BRIDGING COGNITIVE AND PSYCHOMETRIC MODELS OF COGNITIVE ABILITIES

An Investigation Based on Process Overlap Theory

Han Hao
Postdoctoral Researcher
Caliber Lab
New Mexico State University

05/08/2024 @ Tarleton State University

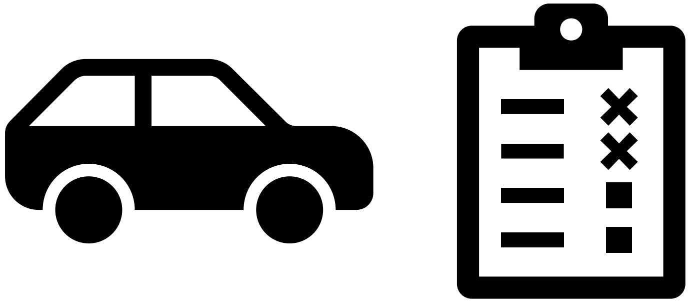
ABILITY

How do we understand an ability?

“Able”, “Can”, “Potential”

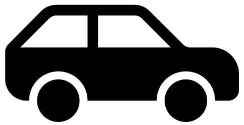
“I have the ability to drive.”

Drive test!

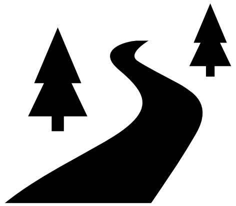


ABILITY

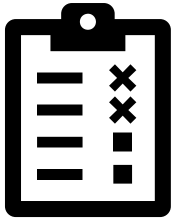
Ability



Test/Measure



Criteria/Differences

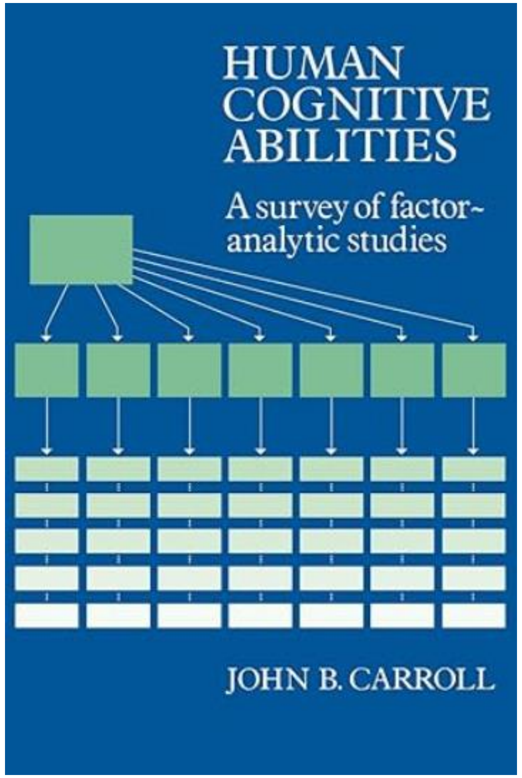


COGNITIVE ABILITIES

Cognitive ability

Cognitive task

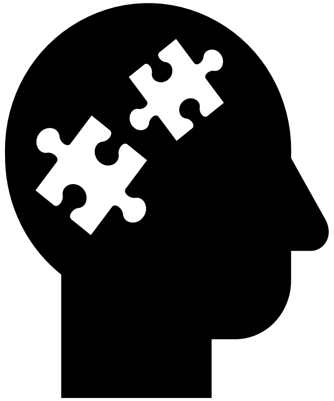
Cognitive process



THEORIES OF COG ABILITIES

- Various definitions and theories
- Robust phenomena and findings
- Positive manifold
- A common underlying cause of positive manifold?

THEORIES OF COG ABILITIES



Psychometric theories

Cognitive theories

PSYCHOMETRIC THEORIES

Investigate correlational relationships and individual differences in performance of (cognitive) tests to understand the "**map of mind**" (Sternberg, 2012, p.19)

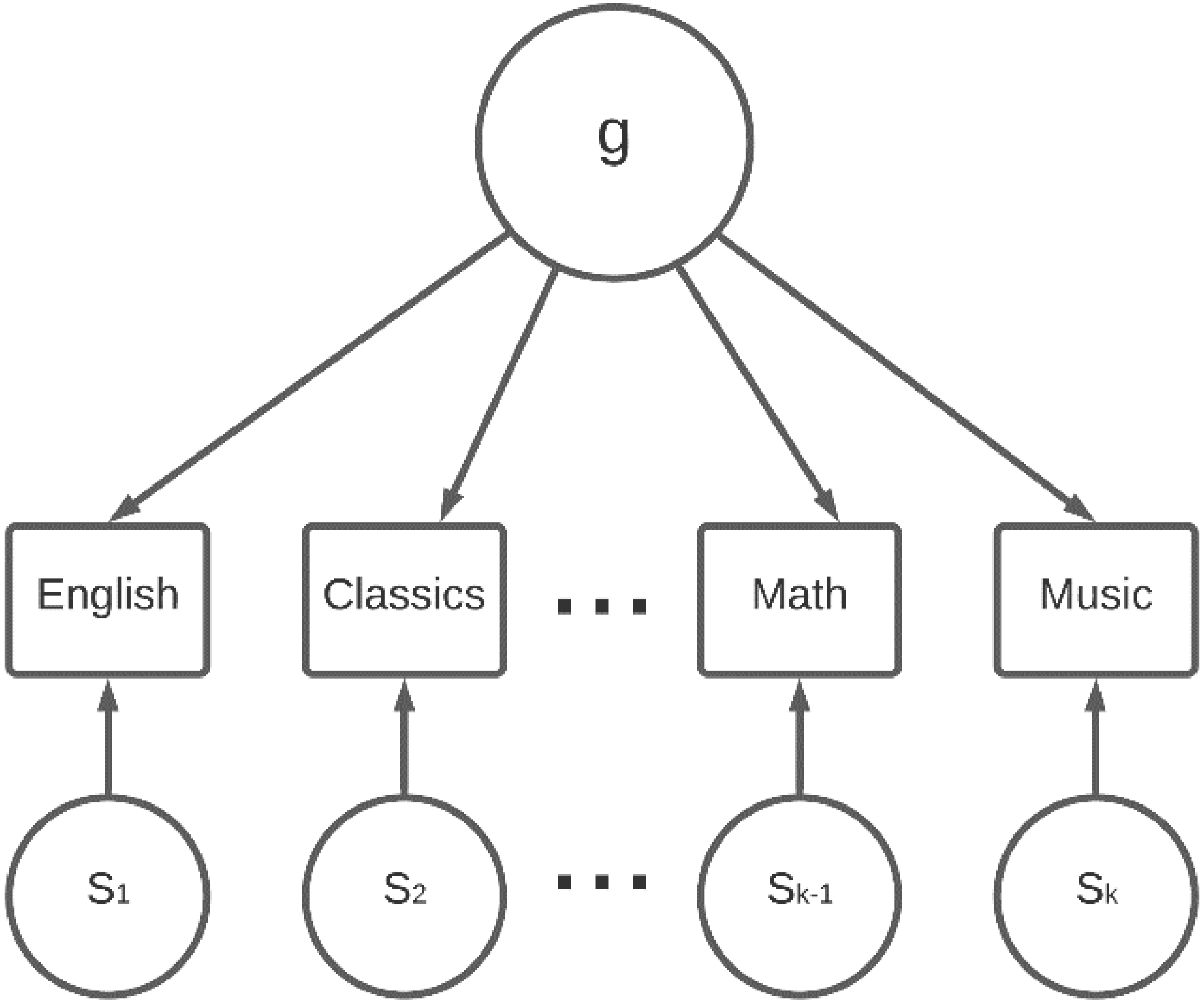
Correlational data, latent variable analyses

Stem from the investigation on (general) intelligence

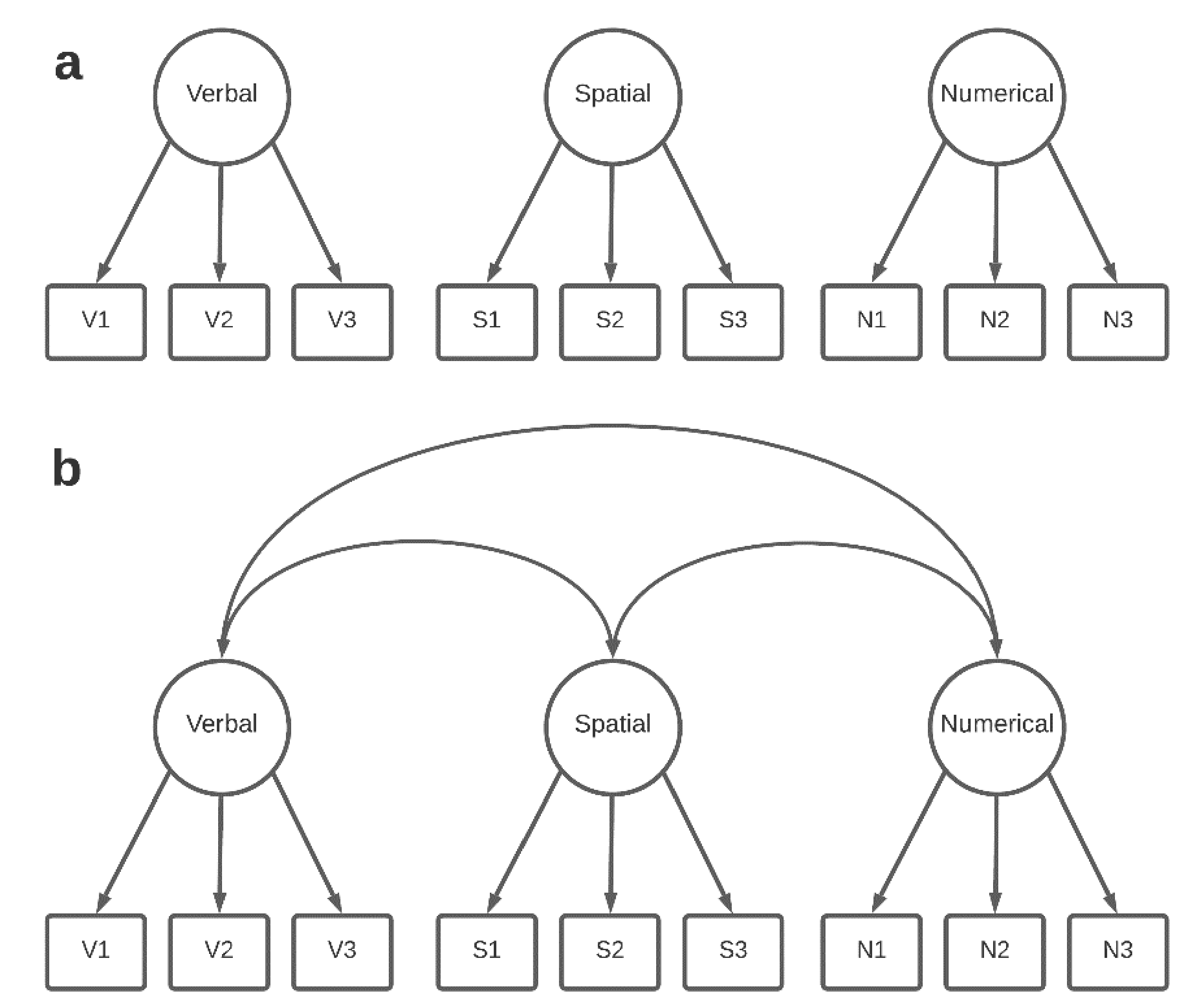
INTELLIGENCE RESEARCH

Spearman's Theory of General Intelligence

- **One-factor model of intelligence**



INTELLIGENCE RESEARCH

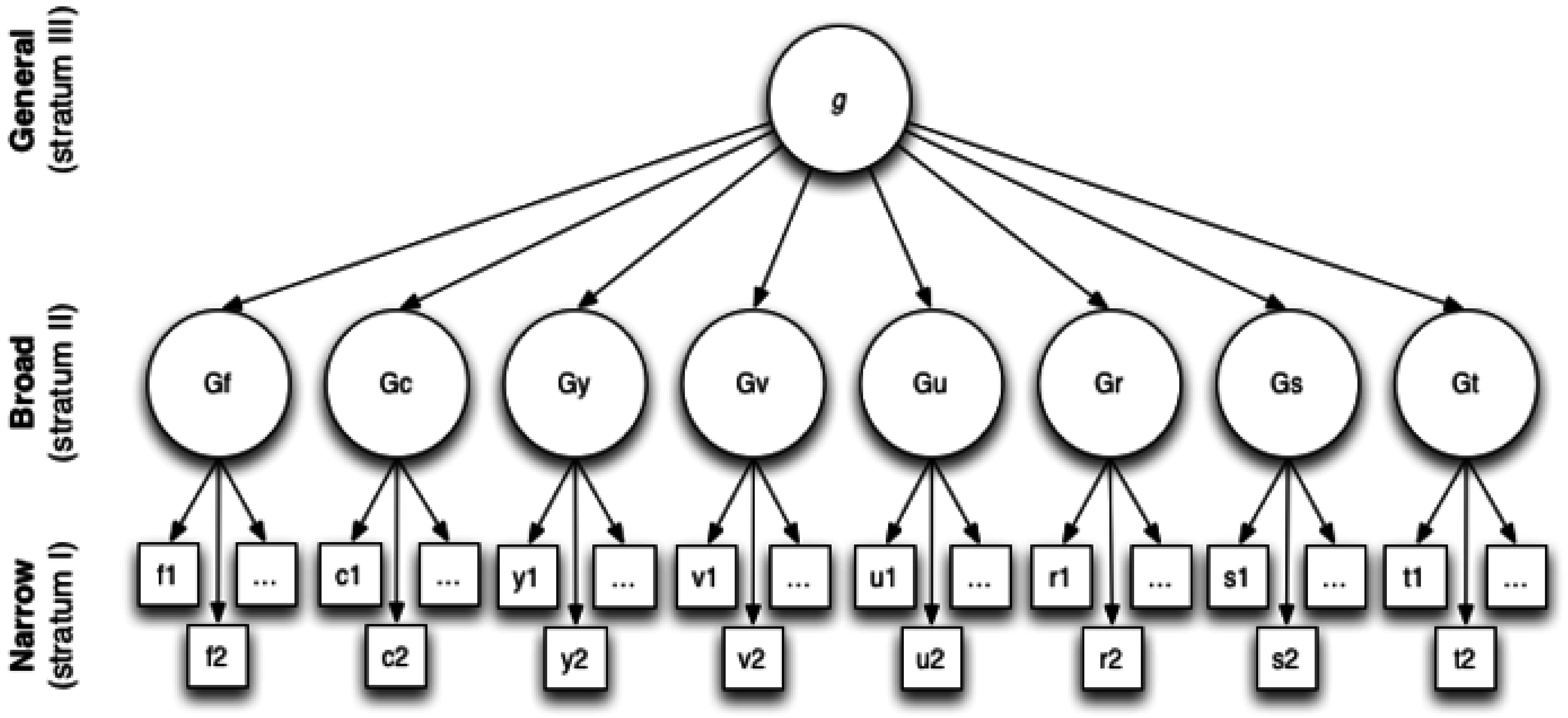


Spearman's Theory of General Intelligence

- One-factor model of intelligence

Thurstone's Primary Mental Abilities

- Multi-Factor Models of Intelligence


General
(stratum III)
Broad
(stratum II)
Narrow
(stratum I)
g
Gf
Gc
Gy
Gv
Gu
Gr
Gs
Gt
f1
...
f2
c1
...
c2
y1
...
y2
v1
...
v2
u1
...
u2
r1
...
r2
s1
...
s2
t1
...
t2

