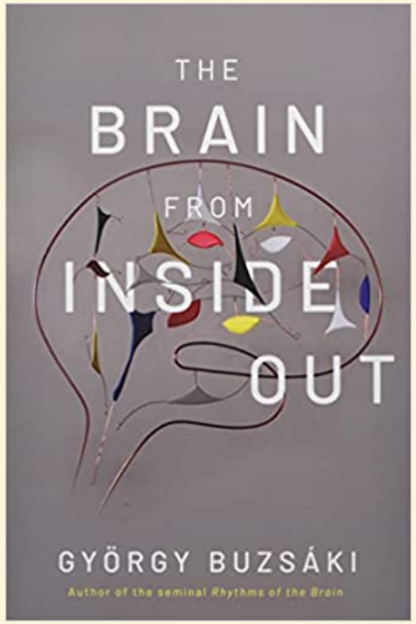
Clarifying Cognitive Constructs by Automated Text Mining of the Literature

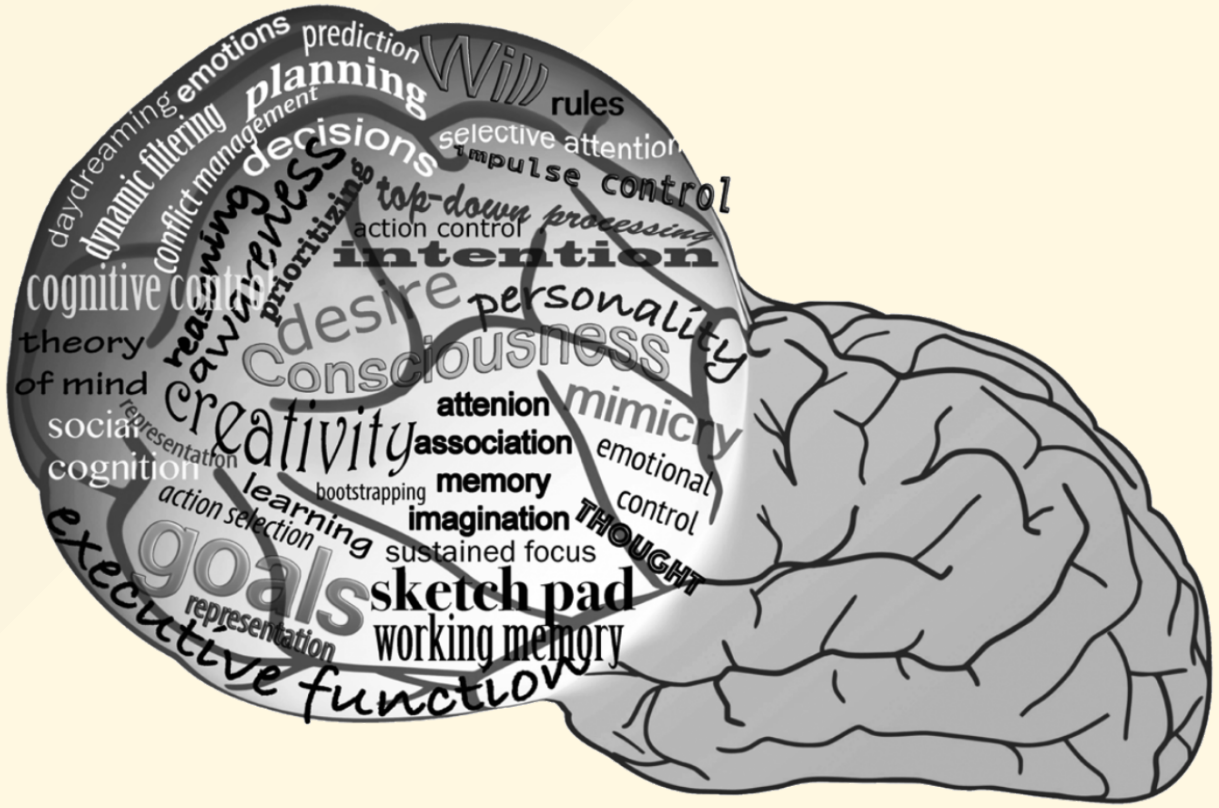
Morteza Ansarinia
Pedro Cardoso-Leite

COSA-LUCET Colloquium
July 14, 2021

This is an ongoing study. Results are still inconclusive.

Cognitive scientists have created numerous concepts, referring to them by arbitrary names.





Images from Buzsaki2019.

😔 There is a many-to-many relationship between tests and constructs; e.g., constructs are not fully independent and the same tests are being loaded into different constructs.

😔 Current state of affairs makes it hard to understand past results and limits scientific progress.

😔 Combining results from multiple disciplines is pretty difficult.

😔 It also makes it hard to develop effective interventions.

🎯 There is a great need for more conceptual clarity.

Example:

Executive Functions (EFs)

Executive functions (Diamond2013), executive attention (Engle2002), executive control (Posner1990), attention control, attentional control (Bavelier2019), cognitive control (Botvinick2016), self-regulation (Eisenberg2019), fluid intelligence (Diamond2013), fluid cognition, working memory (Baddeley1996), updating, shifting, and inhibition (Miyake2000).

To gain clarity, we can...

🪶 **manually** read, synthesize, and criticize the literature to write reviews describing our understanding of the field.

- 👎 It's biased and reflects author's view.
- 📈 Sheer volume of papers published every year (200 EFs papers per month on PubMed).

OR

🤖 **automatically** analyze the scientific texts.

Executive Functions (EFs)

The ability to coordinate complex behaviors in pursuit of goals.

Is a focus of research in many disciplines, including psychology, neuroscience, and artificial intelligence.

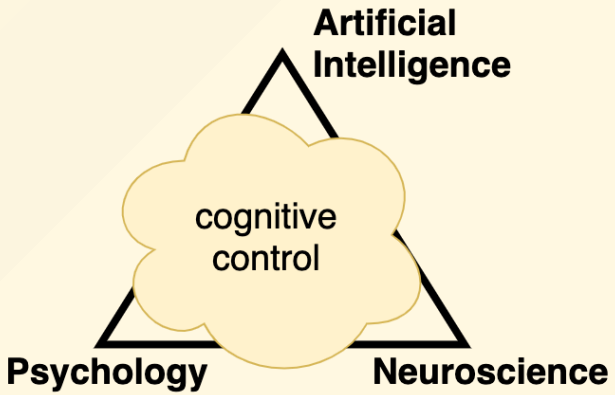
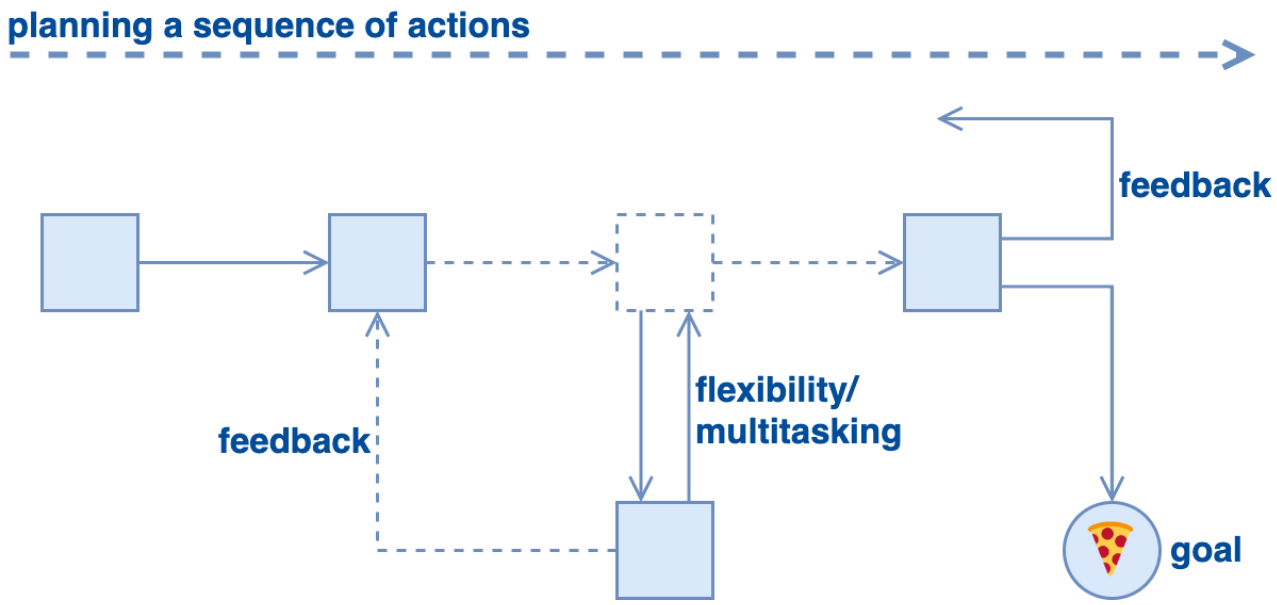


