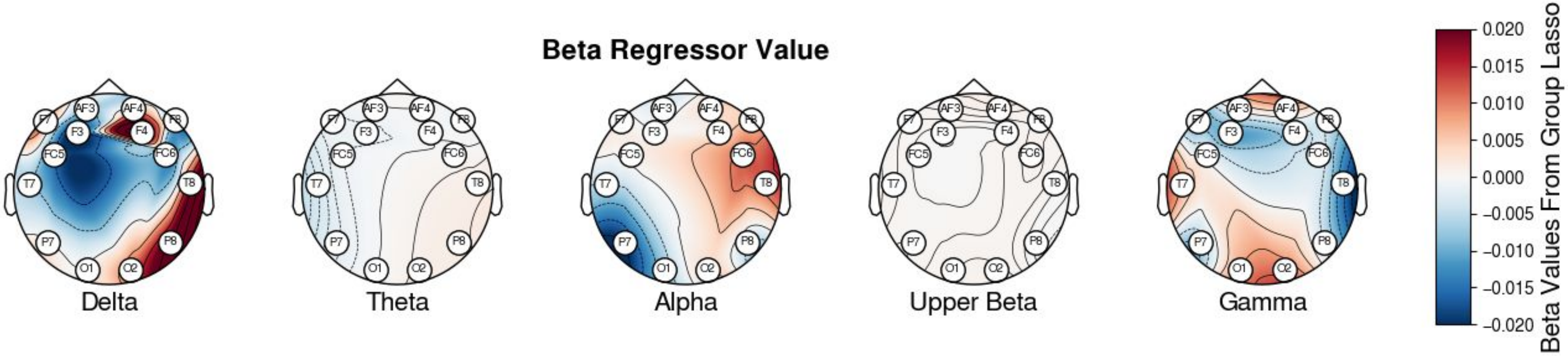


Predicted vs. Observed

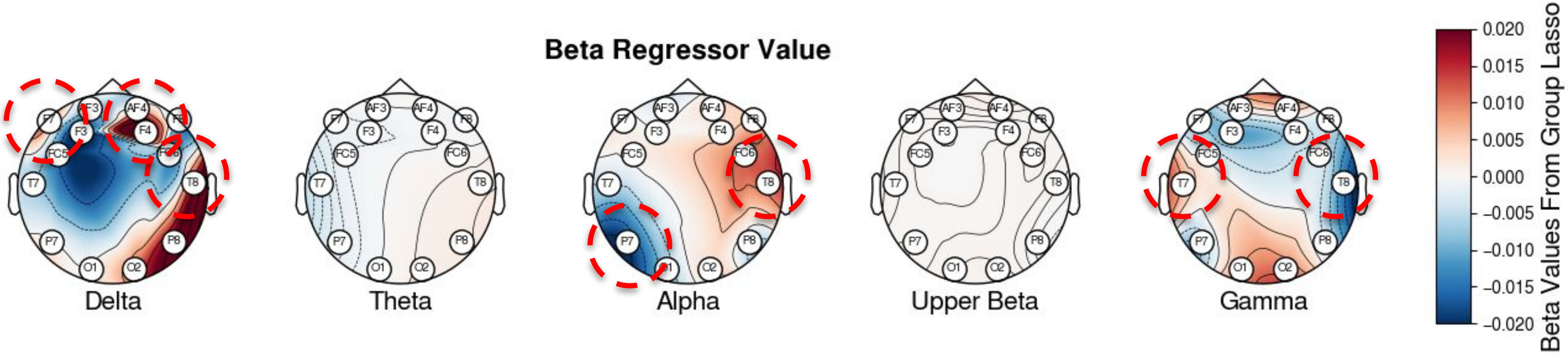




What does it look like?



What does it look like?



Why is this important?

- Forgetting rate has many practical applications
  - Learning skills, scholastic achievement…
  - See work by Hedderik, & Florian, Kevin Gluck, Michael Collins
- But it has implications for disorders that affect memory directly (Alzheimer’s) or indirectly (PTSD)
- **Marieke** will do a way better job talking about this!

Other research (aka: shout-out to my students)

- **Yinan Xu** is applying this approach to fMRI data
- **Briana Smith** is using this approach to predict hippocampal volume in PTSD
  - See her upcoming ICCM talk and paper
- **Patrick Rice** is extending this approach to working memory
  - Mixture of **parameters** and **strategy**

Conclusion

Nomothetic Approach

🚫 Group averages

🚫 Low predictive power (“one size fits all”)

🚫 Stale

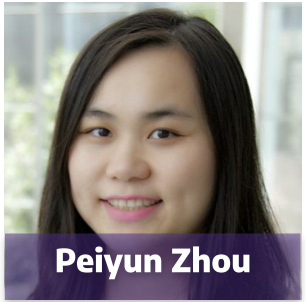
Idiographic Approach

✔️Individual differences

✔️High predictive power

✔️Possibilities are endless

Credits



university of groningen

Y of WASHINGTON