COGNITIVE THEORIES

Investigate the specific roles of important cognitive processes in cognitive activities as basic components in **information processing**

COGNITIVE THEORIES

Investigate the specific roles of important cognitive processes in cognitive activities as basic components in **information processing**

Experimental & correlational approaches

Important cognitive processes

- E.g., Working Memory (WM)

WORKING MEMORY

A system that maintain temporary availability to a limited amount of information for ongoing information processing (Cowan, 2017)

An attentional bottleneck to higher-order cognitive abilities

- **r = .70 to .90 for WM & gF (Kane, Hambrick, & Conway, 2005)**

Associated with real-world cognitive behaviors

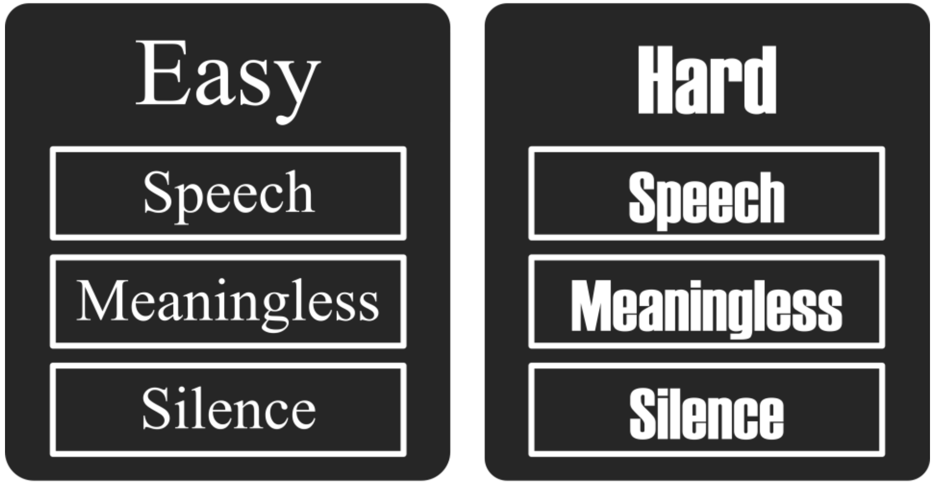
- **Problem solving, Planning, Learning, Metacognition, etc.**

WORKING MEMORY

Hao & Conway (2022)
The Impact of Auditory Distraction on Reading Comprehension: An Individual Differences Investigation

N = 126, 2 X 3 Mixed factorial

- **Perceptual disfluency (Between: Easy vs. Hard)**
- **Background noise (Within: Silence, Meaningless, Speech)**
- **Working memory capacity (WM span tasks)**



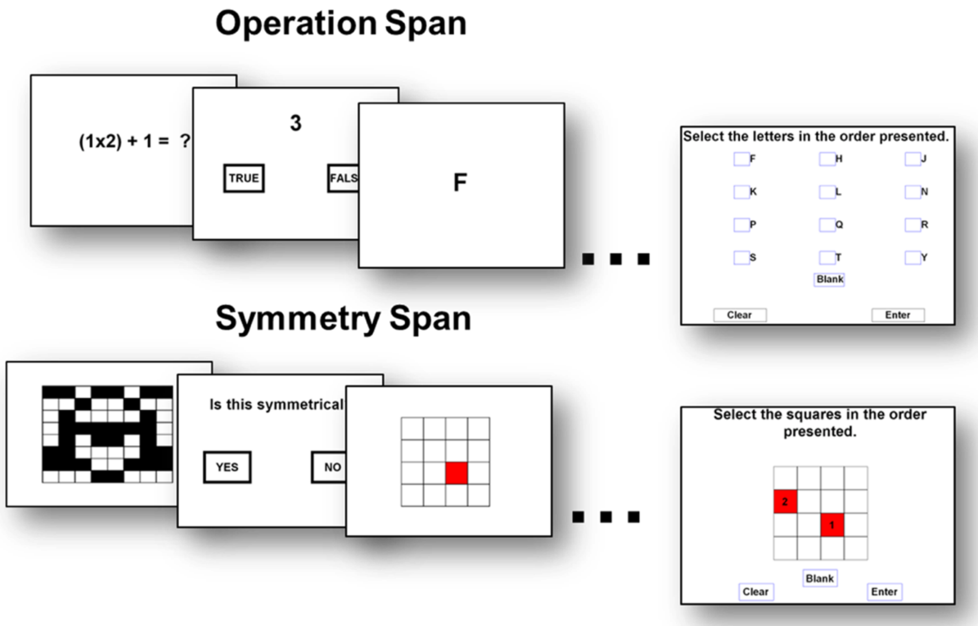
WORKING MEMORY

Hao & Conway (2022)

The Impact of Auditory Distraction on Reading Comprehension: An Individual Differences Investigation

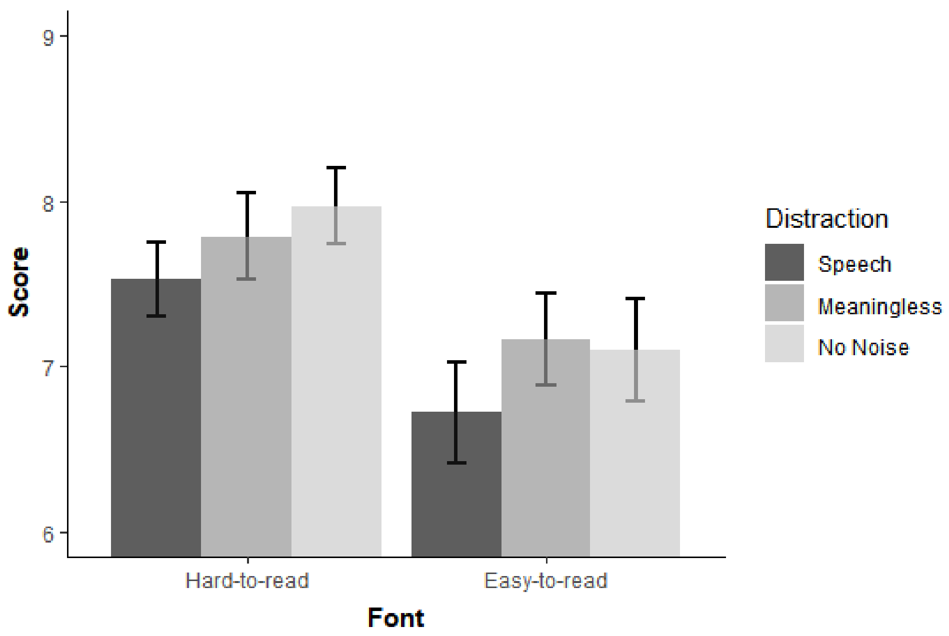
N = 126, 2 X 3 Mixed factorial

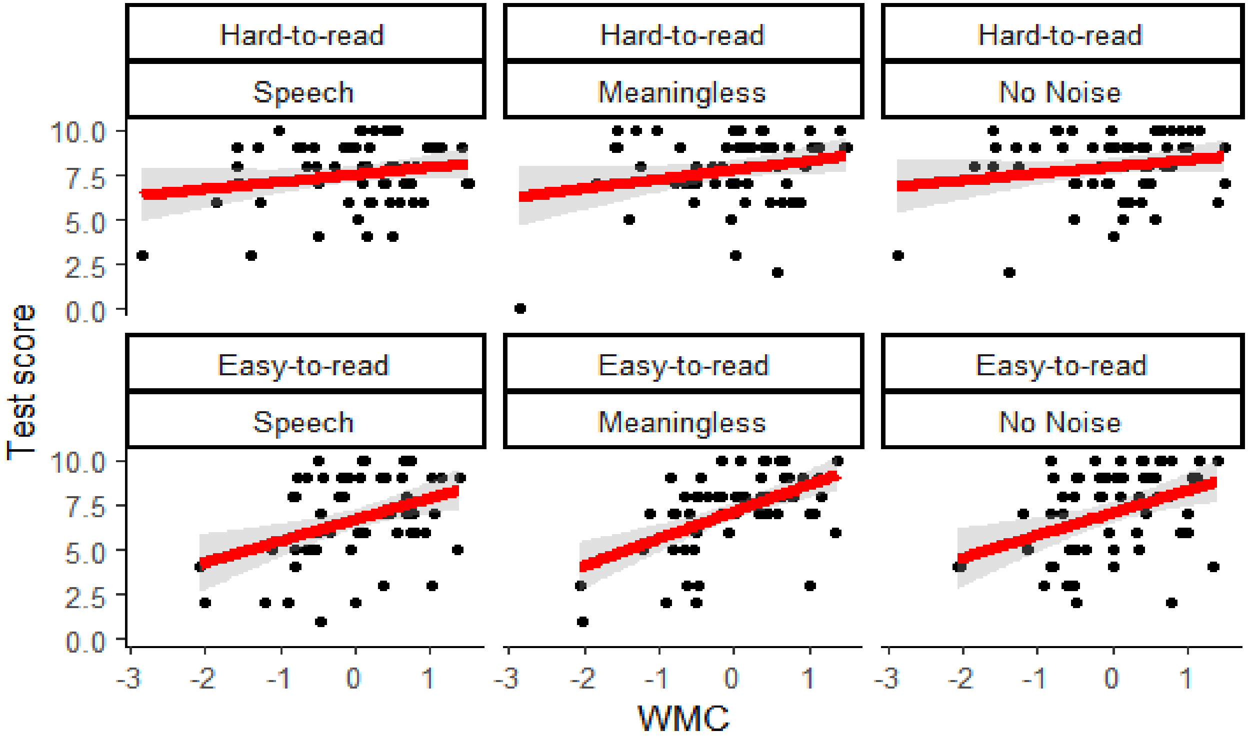
- **Perceptual disfluency (Between: Easy vs. Hard)**
- **Background noise (Within: Silence, Meaningless, Speech)**
- **Working memory capacity (WM span tasks)**



WORKING MEMORY

Hao & Conway (2022)





PROBLEMS OF THE CONVENTIONAL THEORIES

Psychometric & Cognitive Theories

PROBLEMS

Psychometric theories: “The common-cause premise problem”

Latent factor ≠ a common cognitive process/mechanism

Cognitive theories: “The measurement problem”

No task is process-pure

PROBLEMS

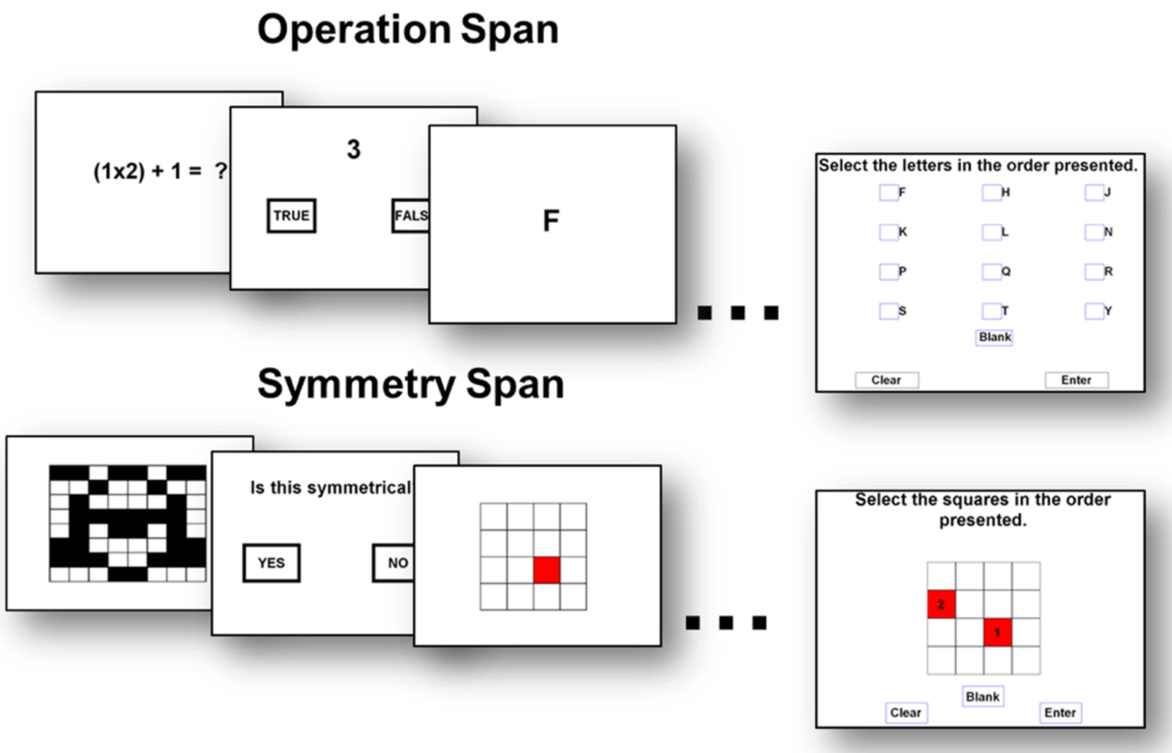
Working Memory Span Tasks (Complex Span Tasks)

Domain-general mechanism:

Attention control

Domain-specific storage:

Numerical, Spatial, etc.



PROBLEMS

Navarro, Hao, Rosales, & Conway (2023)

An IRT Approach to The Measurement of Working Memory Capacity

Item response theory: Psychometric properties of items (discrimination, difficulty, etc.)