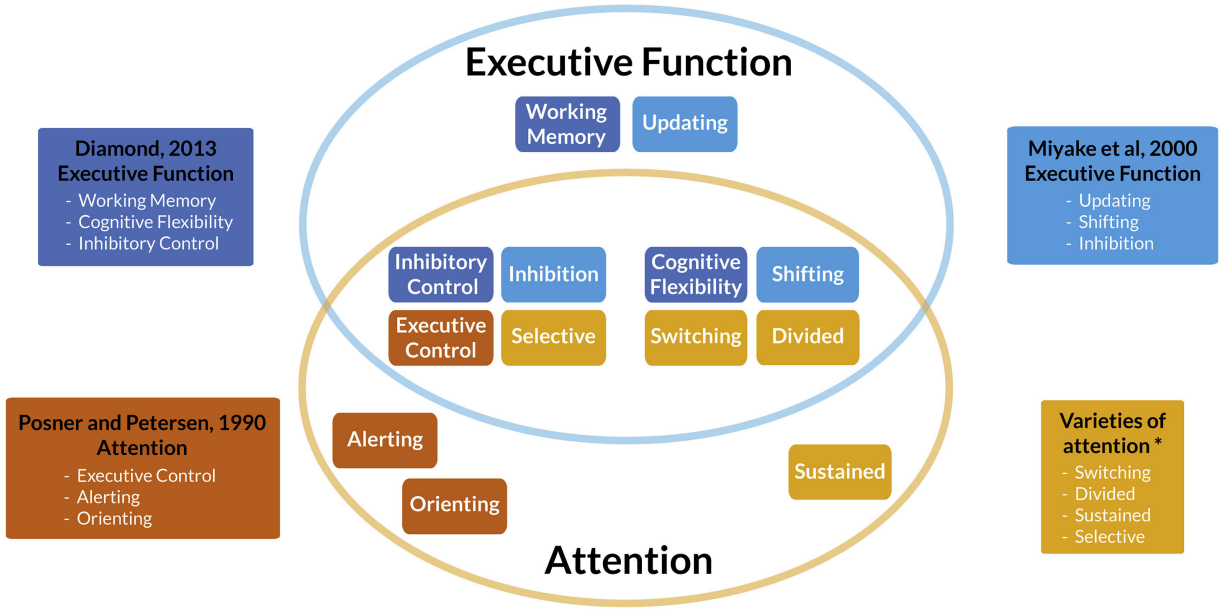
Image reproduced from Botvinick at Triangulating Intelligence (2020).

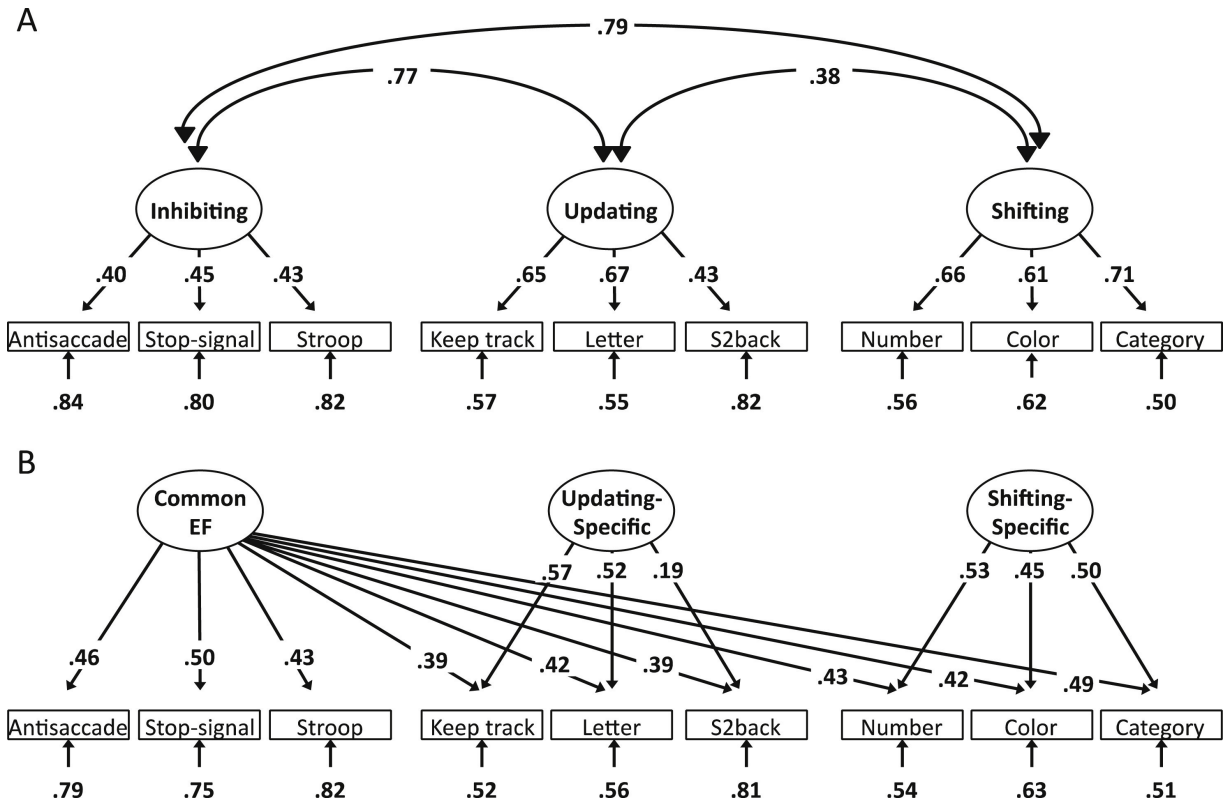
Executive Functions (EFs)

🍕 One example of such complex behaviors would be, for instance, cooking a pizza.

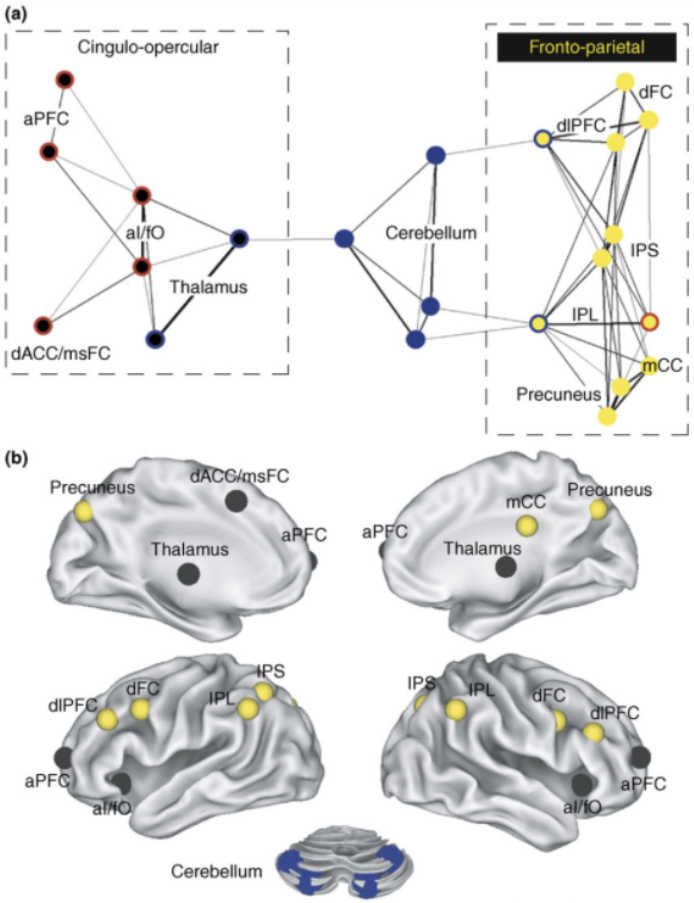


Models of EFs



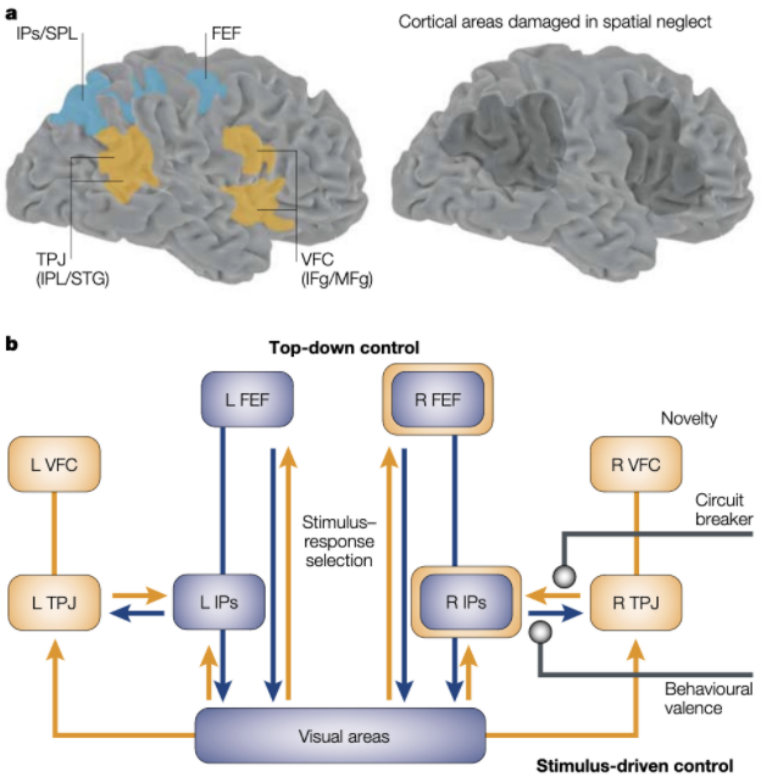


Images from Bavelier2019 and Miyake2017.



Distinct brain networks for adaptive and stable task control in humans

Nico U. F. Dosenbach*†, Damien A. Fair‡, Francis M. Miezin*‡, Alexander L. Cohen*, Kristin K. Wenger‡, Ronny A. T. Dosenbach*‡, Michael D. Fox*, Abraham Z. Snyder*‡, Justin L. Vincent*, Marcus E. Raichle*†‡§, Bradley L. Schlaggar*‡§¶, and Steven E. Petersen*†‡§‖



CONTROL OF GOAL-DIRECTED AND STIMULUS-DRIVEN ATTENTION IN THE BRAIN

Maurizio Corbetta and Gordon L. Shulman

Images from Dosenbach2007 and Corbetta2008.

