Decoding Idiographic ACT-R Parameters from Brain Data

Andrea Stocco¹, Peiyun Zhou¹, Florian Sense², & Hedderik van Rijn²
¹ University of Washington, Seattle
² University of Groningen, The Netherlands
Idiographic vs. Nomothetic

Diagram:
- Manual (M)
- Imaginal
- Procedural
- Retrieval (R)
- Visual (V)
- Goal (G)
Idiographic vs. Nomothetic

Goal

Retrieval

Manual

Imaginal

Procedural

Visual

Nomothetic Approach

Group average

W
Idiographic vs. Nomothetic

Idiographic Approach

Nomothetic Approach

Group average
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

$R = 0.86^{**}$
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

- Raw data
- Divide into 2-sec, overlapping epochs
- Remove epochs with artifacts

Decompose each epoch into frequencies:
- Gamma: 30 – 40 Hz
- Beta: 13 – 30 Hz
- Alpha: 8 – 13 Hz
- Theta: 4 – 8 Hz
- Delta: < 4 Hz

Mean power of each frequency:

Individual spectrogram
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:

\[ A = \sum_j t_j^{-d} \]

\[ d = ce^A + \alpha \]

- Consistent across very short and very long intervals
- Accounts for spacing effects
How is $\alpha$ Measured?
Predict when the chunk is forgotten
If the chunk is remembered, reduce $\alpha$
If the chunk was forgotten, increase $\alpha$
Hedderik and Florian have done massive work showing that $\alpha$ ...

- 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, $\alpha$ is psychological “trait”.
Can we decode $\alpha$ 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 $\alpha$ 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
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 \(\lambda\) accuracy and simplicity
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

- Used **cross-validation**
  - fit vs. generalizability
- Used Leave-One-Out (LOOV)
- Guarantees ideal lambda is actually best estimate (minimum absolute loss)

\[ \lambda \to 0 \quad \text{(simple regression)} \]

\[ \lambda \to \infty \quad \text{(only intercept)} \]
Predicted vs. Observed

Cross-Validation:
Predicted vs. Observed Rates of Forgetting

Condition
- Observed
- Predicted

$r = 0.9$
Predicted vs. Observed

Cross-Validation: Predicted vs. Observed Rates of Forgetting

Cross-Validation

Rate of Forgetting (Observed)

Rate of Forgetting (Predicted)

r = 0.9
What does it look like?

Delta

Theta

Alpha

Upper Beta

Gamma

Beta Regressor Value

Beta Values From Group Lasso

W
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

Peiyun Zhou
Florian Sense
Hedderik van Rijn