How verbal and spatial complex span tasks assess domain-general WM at the item-level

Differences in item properties reveal influences of domain-general WM and domain-specific storage for different task types (cognitive mechanisms)

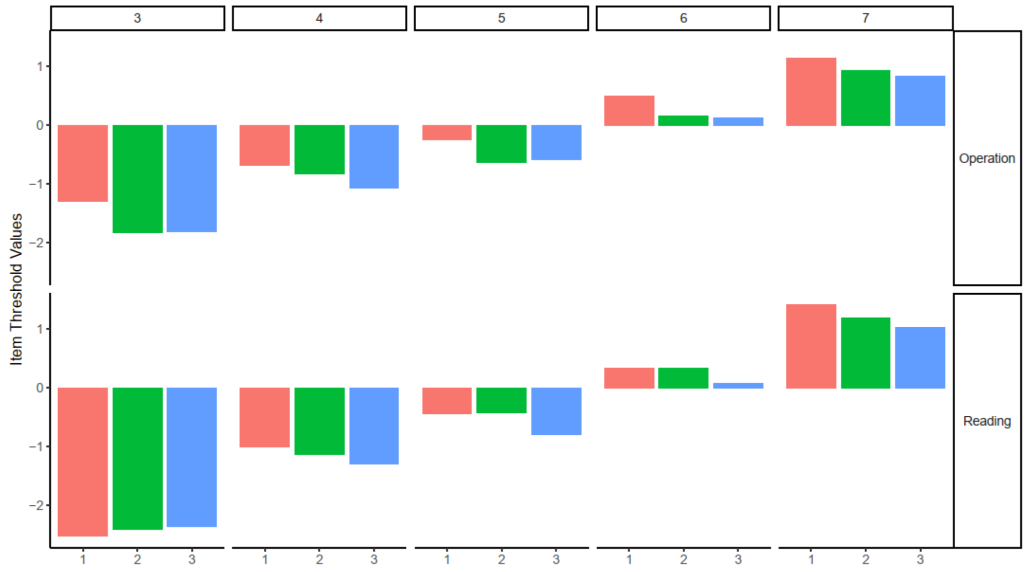
PROBLEMS

Navarro, Hao, Rosales, & Conway (2023)

Item Difficulty (Y-Axis) Plots by Blocks (Colors) and Item Sizes (Panels)

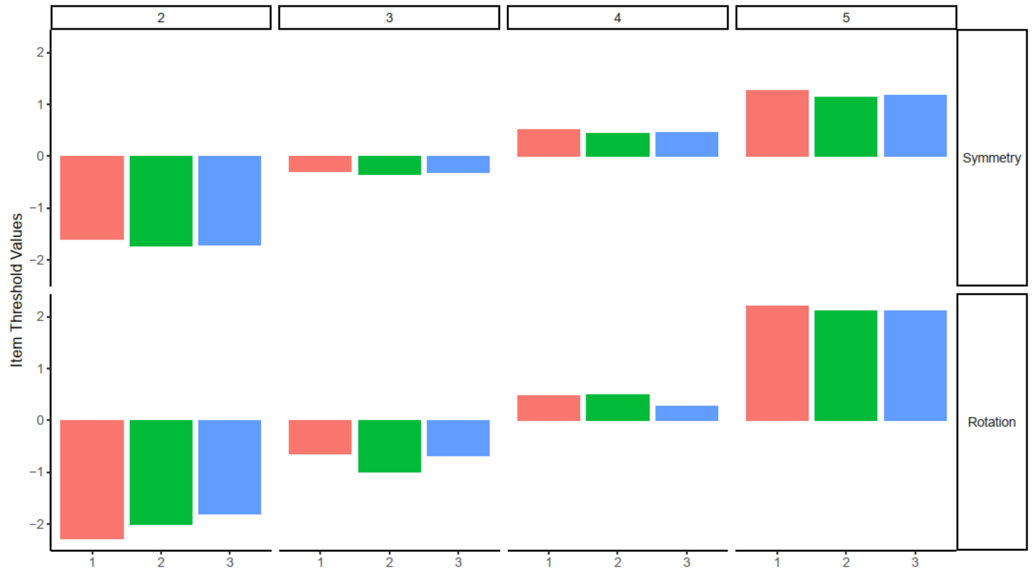
OSpan & RSpan

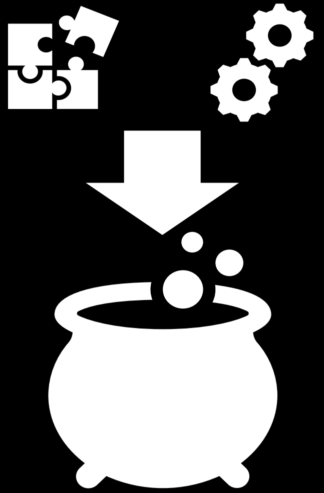
Figure 1. Visualizing Item Threshold Parameters for Verbal/Operational Tasks by Blocks and Set Sizes



SymSpan & RotSpan

Figure 2. Visualizing Item Threshold Parameters for Spatial Tasks by Blocks and Set Sizes





PROCESS OVERLAP THEORY

Bridge the gap between psychometric theories and cognitive theories

PROCESS OVERLAP THEORY

Kovacs & Conway (2016; 2019)

Attempts to explain **inter-individual** differences in cognitive abilities in terms of **intra-individual** psychological processes

Proposes an alternative cognitive foundation of the positive manifold of intelligence (**formative *g***)

PROCESS OVERLAP THEORY

A unified theory of intelligence based on the sampling theories (Thomson, 1916)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	+	+	+						+	+							
2		+	+	+						+	+						
3	+			+	+		+	+				+					
4	+	+		+	+	+						+	+				
5					+	+	+			+					+		+
6						+	+	+	+						+		

POT SAMPLING MECHANISMS (POT-V)

1. Domain-general and domain-specific processes are sampled in **an overlapping manner** across tests, no process is sampled in all tests
2. Domain-general processes are sampled more often than domain-specific processes across different tasks
3. Domain-general processes are also sampled more often in fluid reasoning tasks than in domain-specific tasks
4. The sampled processes are **compensatory** within each domain and **non-compensatory** across domains

POT IRT MODEL (POT-I)

$$P(U_{pi} = 1 \mid \Theta_{plm}, a_{il}, b_{il}) = \prod_{l=1}^{D} \frac{e^{\sum_{m=1}^{C} a_{il}(\Theta_{plm} - b_{il})}}{1 + e^{\sum_{m=1}^{C} a_{il}(\Theta_{plm} - b_{il})}}$$

where:

Θ_{plm} = the process score for the p^{th} person on the m^{th} process of the l^{th} domain

a_{il} = the discrimination parameter for the l^{th} domain on the i_{th} item

b_{il} = the difficulty parameter for the l^{th} domain on the i_{th} item

D = number of domains tapped by the item

C = number of processes in the given domain tapped by the item

POT STRUCTURAL MODEL (POT-S)

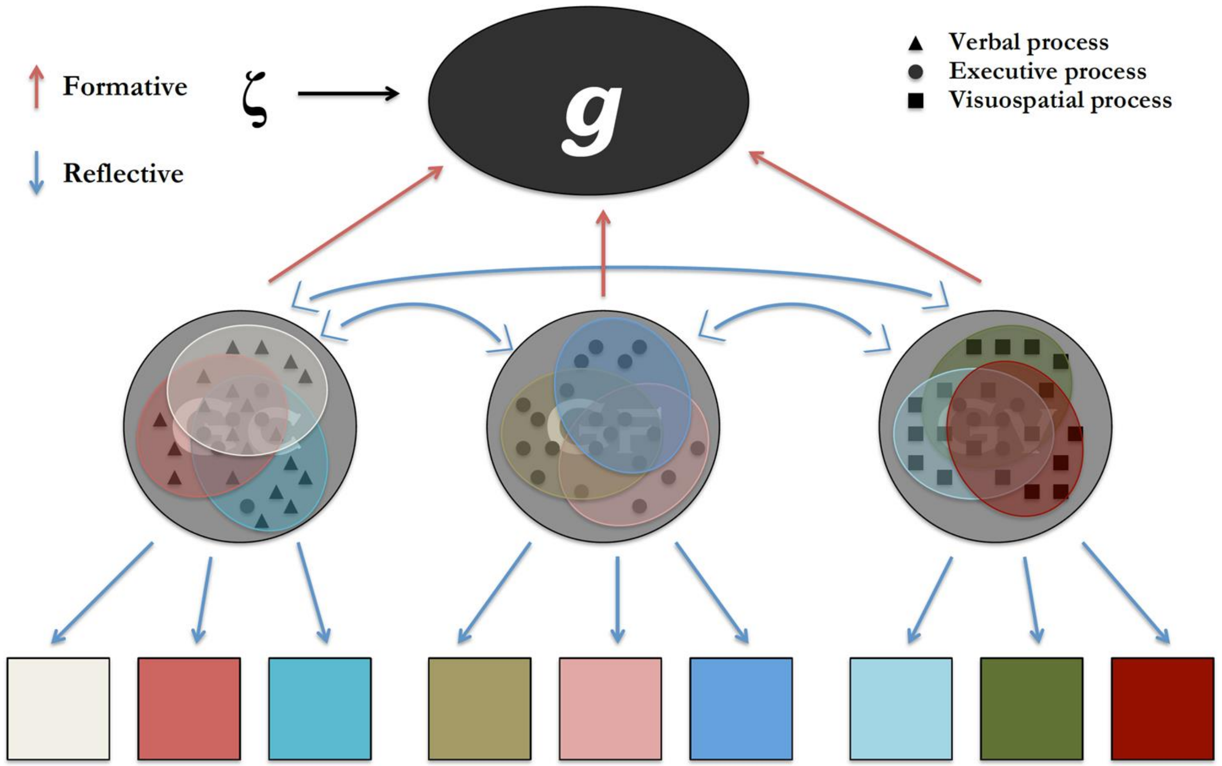


Figure 8, Kovacs & Conway (2016)

SIMULATING THE PROCESS OVERLAP THEORY

A Unified Framework Bridging Psychometric and Cognitive Perspectives

Hao, Conway, Kovacs, & Snijder (2023)

SIMULATING POT

The simulation translates the conceptual and IRT model of POT to sampling algorithms on simulated matrices and demonstrates:

- A) "*g*" can emerge from the simulated test scores in the absence of a general cognitive ability
- B) the broad ability factors can emerge by introducing a distinction between domain-general and domain-specific processes, and how they are sampled by different types of tests

SIM PROCEDURES

Simulate a sample of 1000 subjects performing 9 tests

- Fluid, Verbal, Spatial (3 for each type)

Each subject has a set of 60 cognitive processes

- EF, Reasoning, Verbal, Spatial (15 for each type)

Apply 2 specific sampling algorithms to the simulated processes

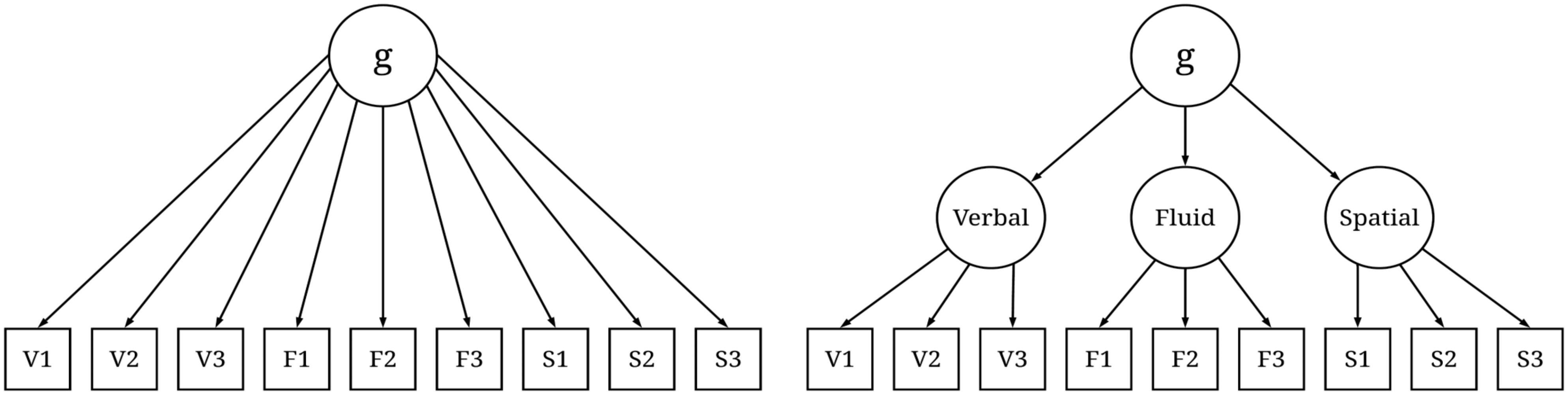
- The general sampling algorithm (Thomson) vs The POT algorithm (POT)

6 processes/test; none of the cognitive processes was sampled in all tests

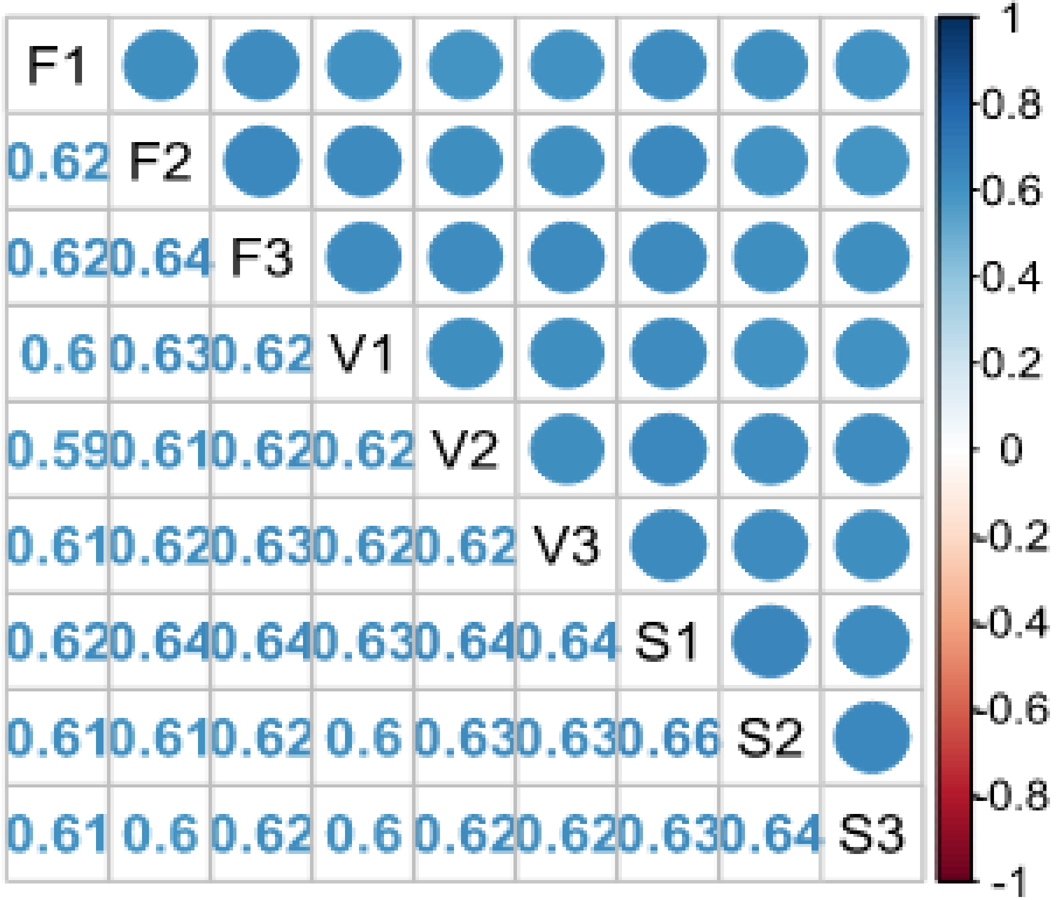
- So no general cognitive ability!