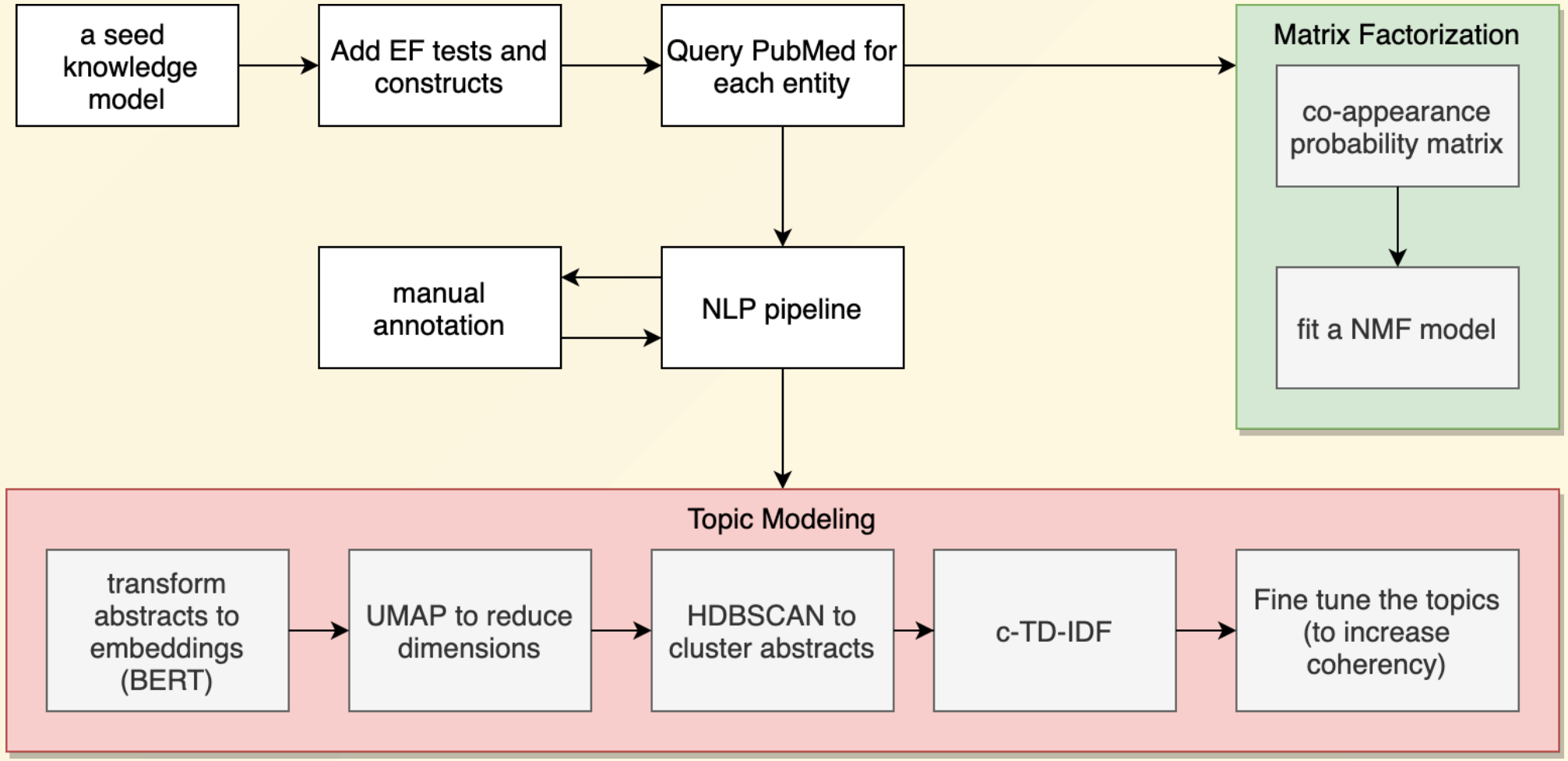
Current project

A text-based method to gain clarity on the meaning of EFs, including:

- how they should be **operationalized** (tests, batteries, instructions)
- how they should be **conceptualized** (constructs)
- how they should be **observed** (methods and brain mechanisms)

Ultimately, develop a shared latent space that connects many tests, constructs, and brain mechanisms.

Method

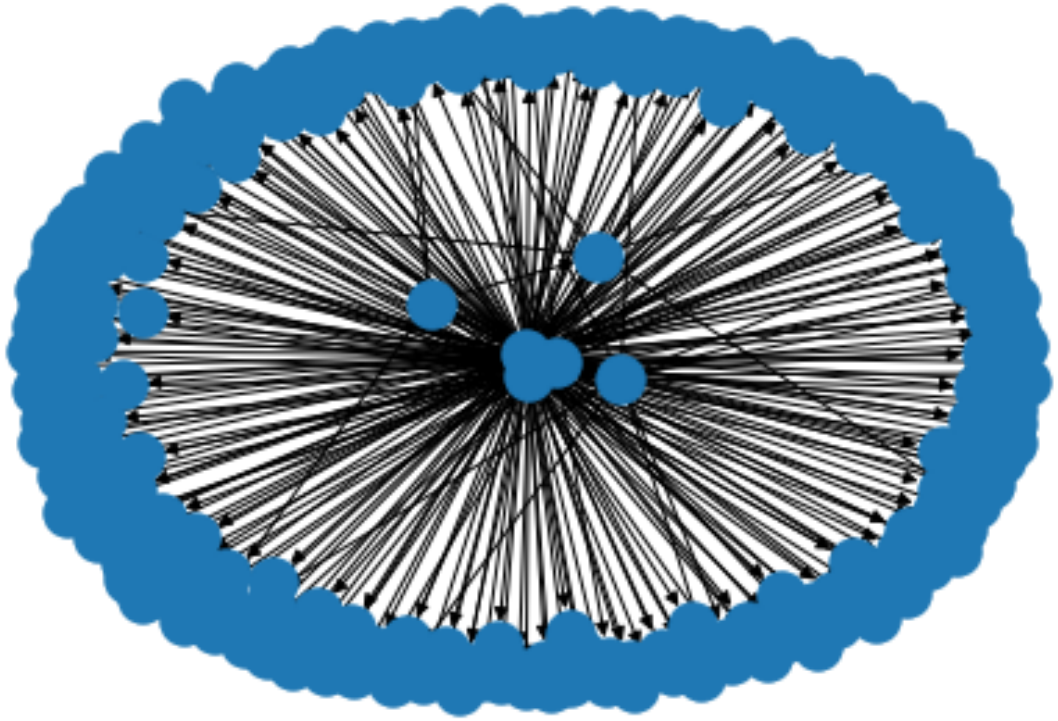

a seed knowledge model
Add EF tests and constructs
Query PubMed for each entity
Matrix Factorization
co-appearance probability matrix
fit a NMF model
manual annotation
NLP pipeline
Topic Modeling
transform abstracts to embeddings (BERT)
UMAP to reduce dimensions
HDBSCAN to cluster abstracts
c-TD-IDF
Fine tune the topics (to increase coherency)

Method

- Develop a knowledge model of what we know about EF-related constructs, tests, and brain mechanisms.
- Collect publications related to the constructs and tests.
- Parse the texts.
- Use computational methods to develop:
  1. descriptive statistics
  2. latent model of tests/constructs co-appearance
  3. latent topic model

Executive Functions Ontology (EFO)

A machine-readable graph-based knowledge model of what we know about *Executive Functions*.