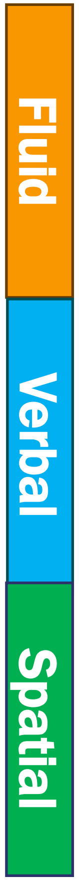
SIM PROCEDURES

Fit psychometric models to simulated data (200 iterations)

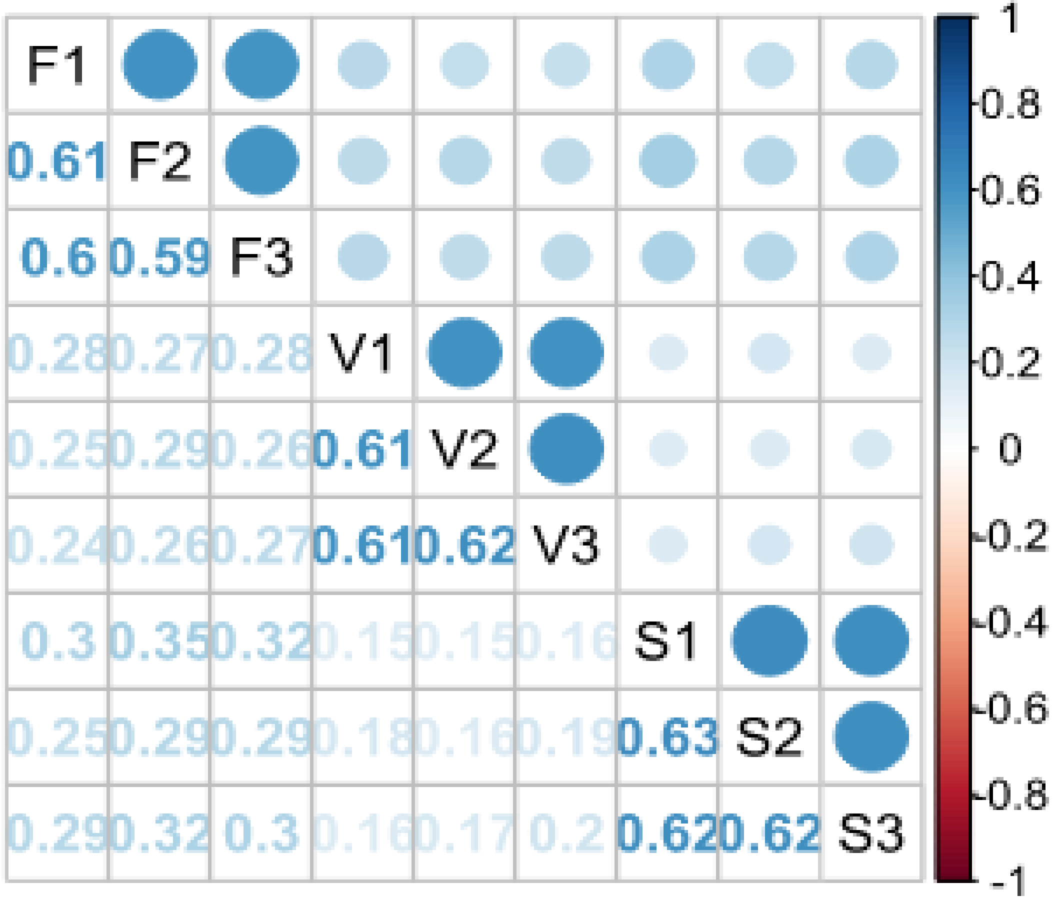


Correlation Matrix for Thomson Algorithm



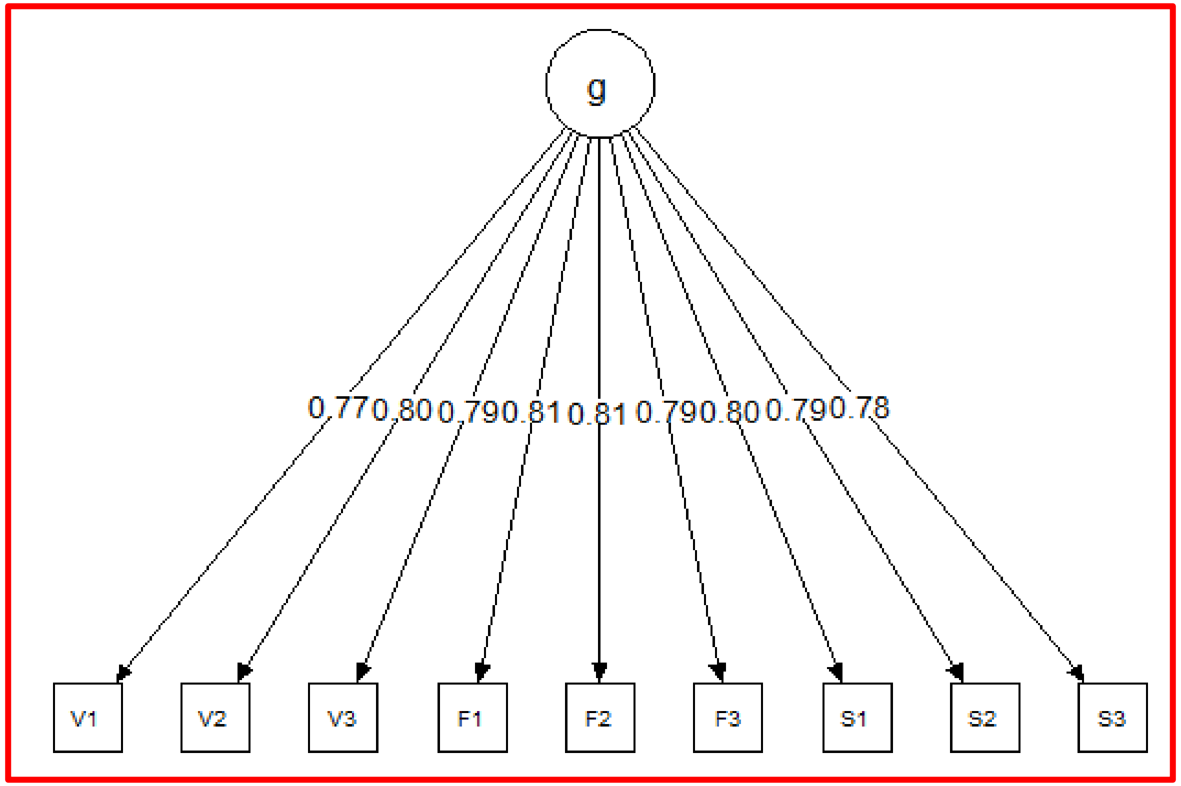


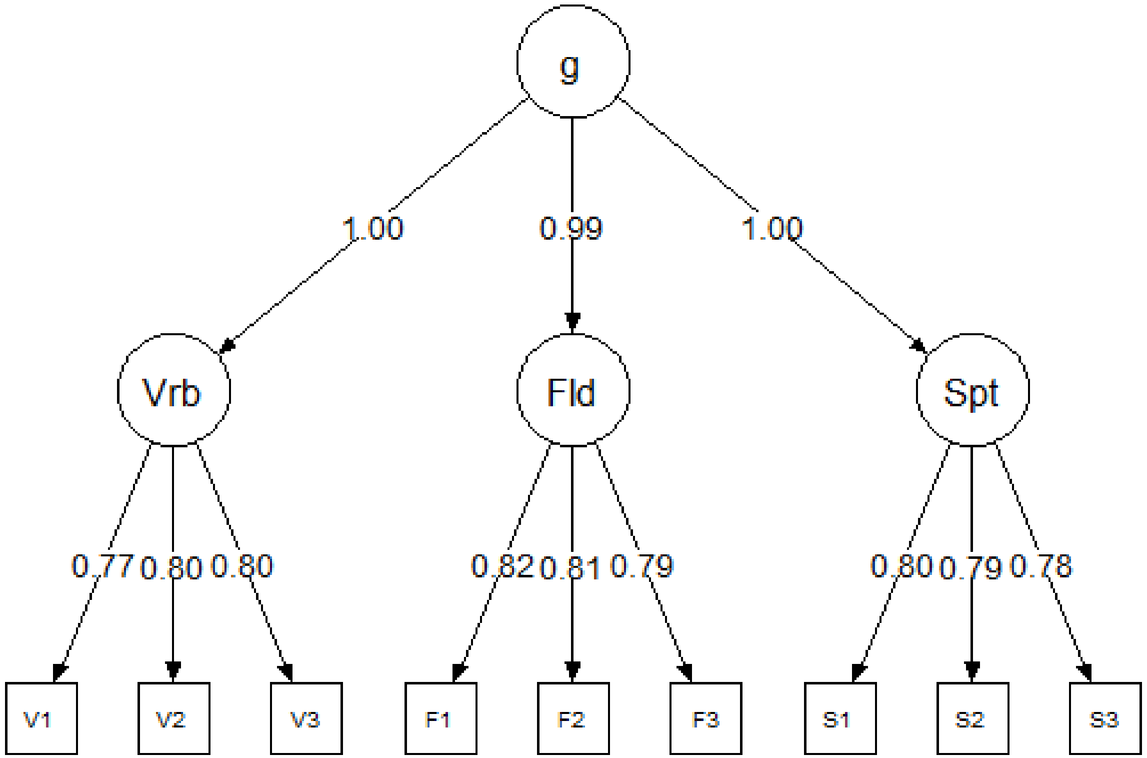
Correlation Matrix for POT Algorithm



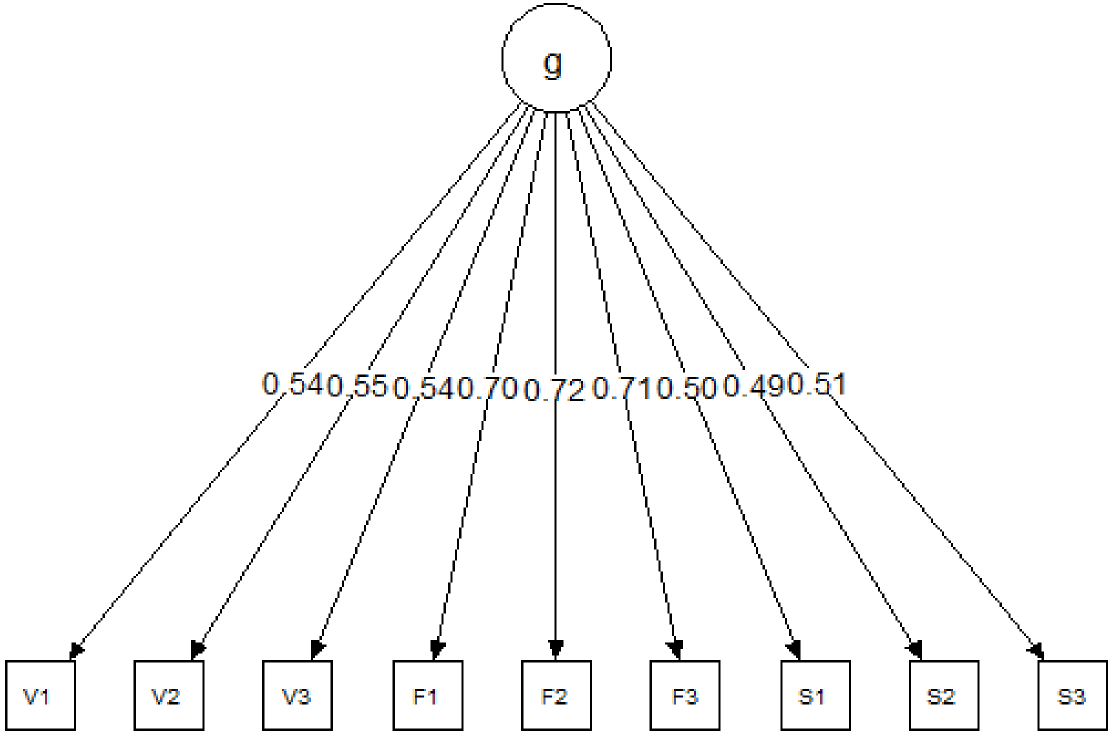
Iterations: 200

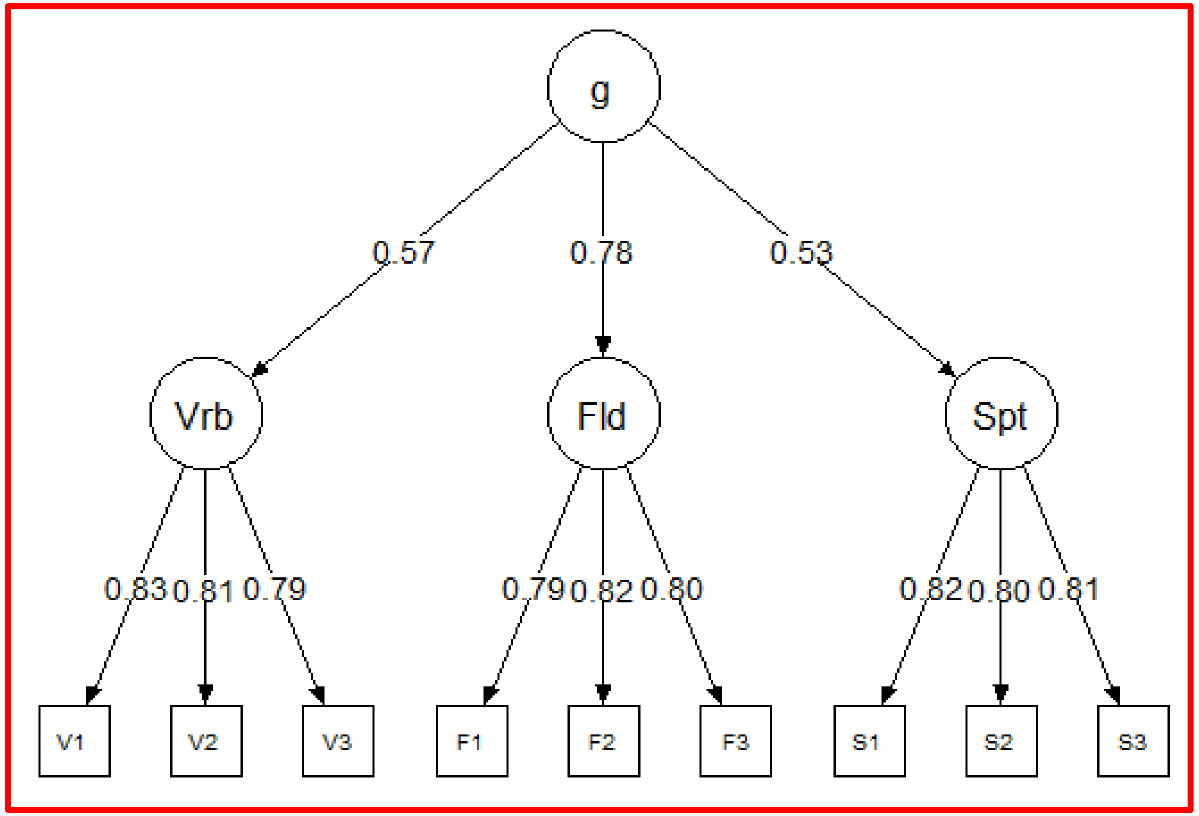
SIM RESULTS (THOMSON ALGORITHM)