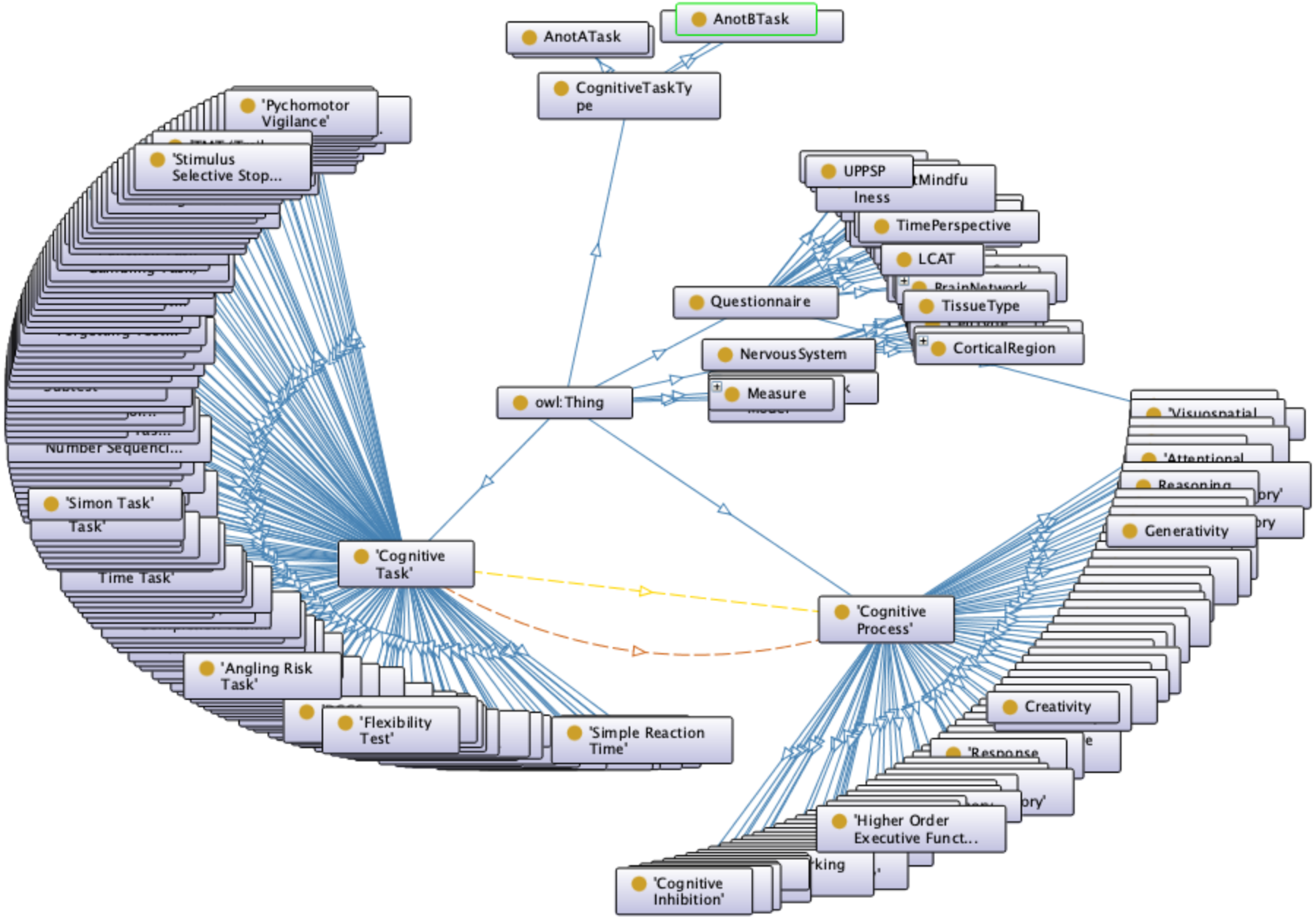


Executive Functions Ontology (EFO)

- We created a seed ontology and then improved it by manually adding tests, constructs, brain mechanism, models, and questionnaires from highly cited review papers[1].
  - 102 EF tests.
  - 67 EF constructs.
  - *brain mechanisms, EF models, questionnaires, etc.*

[1] Diamond2013, Miyake2000, Baggetta2016, Enkavi2019, Eisenberg2019, CogPo, and CogAt.

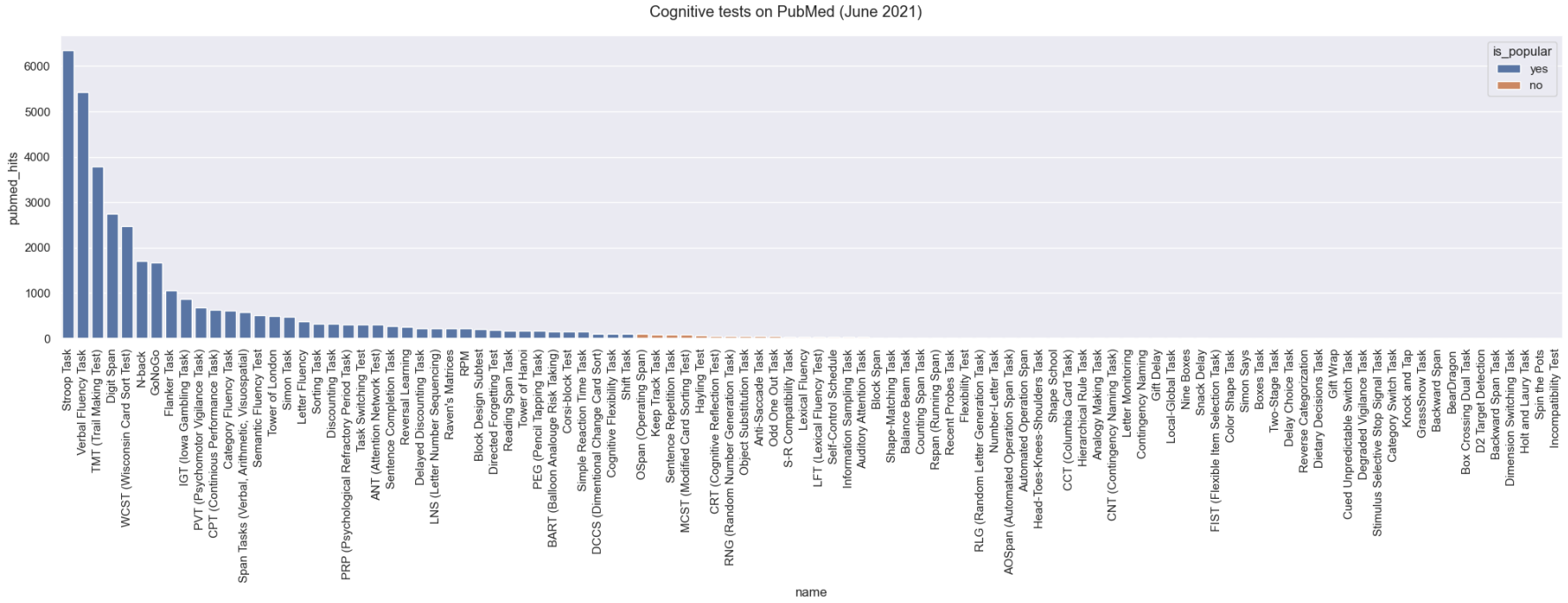
EFO

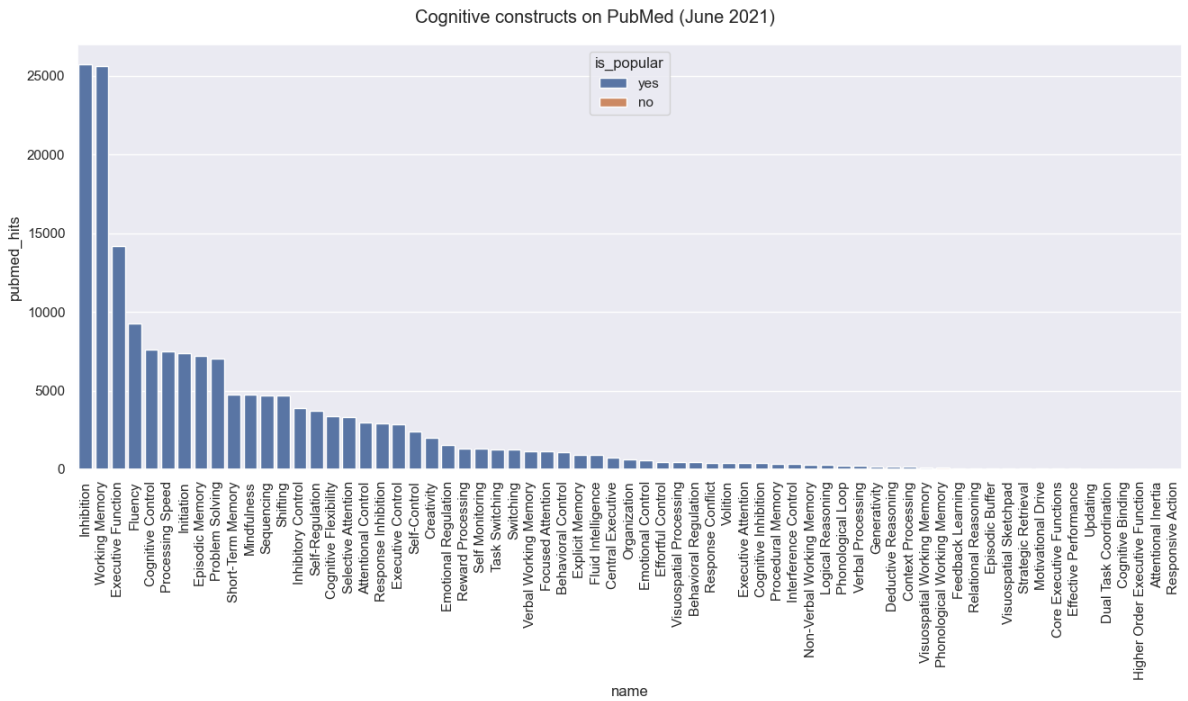

AnotBTask
AnotATask
CognitiveTaskTy
pe
'Pychomotor
Vigilance'
'Stimulus
Selective Stop...
UPPSP
Iness
TimePerspective
LCAT
Questionnaire
TissueType
CorticalRegion
NervousSystem
Measure
owl:Thing
Number Sequenci...
'Simon Task'
Task'
Time Task'
'Cognitive
Task'
'Cognitive
Process'
Generativity
'Angling Risk
Task'
'Flexibility
Test'
'Simple Reaction
Time'
Creativity
'Higher Order
Executive Funct...
'Cognitive
Inhibition'

Preliminary Results

Frequency of tasks and constructs

- Many tests and constructs, but few are used (power law).





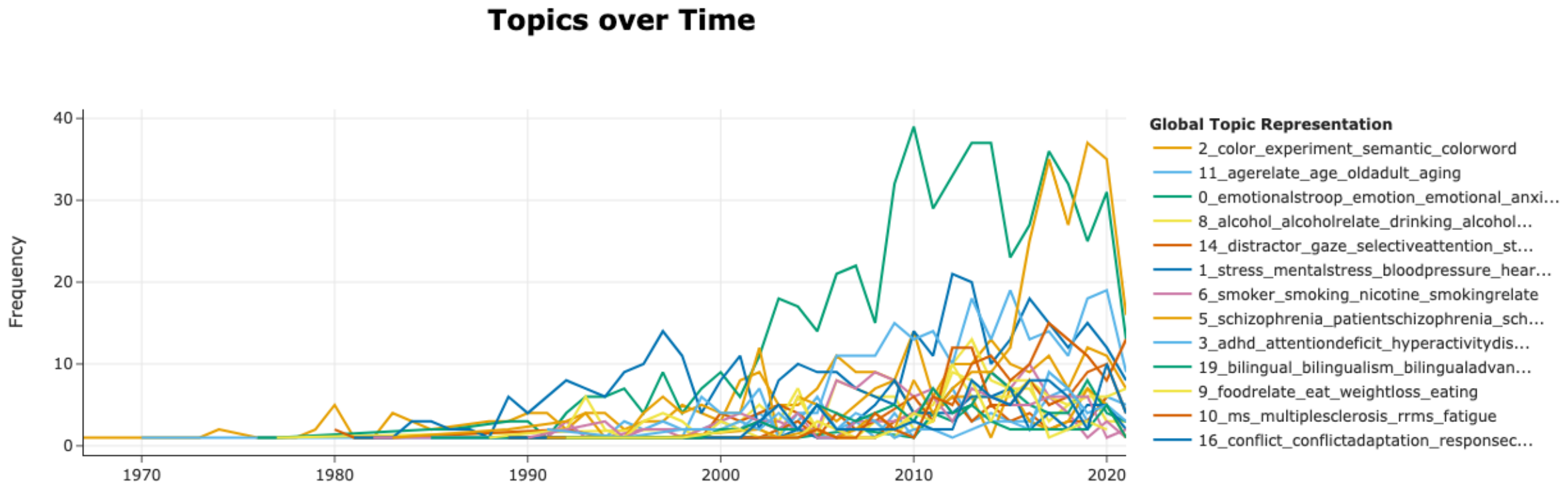
PubMed-EF corpus: word cloud

SKIP

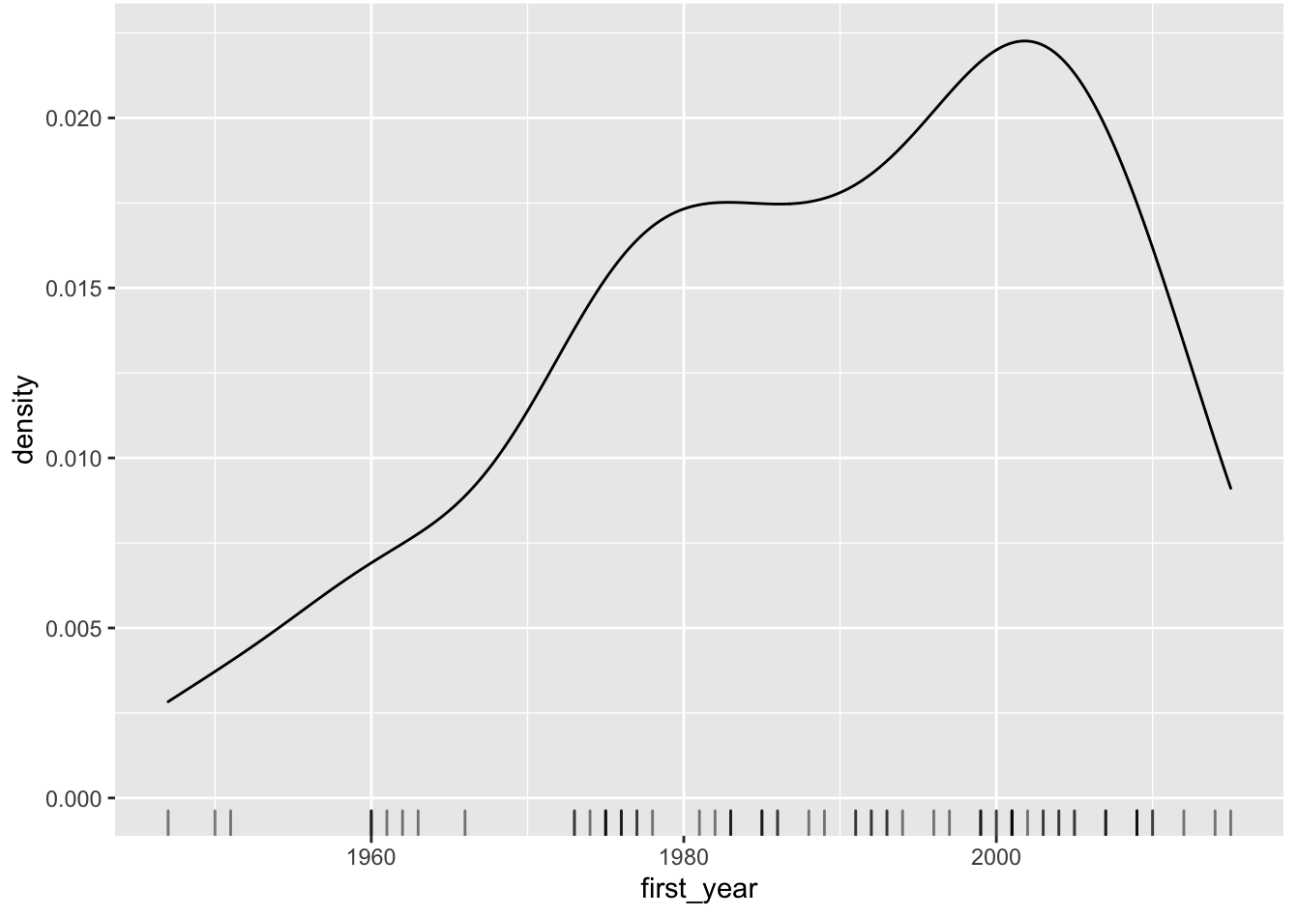
- number of tasks per paper (x), percentage of papers (y)
- how many papers used more than one task
- co-occurrence of tasks