SIM RESULTS (POT ALGORITHM)





SUMMARY OF FIT INDICES

Means (Standard Deviations) for the Fit Indices from the Models Based on Simulated Data from 200 Iterations (by the Algorithm and the Model Structure)

		χ^2	CFI	RMSEA	SRMR	AIC
POT	Higher-Order	26.21 (6.57)	1.00 (< .01)	.01 (.01)	.02 (<.01)	58565.72 (138.62)
	One-Factor	1637.46 (87.18)	.56 (.02)	.24 (.01)	.15 (.01)	60170.97 (169.36)
GSM	Higher-Order	26.77 (7.49)	.99 (< .01)	.01 (.01)	.01 (<.01)	56411.03 (127.53)
	One-Factor	27.69 (7.65)	1.00 (< .01)	.01 (.01)	.01 (<.01)	56405.96 (127.42)

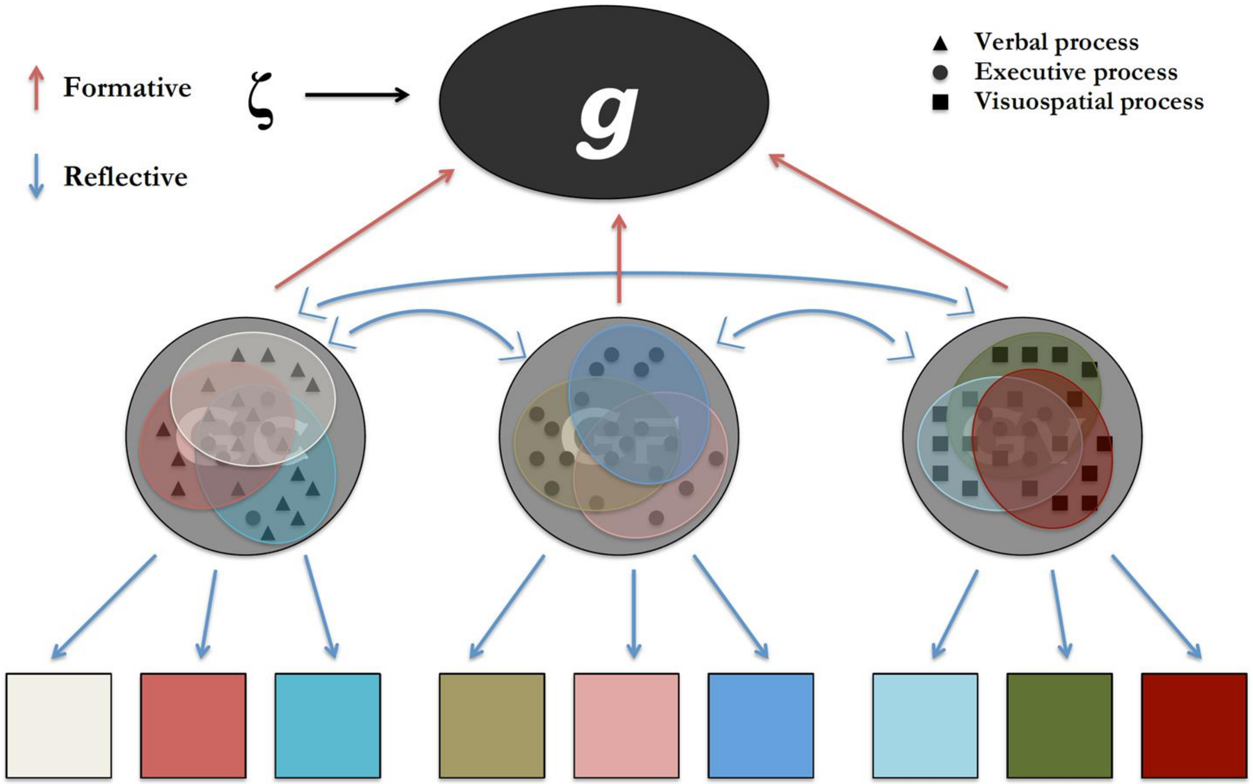
INTERPRETATION

For both POT and GSM algorithms, a positive manifold emerged from the simulated test scores **in the absence of a general cognitive ability**

Results from the POT algorithm is aligned with real-world observations:

- The **higher-order** structure of cognitive abilities
- The **high loading of fluid** subfactor on higher-order *g*

POT STRUCTURAL MODEL



A PSYCHOMETRIC NETWORK OF THE POT SIMULATION

Apply a Psychometric Network to the Simulation (POT-N)

Extending the original simulation results by applying a network structure to the psychometric model of POT (**POT-N**)

A NETWORK MODEL OF POT

Conway, Kovacs, Hao, Goring, Schmank, 2020

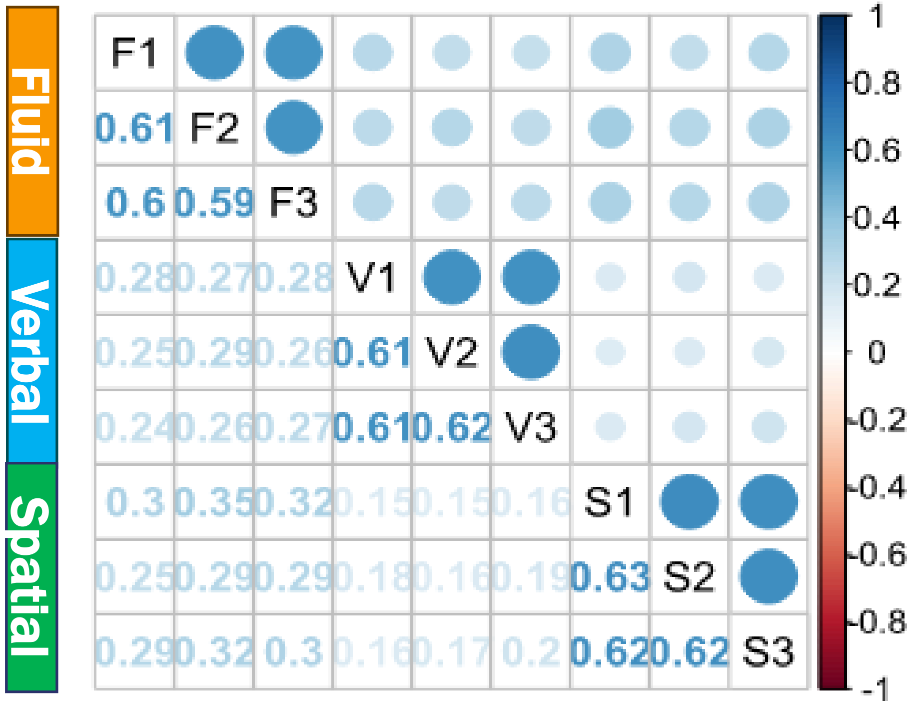
The Struggle Is Real: Challenges & Solutions in Theory Building

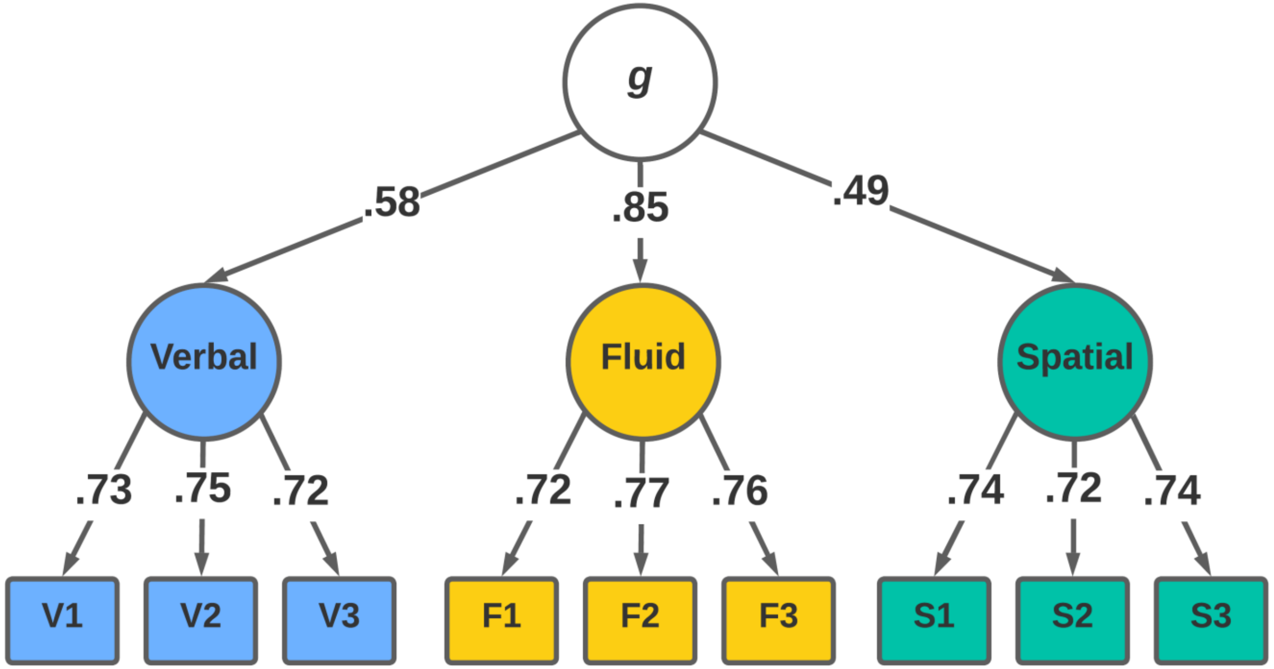
Why POT-N?

Theory Building: Factor Models vs. Network Models

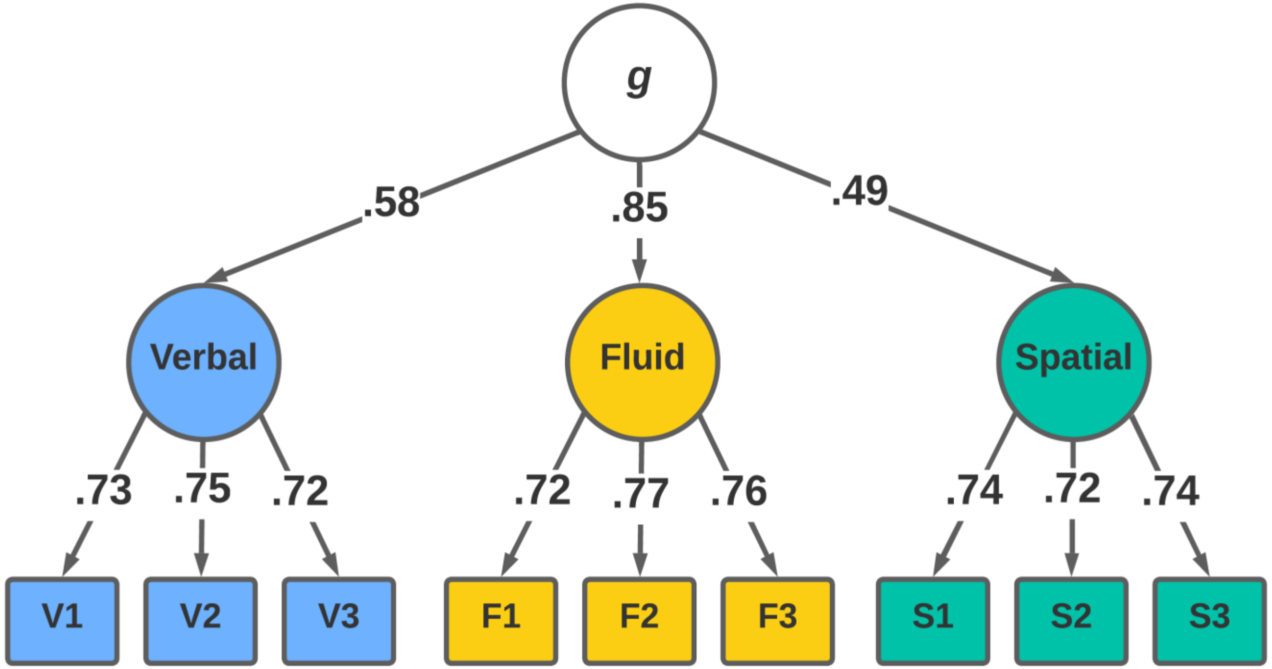
- An alternative representation to the positive manifold
- Shifts the main emphasis from a common cause to the direct mutual associations among specific cognitive measures