Example Corpus: Stroop Task

Example Corpus: Stroop Task



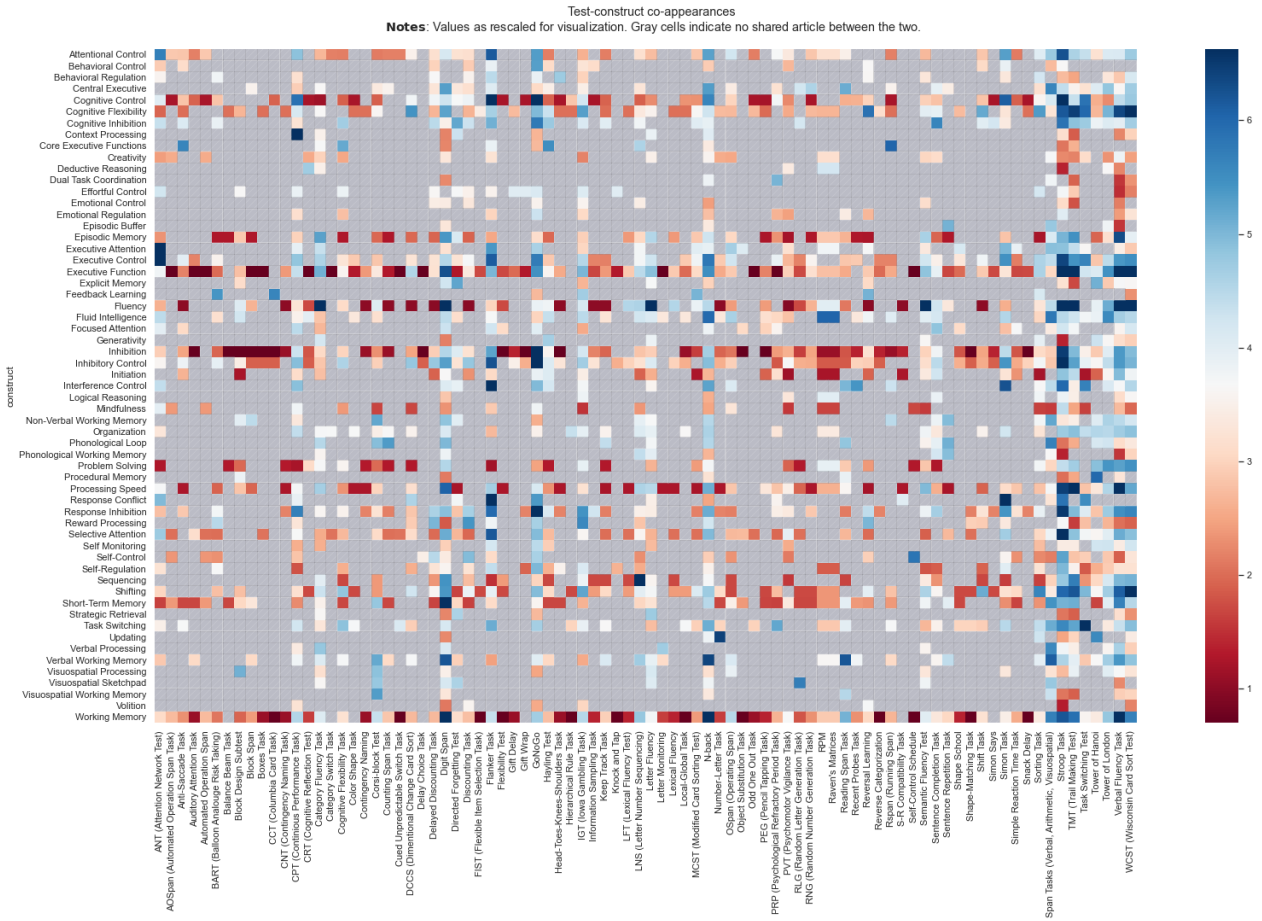
New standard tests



Co-appearance of tests and constructs

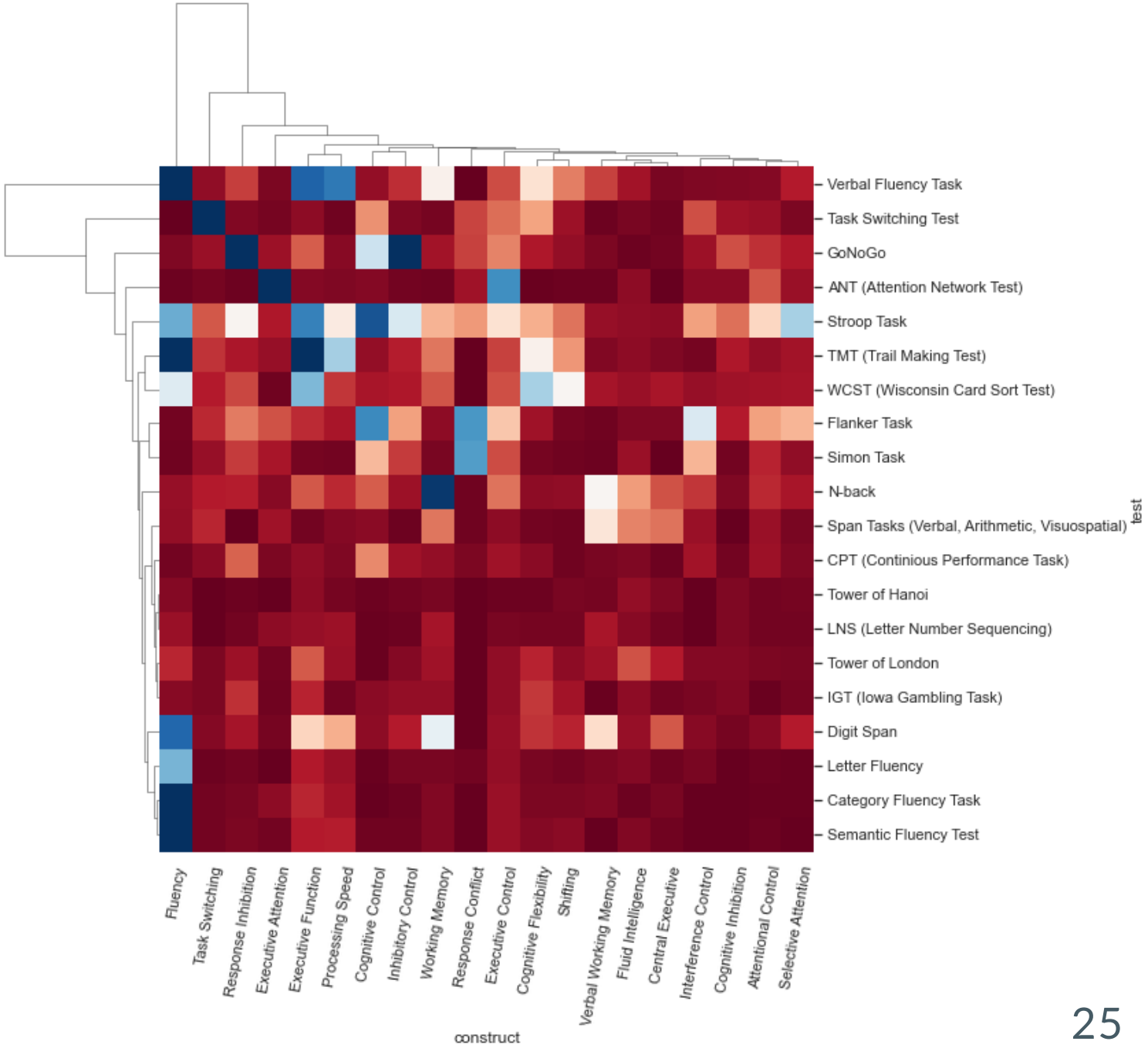
- specificity of tests

- some are specific to a single construct
- some are generic as for many constructs



Co-appearance of tests and constructs

- Popular tests and constructs[1]



This is a subset of previous heat map.

A latent space of tasks and constructs

💭 We are interested in a shared space of tests and constructs that encapsulates the similarity between one type given the other.

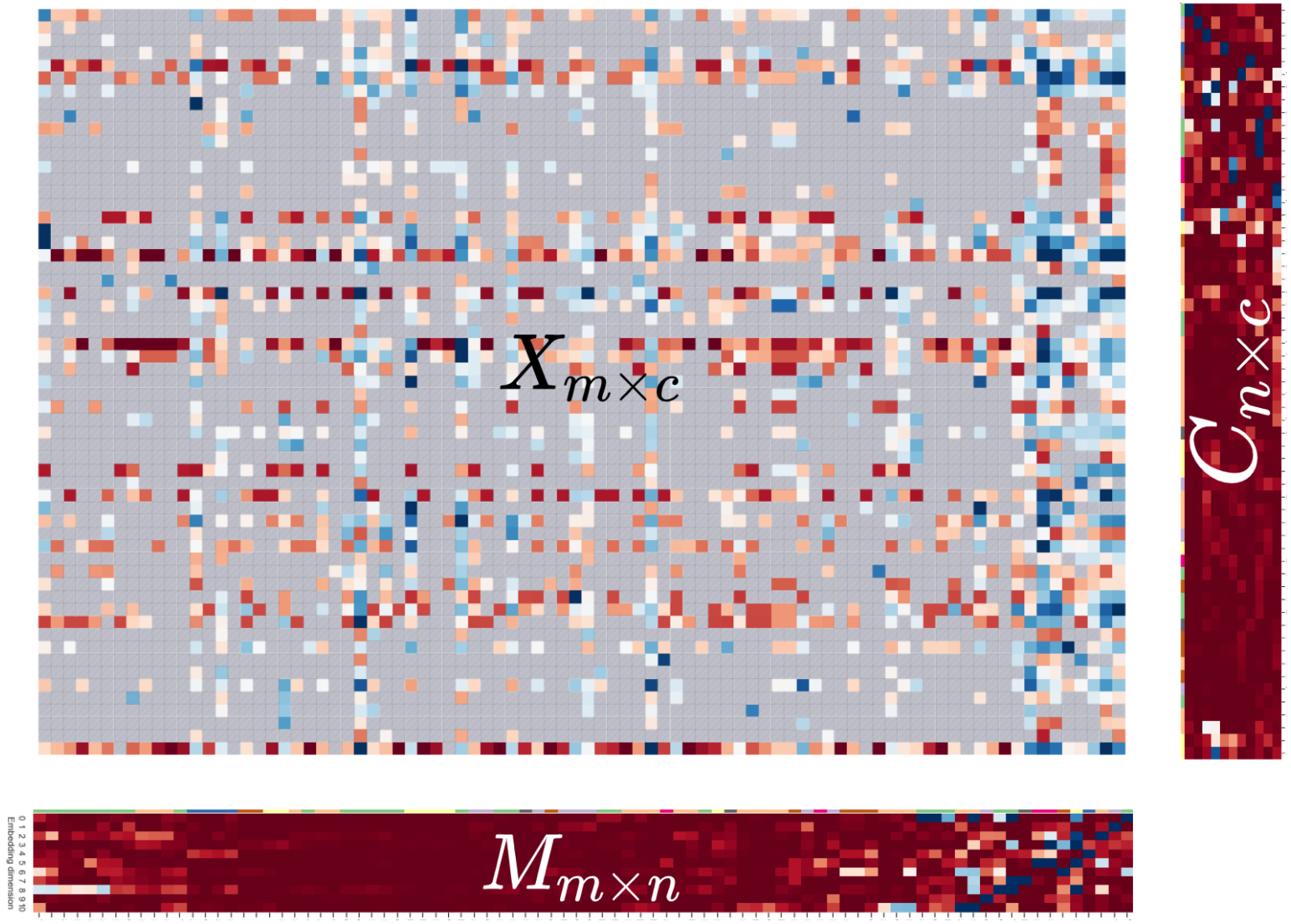
Similar to Peterson2019 and Bavelier2019, but driven by data.

- Method 1: factorize the co-appearance matrix

$$X_{m \times c} \approx M_{m \times n} \times C^T_{n \times c}$$

- SKIP Method 2: topic modeling

X is the co-appearances, n is embedding dimension, M is test, and C is construct.


$X_{m \times c}$
$C_{n \times c}$
$M_{m \times n}$
Embedding dimension
0 1 2 3 4 5 6 7 8 9 10