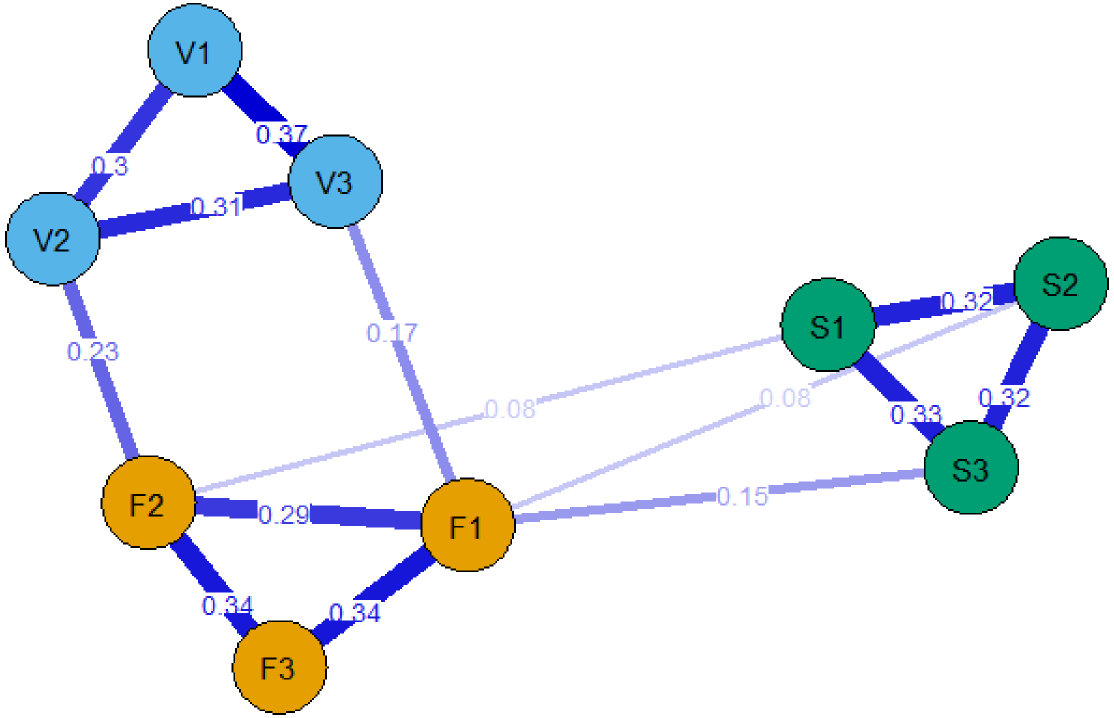
A NETWORK MODEL OF POT





A NETWORK MODEL OF POT





TAKEAWAYS

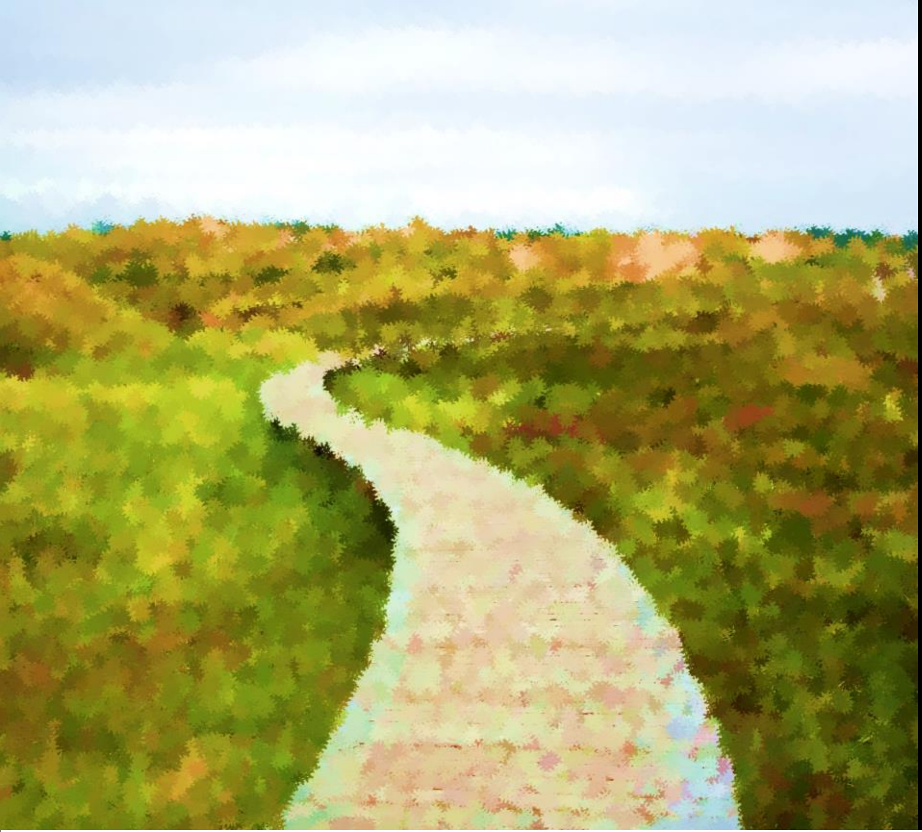
The simulation algorithm illustrated a cognitive mechanism based on POT and reflected it in psychometric models

The positive manifold and the higher-order g can be achieved without a general cognitive process as the common cause

Domain-general processes such as those in WM and EF are central to various cognitive behaviors

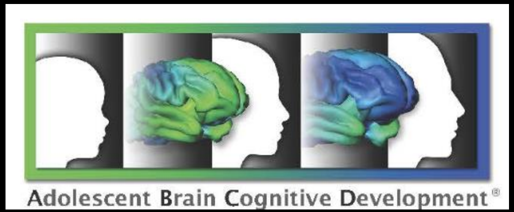
The network model proposed an alternative psychometric interpretations of individual differences in cognitive abilities based on POT

MOVING FORWARD..



- Improve the simulation framework
  - Reaction time, drift-diffusion model (In prep.)
- Understanding cognitive processes
  - Experimental tasks on cognitive processes (In prep.)
- Psychometric Network modeling
  - Empirical data (Manuscript)

MOVING FORWARD..

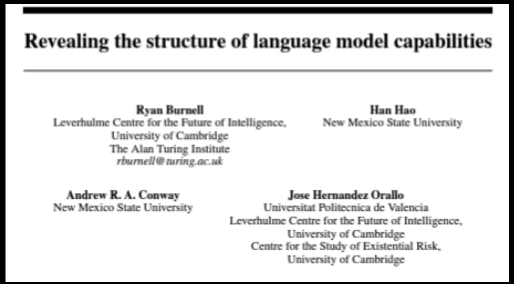


Cognitive development

Longitudinal analyses w/ ABCD Study (In prep.)

Machine psychology

Cognitive behaviors of AI (Preprints)



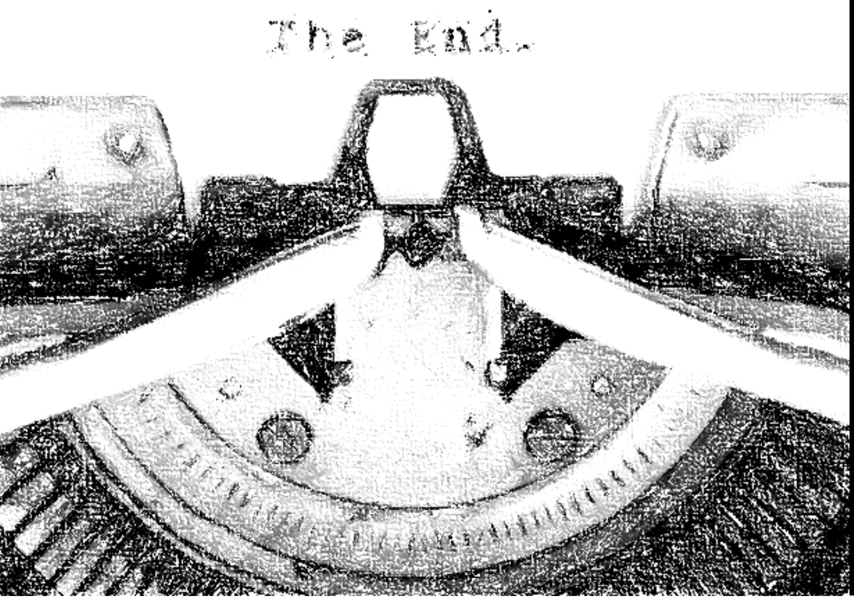
Revealing the structure of language model capabilities

Ryan Burnell
Leverhulme Centre for the Future of Intelligence,
University of Cambridge
The Alan Turing Institute
rburnell@turing.ac.uk

Han Hao
New Mexico State University

Andrew R. A. Conway
New Mexico State University

Jose Hernandez Orallo
Universitat Politecnica de Valencia
Leverhulme Centre for the Future of Intelligence,
University of Cambridge
Centre for the Study of Existential Risk,
University of Cambridge



THANK YOU!

Han Hao **05/08/2024**

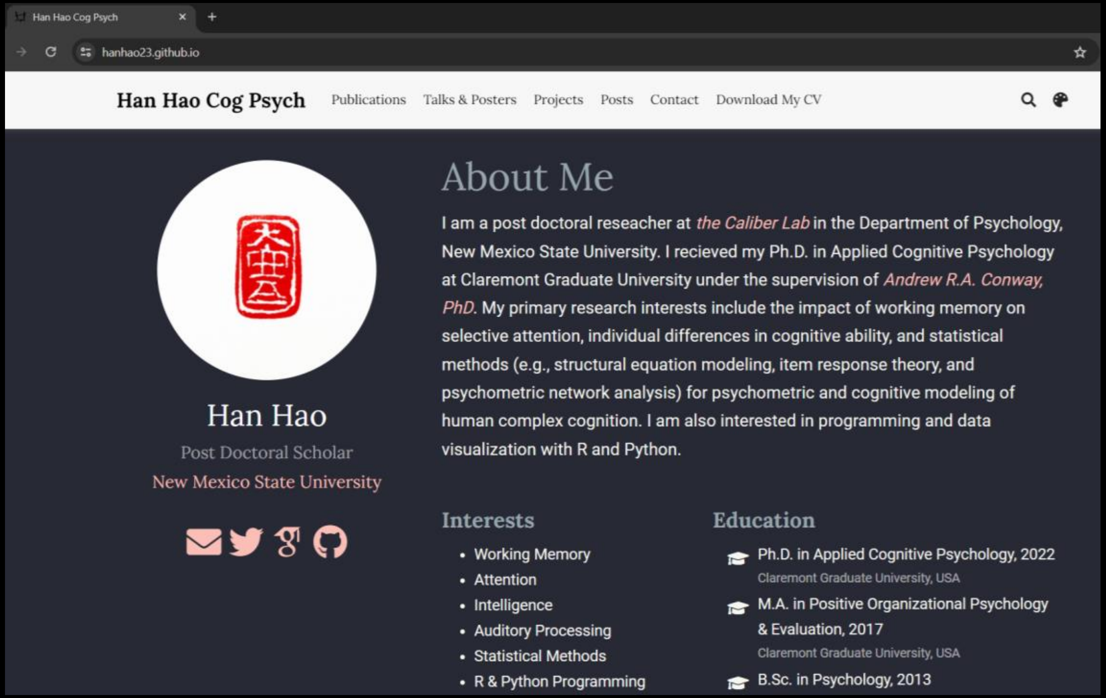
Slides & Materials Available:

https://hanhao23.github.io/talk/tarletonjobtalk/

Web: https://hanhao23.github.io

Email: hanhao@nmsu.edu

X: https://twitter.com/HanHao23



Han Hao Cog Psych

hanhao23.github.io

Han Hao Cog Psych — Publications | Talks & Posters | Projects | Posts | Contact | Download My CV

Han Hao

Post Doctoral Scholar

New Mexico State University

About Me

I am a post doctoral reseacher at *the Caliber Lab* in the Department of Psychology, New Mexico State University. I recieved my Ph.D. in Applied Cognitive Psychology at Claremont Graduate University under the supervision of *Andrew R.A. Conway, PhD*. My primary research interests include the impact of working memory on selective attention, individual differences in cognitive ability, and statistical methods (e.g., structural equation modeling, item response theory, and psychometric network analysis) for psychometric and cognitive modeling of human complex cognition. I am also interested in programming and data visualization with R and Python.

Interests

- Working Memory
- Attention
- Intelligence
- Auditory Processing
- Statistical Methods
- R & Python Programming

Education

- Ph.D. in Applied Cognitive Psychology, 2022
  Claremont Graduate University, USA
- M.A. in Positive Organizational Psychology & Evaluation, 2017
  Claremont Graduate University, USA
- B.Sc. in Psychology, 2013

THE CALIBER LAB

https://caliberlab.wixsite.com

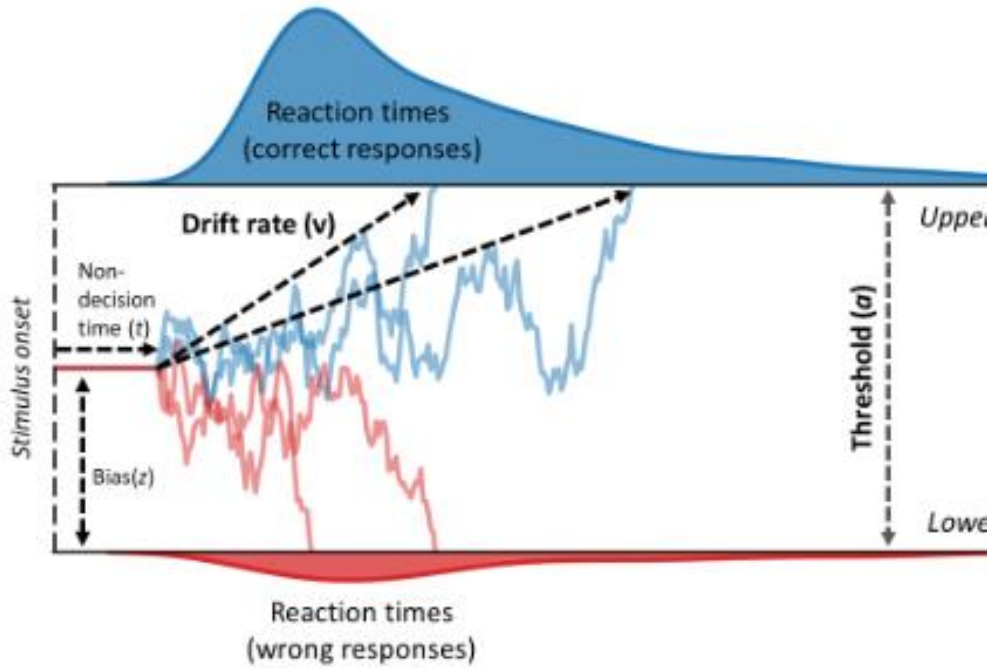
MOVING FORWARD...

For the simulation algorithm:

Improving the sampling algorithm of POT

Incorporating the drift-diffusion model to account for reaction time measures

Assumption tests based on algorithms and parameters

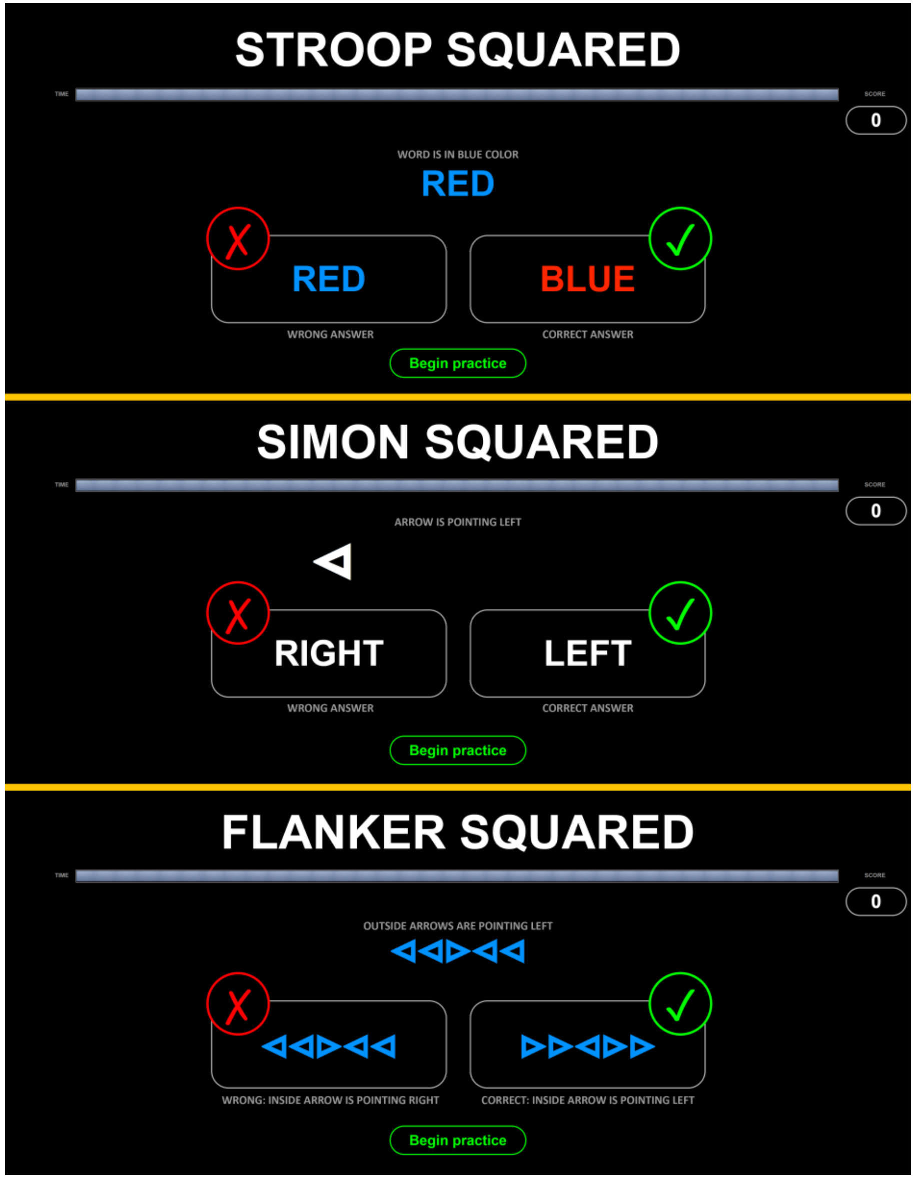


MOVING FORWARD...

For the experimental approach:

Deconstructing the latent construct of attention control measured by the Squared Tasks (Burgoyne et al., 2023)

A “square root project”

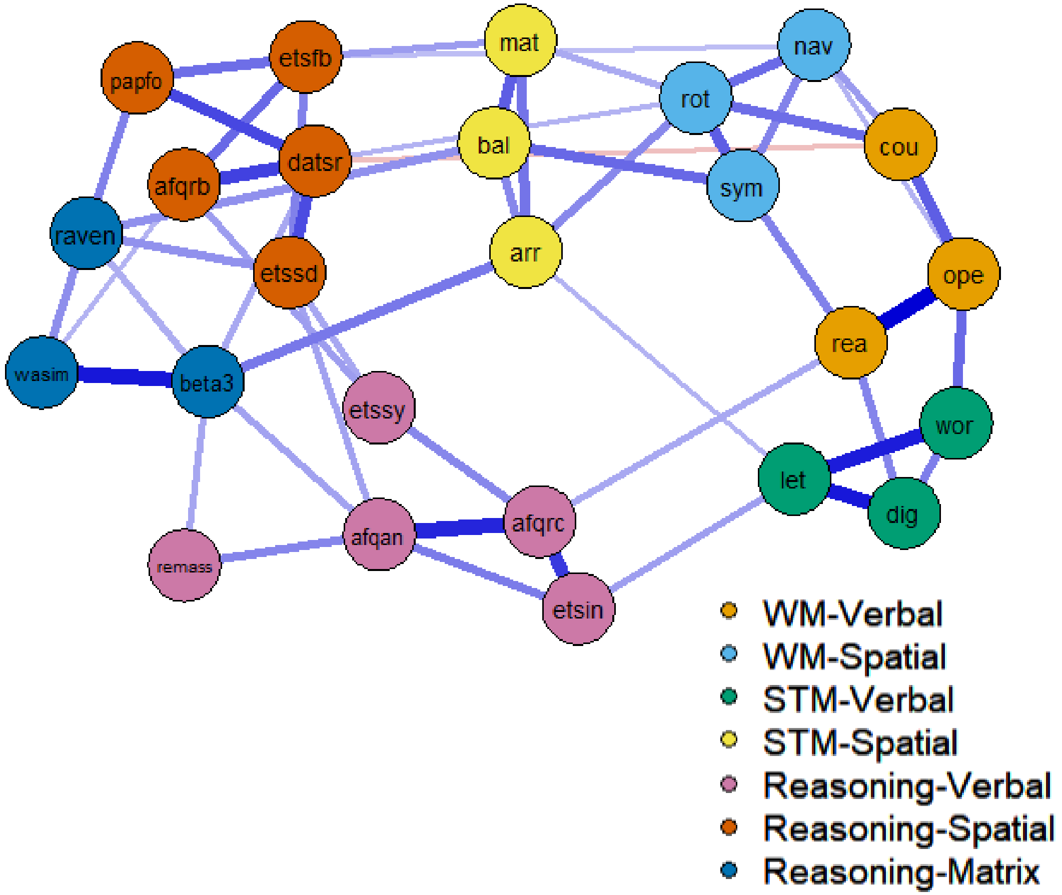


MOVING FORWARD...

For the psychometrical approach:

Network Modeling on Empirical data of cognitive abilities

Working memory & Reasoning (Kane et al., 2004)

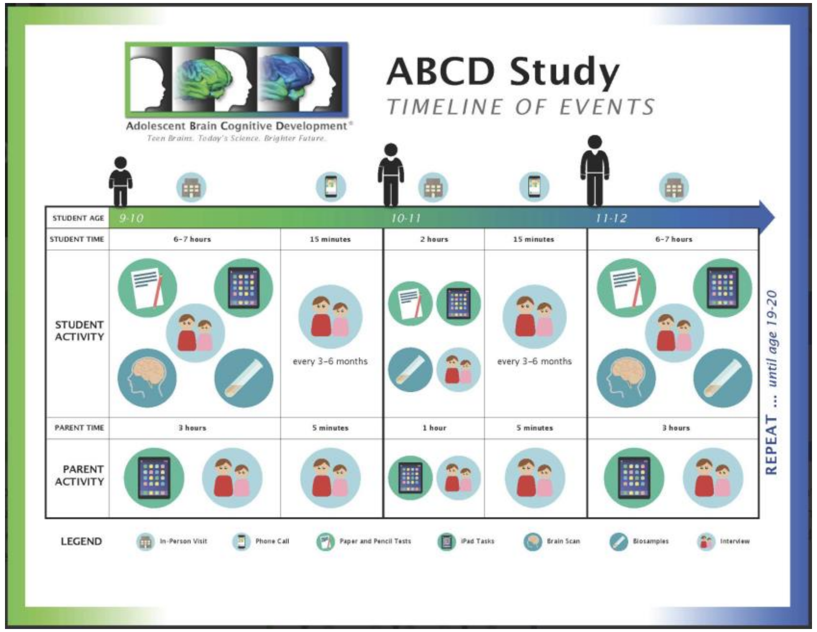


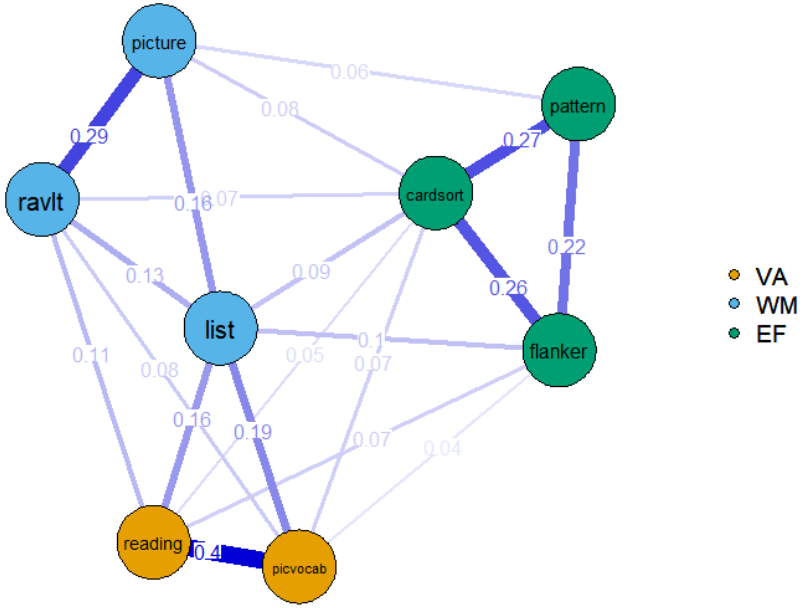
MOVING FORWARD…

For a longitudinal approach:

The Adolescent Brain Cognitive Development (ABCD) Study

Network inspection on adolescents' cognitive development





MOVING FORWARD…

A machine cognition approach:

The understanding of Cognitive abilities of AI could help the understanding of human cognition

Psychometrics

Cognitive behaviors (ToM)

Revealing the structure of language model capabilities

Ryan Burnell
Leverhulme Centre for the Future of Intelligence,
University of Cambridge
The Alan Turing Institute
rburnell@turing.ac.uk

Han Hao
New Mexico State University

Andrew R. A. Conway
New Mexico State University

Jose Hernandez Orallo
Universitat Politecnica de Valencia
Leverhulme Centre for the Future of Intelligence,
University of Cambridge
Centre for the Study of Existential Risk,
University of Cambridge

Abstract

Building a theoretical understanding of the capabilities of large language models (LLMs) is vital for our ability to predict and explain the behavior of these systems. Here, we investigate the structure of LLM capabilities by extracting latent capabilities from patterns of individual differences across a varied population of LLMs. Using a combination of Bayesian and frequentist factor analysis, we analyzed data from 29 different LLMs across 27 cognitive tasks. We found evidence that LLM capabilities are not monolithic. Instead, they are better explained by three well-delineated factors that represent reasoning, comprehension and core language modeling. Moreover, we found that these three factors can explain a high proportion of the variance in model performance. These results reveal a consistent structure in the capabilities of different LLMs and demonstrate the multifaceted nature of these capabilities. We also found that the three abilities show different relationships to model properties such as model size and instruction tuning. These patterns help refine our understanding of scaling laws and indicate that changes to a model that improve one ability might simultaneously impair others. Based on these findings, we suggest that benchmarks could be streamlined by focusing on tasks that tap into each broad model ability.

Sim Procedure – 1/4

- **Step 1: Specify the cognitive processes and tests**
- Simulate a sample of **1000** subjects performing **9** tests, each has **100** items
  - **3** fluid reasoning tests, **3** verbal tests, **3** spatial tests

- Each subject has a set of **60** cognitive processes
  - **15** Executive Function (EF) Processes
  - **15** Fluid Reasoning Processes, **15** Verbal Processes, and **15** Spatial Processes

- Each individual subject has an ability level on each process (orthogonal and normally distributed)
  - A 1000 x 60 Matrix

Sim Procedure – 2/4

- **Step 2: Apply 2 specific sampling algorithms to the simulated processes (POT and GSM)**
- The general sampling algorithm (GSM):
  - All **60** processes are sampled with equal probability ($p = .10$) across every task and item
  - For a specific item, about **60*0.10 = 6** processes are expected to be sampled

Sim Procedure – 2/4 (continued)

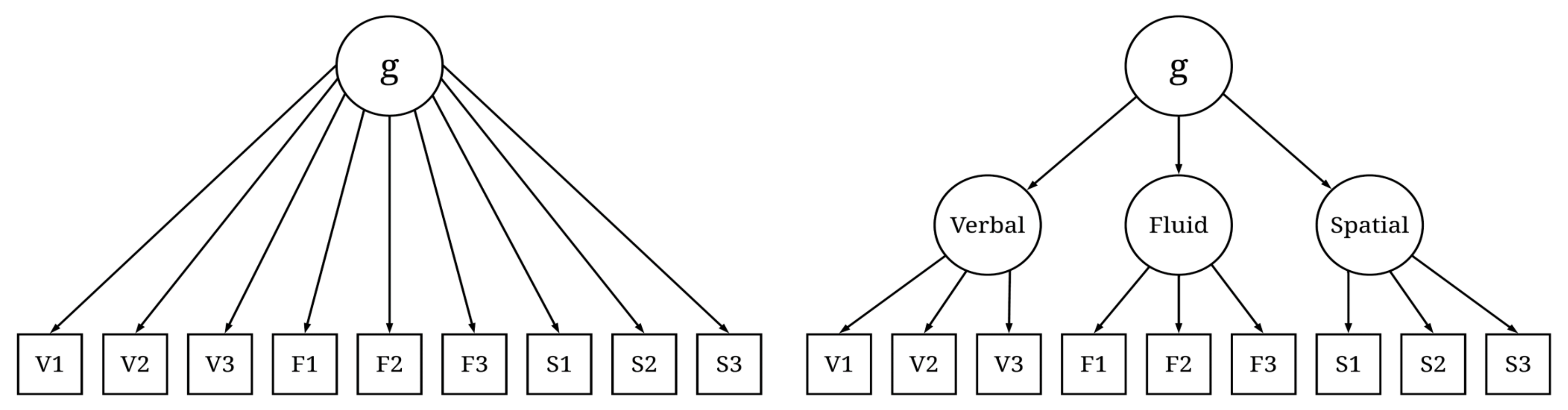
- **Step 2: Apply 2 specific sampling algorithms to the simulated processes (POT and GSM)**
- The POT algorithm:
  - For an item in Gf tests, domain-general (EF) processes are sampled with greater probability (p = **.28**) than domain-specific (Fluid Reasoning) processes (p = **.12**)
  - On average, **4** EF (15*0.28) + **2** Fluid (15*0.12) processes are expected to be sampled for an item
  - For an item in verbal/spatial tests, domain-general (EF) processes are sampled with smaller probability (p = **.12**) than domain-specific (Verbal/Spatial) processes (p = **.28**)
  - On average, **2** EF (15*0.28) + **4** specific (15*0.12) processes are expected to be sampled for an item

Sim Procedure – 3/4

- **Step 3: Calculate item scores from the 2 algorithms**
  - The GSM Algorithm:
    - all sampled processes are summed and standardized to calculated the corresponding "latent trait" required for an item
  - The POT Algorithm:
    - the processes within a domain are summed and standardized as the dimensional "latent trait"
  - The "Latent traits" are converted to probabilities by IRT functions (logistic functions) and are used to generate binary responses of items (0s and 1s)
  - 1000 (subjects) × 9 (tests) × 100 (items)

Sim Procedures – 4/4

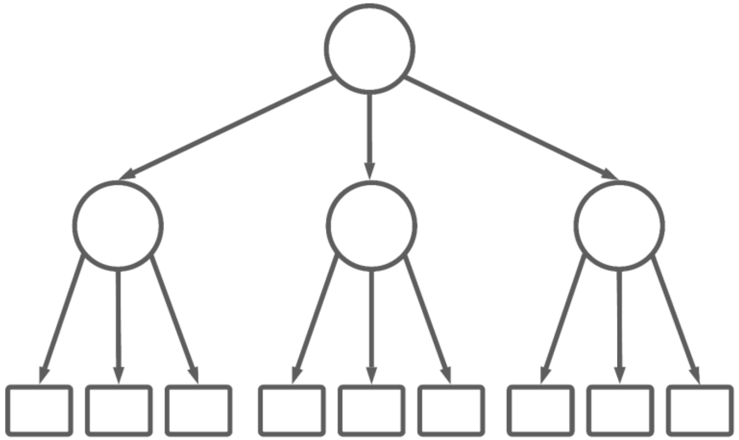
- **Step 4: Fit psychometric models to simulated data**