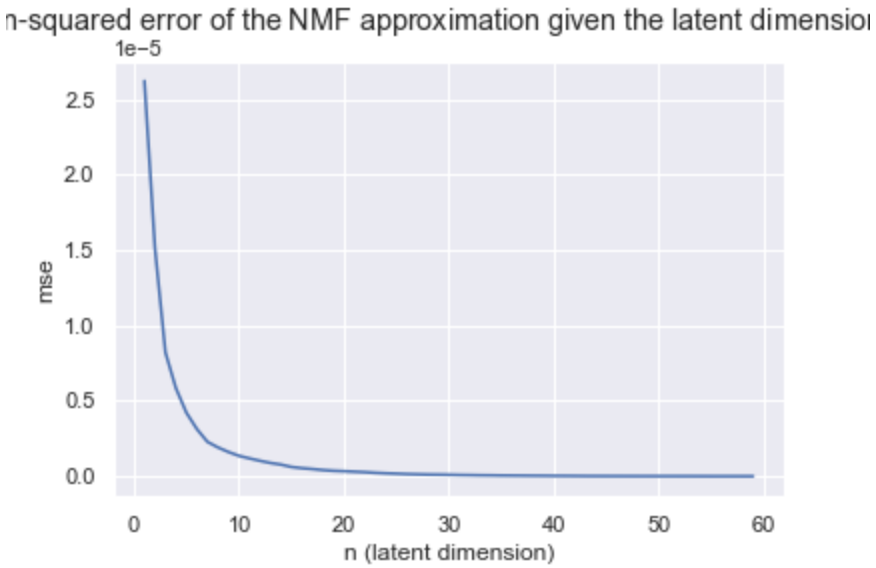
Model fitting

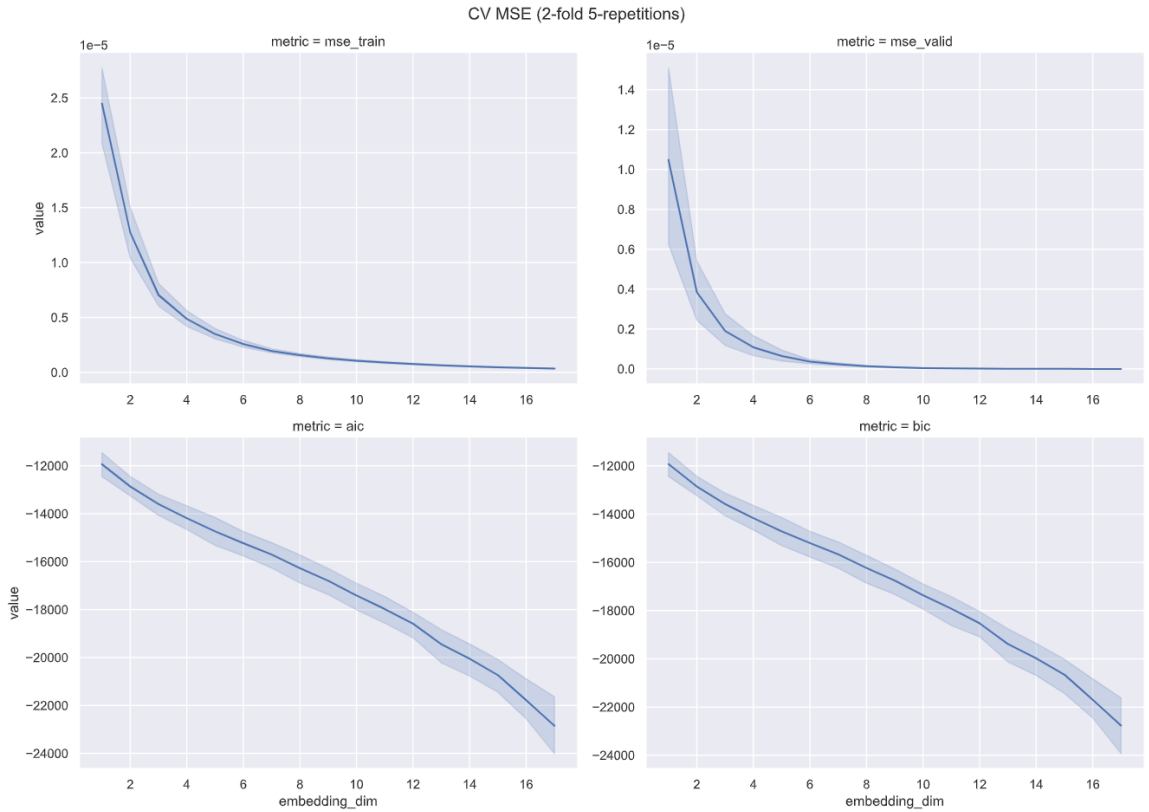
- Non-negative factorization was used to infer latent matrices from a log-transformed probability matrix.

$$\mathrm{f}(n) = \alpha \times \min(\mathbf{MSE}_{1..\min(m,c)}) - \mathbf{MSE}_n(X - M \times C^T)$$

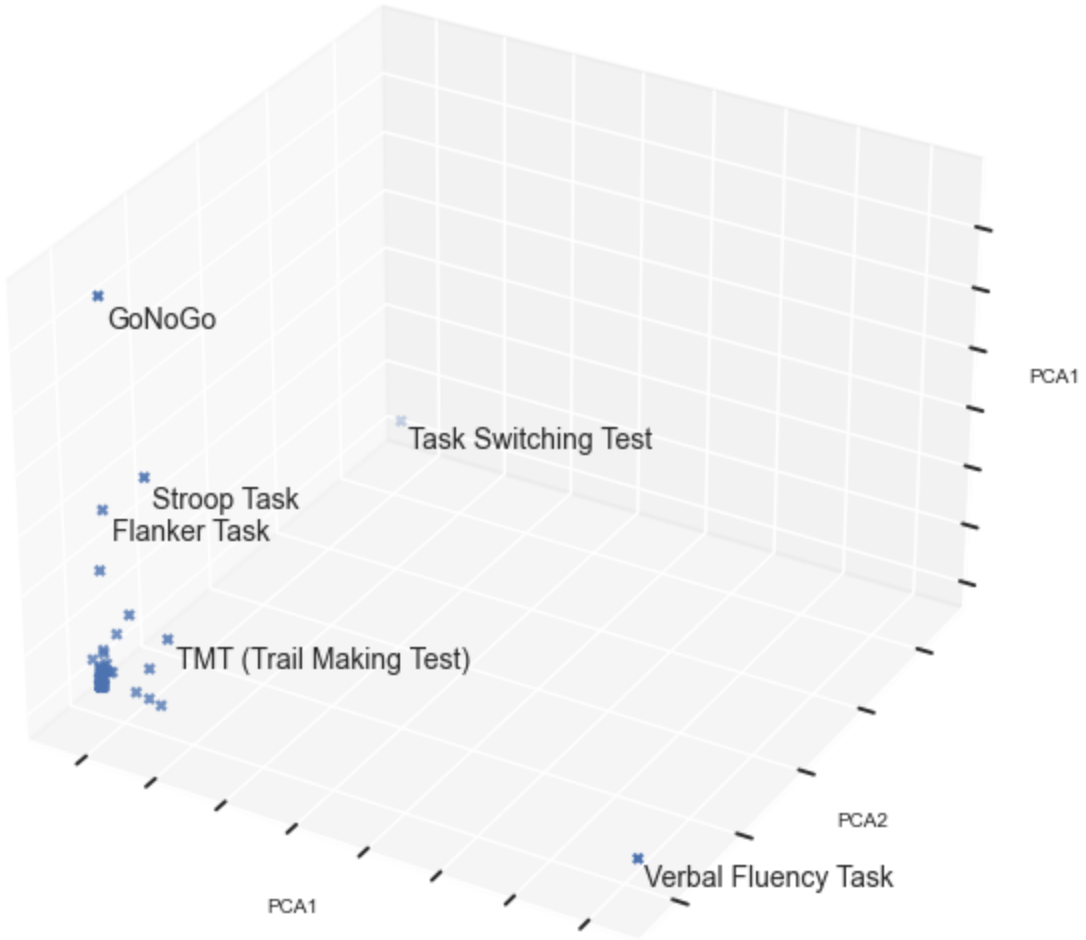
$$\underset{n}{\operatorname{argmin}}(f(n)), f > 0$$

Model fitting

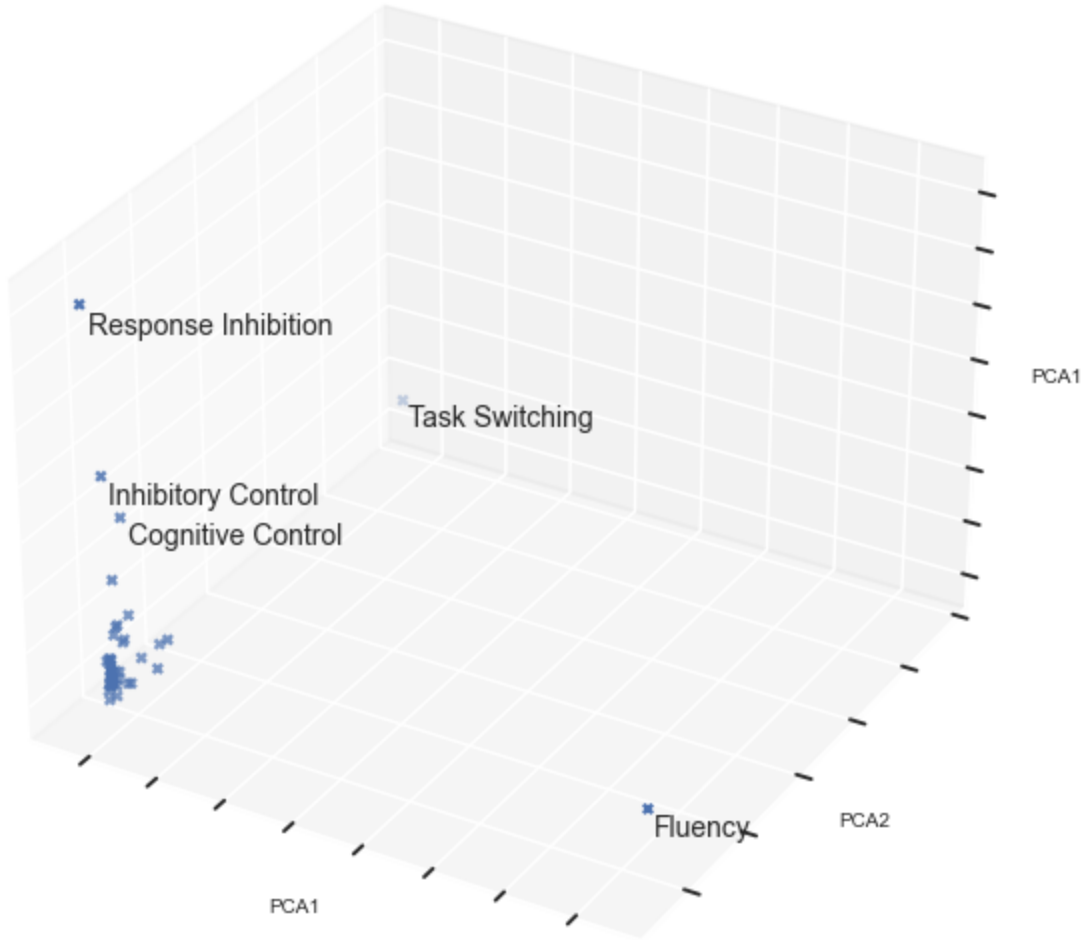




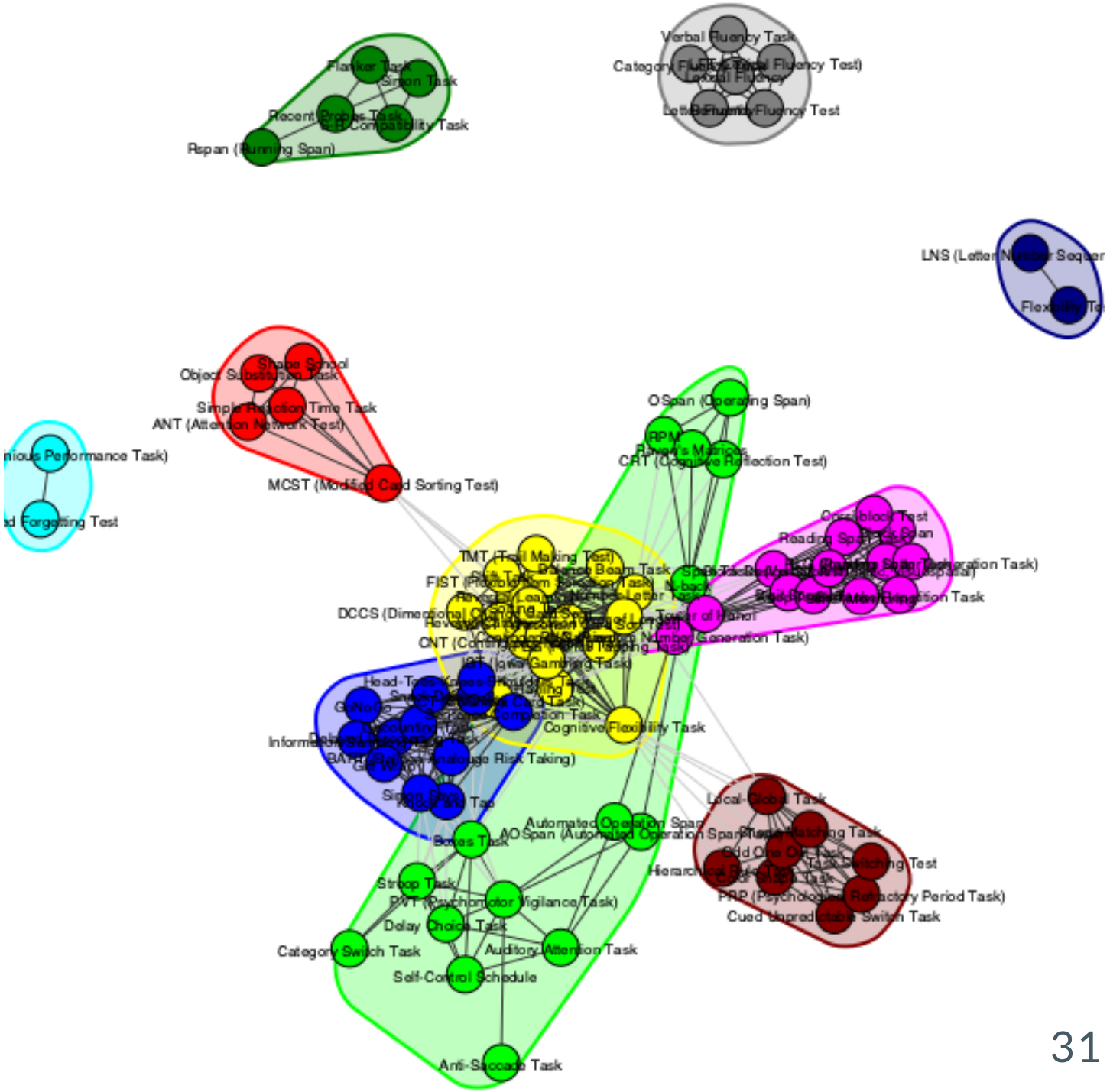
PCA projection of cognitive test embeddings



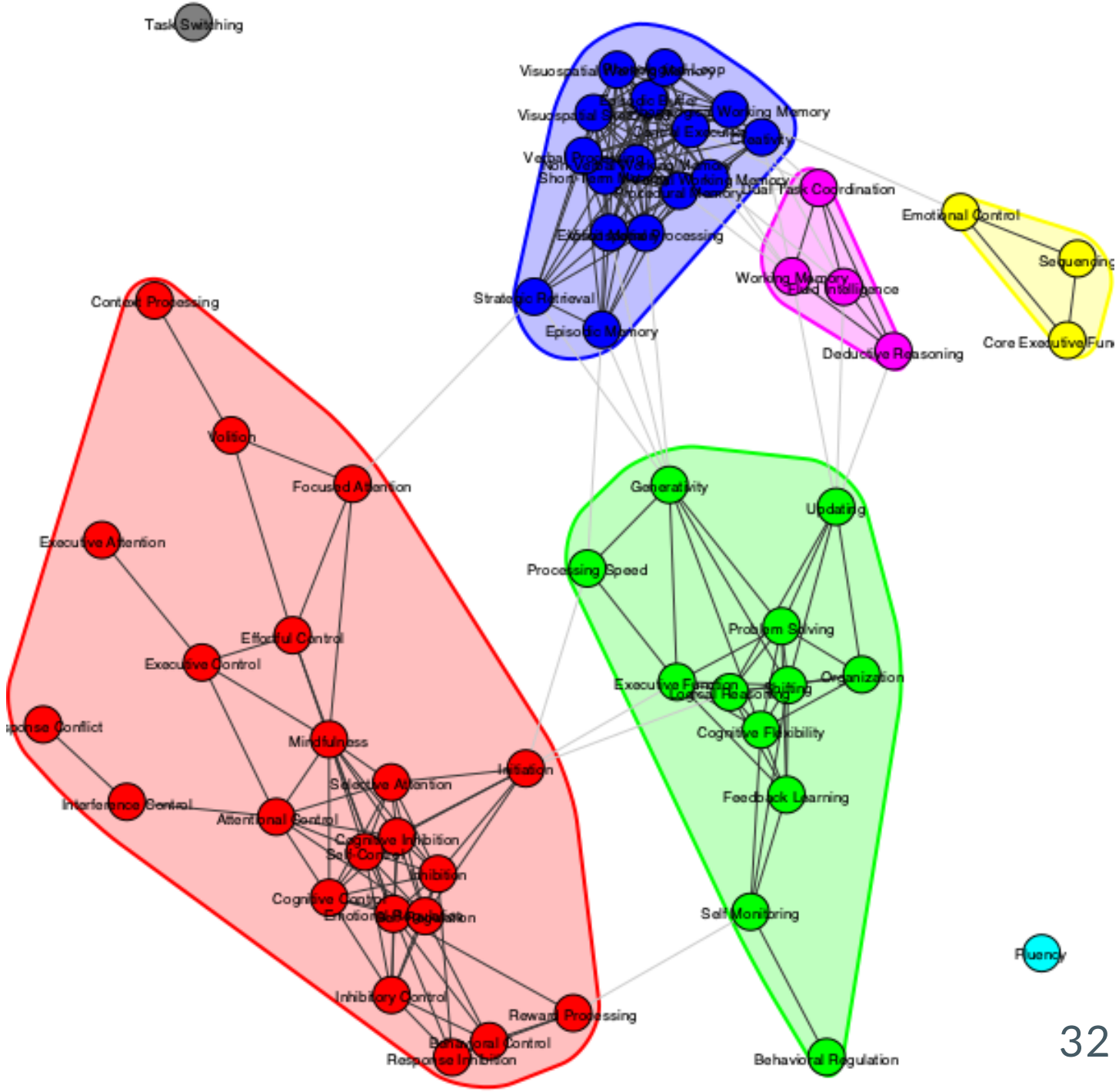
PCA projection of cognitive construct embeddings



Cognitive tests similarity map



Cognitive constructs similarity map



Recap

- We created an ontology of EF-related cognitive tests and constructs,
- collected ~7GB of PubMed abstracts,
- developed tools to automatically analyze the texts,
- extracted a latent space that was shared between tests and constructs, and
- all to reduce the confusion.

Conclusion

- EFs are redundant and unclear, yet analysis of the test/construct latent space reduces the confusion.
- Need for more rigorous methods to avoid confusion (e.g., ontologies, constrained definition of constructs, data-driven knowledge models, and a theory about cognitive tests).
- Instead of focusing on confirmatory analysis, we can focus on cognitive tests (.e.g, use data to develop a battery that covers many cognitive constructs).

Future works

- Improved topic modeling using transformers.
- Questionnaires in the ontology.
- Human-in-the-loop (manually annotate ambiguous parts of the corpus).
- A website for the interactive visualizations.
- Named Entity Tagging (NER) to identify new tests and constructs.

Reproducibility and open science

Outcomes of this study include:

- EF ontology,
- collected PubMed corpus,
- trained NLP models and derived data,
- codes.

Currently are all the materials are available on **Uni HPC GitLab**.
In the future it will be openly available on GitHub.