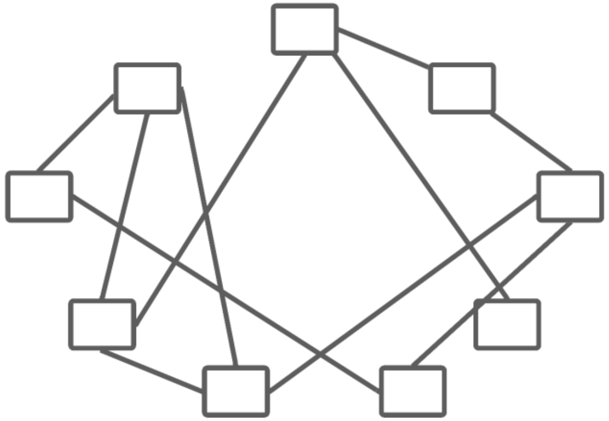


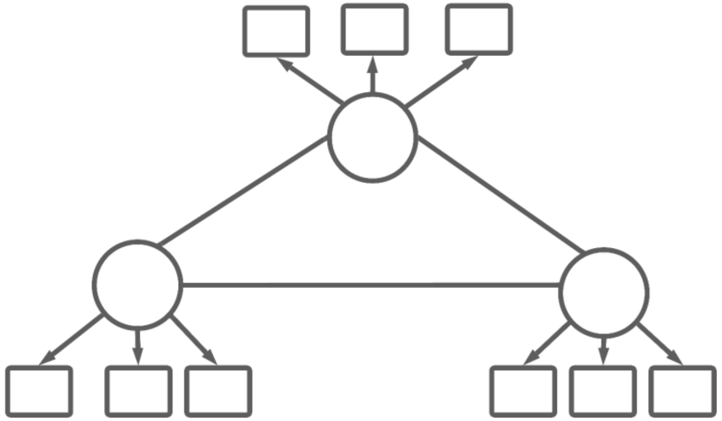
- **GSM – The One-Factor Model**
- **POT data – The Higher-Order Model**

PSYCHOMETRIC NETWORK

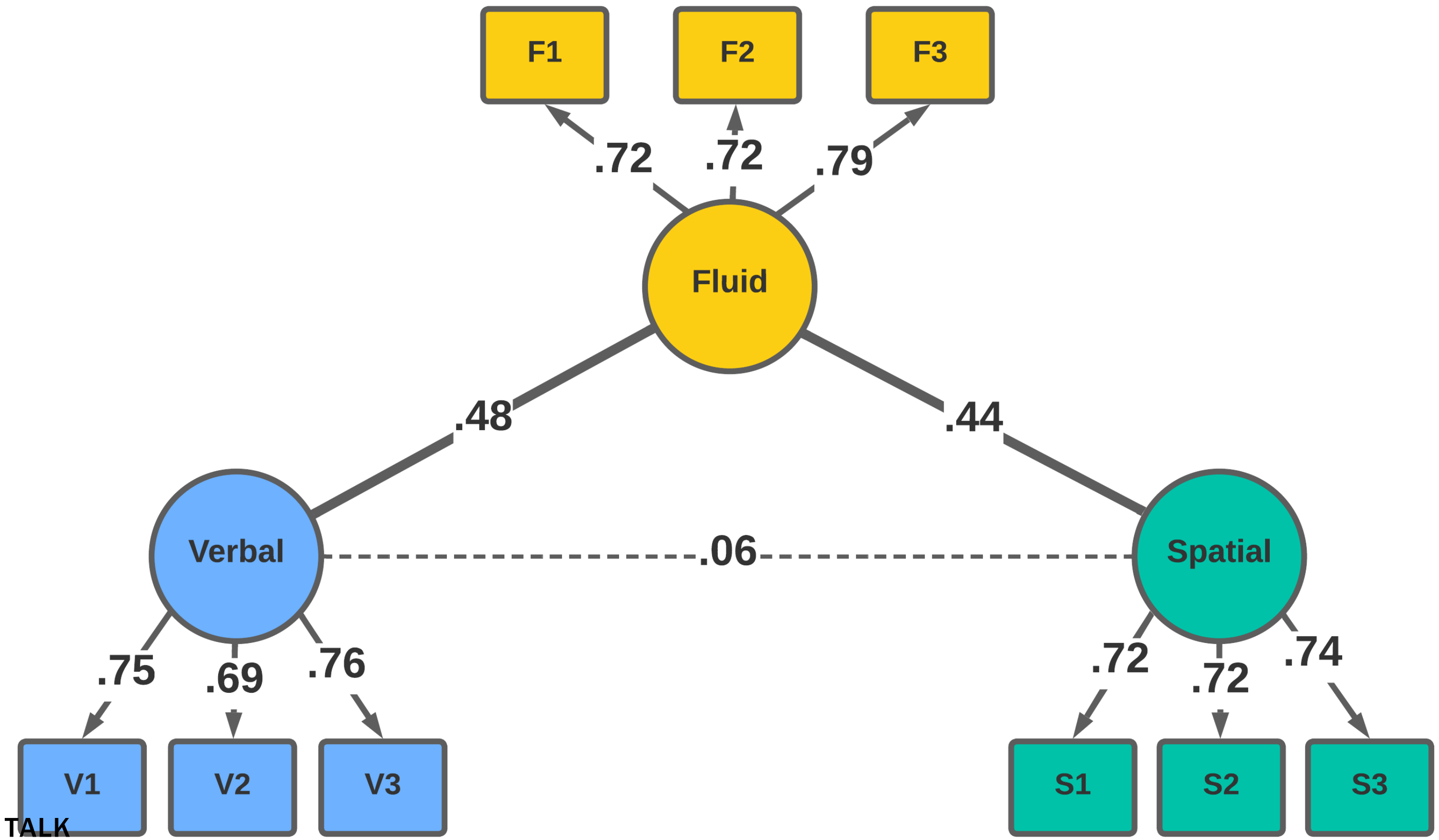
Three types of models on a simulated dataset of POT algorithm



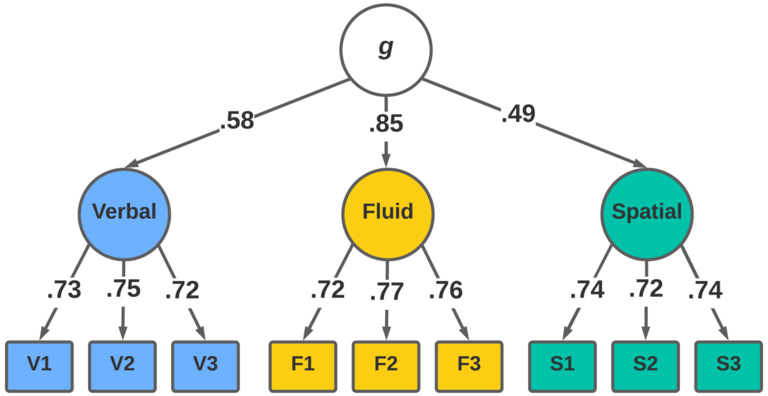


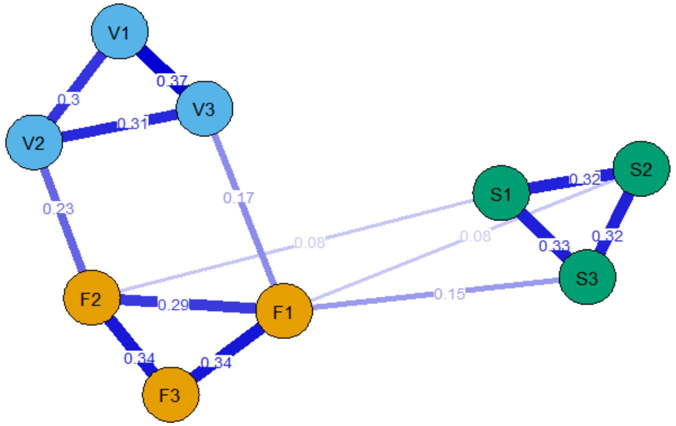


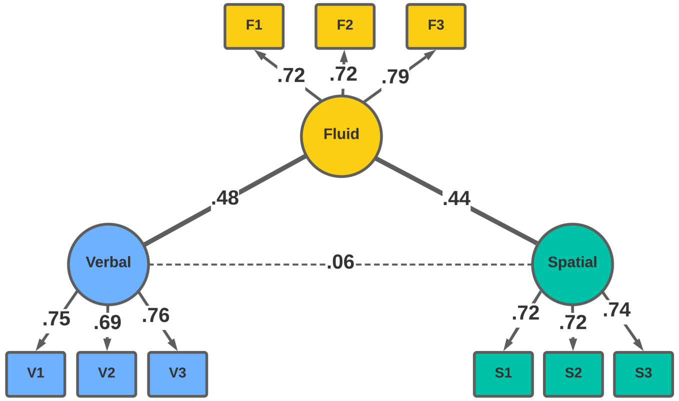
RESULTS - LATENT NETWORK



COMPARING THE LATENT CONSTRUCTS

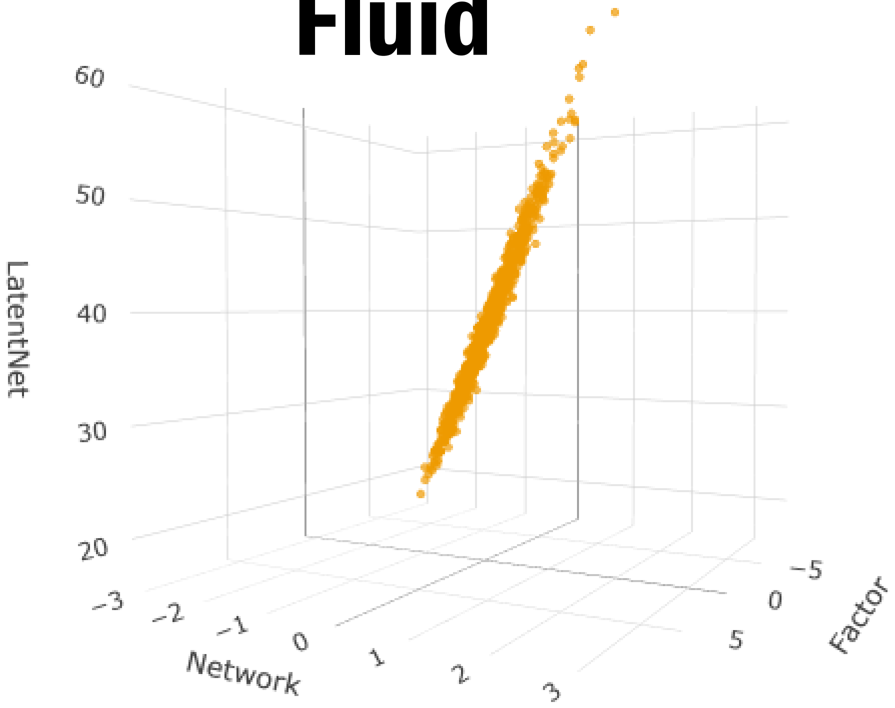




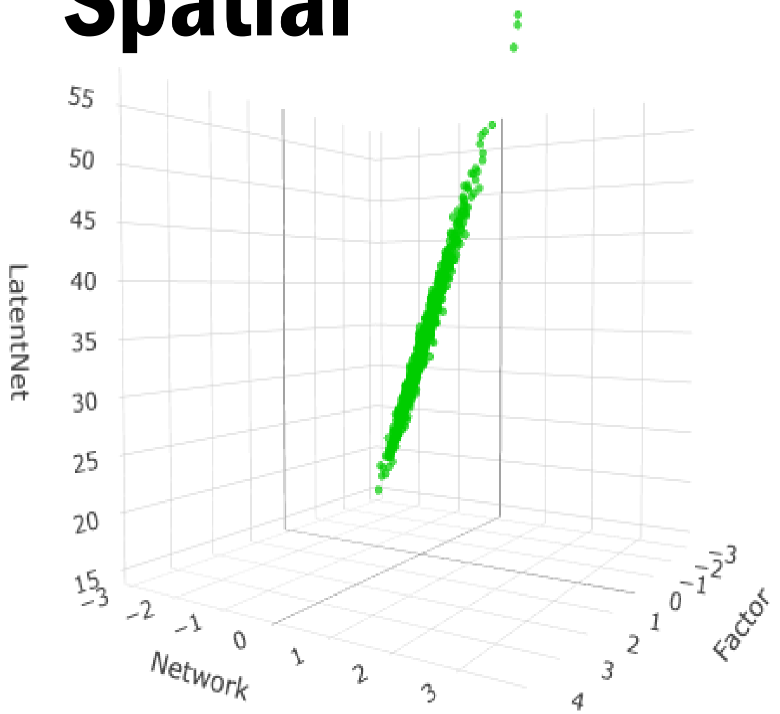


RESULTS – FACTOR/CLUSTER SCORES

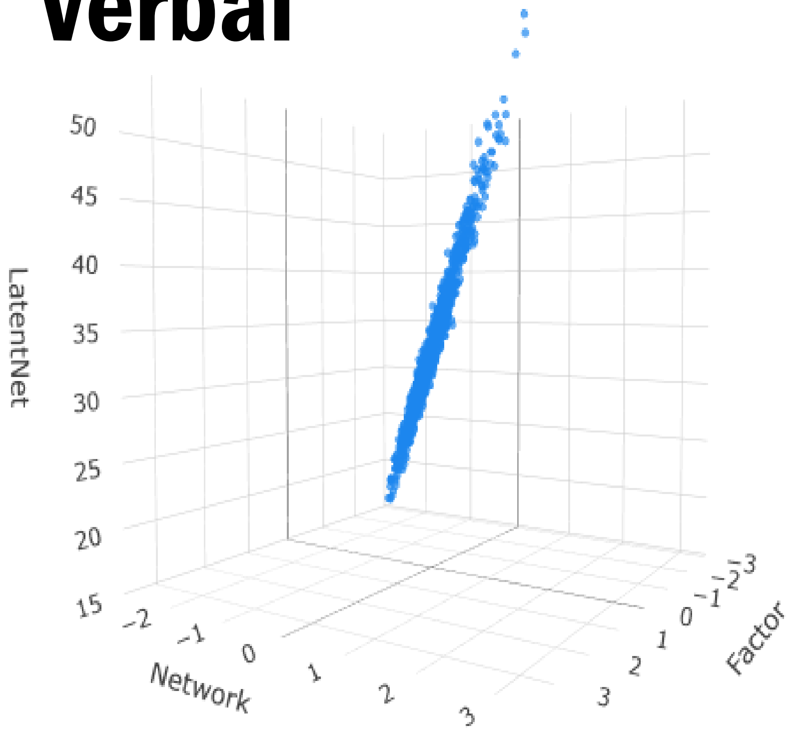
Fluid



Spatial



Verbal



RESULTS – FACTOR/CLUSTER SCORES

		Network Cluster Scores		
		Fluid	Verbal	Spatial
Latent Factor Scores	Fluid	**0.99**	0.62	0.61
	Verbal	0.62	**0.99**	0.36
	Spatial	0.61	0.36	**0.99**

		LatentNet Factor Scores		
		Fluid	Verbal	Spatial
Latent Factor Scores	Fluid	**0.99**	0.60	0.58
	Verbal	0.59	**0.99**	0.33
	Spatial	0.58	0.33	**0.99**

		LatentNet Factor Scores		
		Fluid	Verbal	Spatial
Network Cluster Scores	Fluid	**0.99**	0.62	0.61
	Verbal	0.61	**0.99**	0.36
	Spatial	0.62	0.36	**0.99**

EXECUTIVE FUNCTIONS

Executive functions (Diamond, 2013; Frischkorn et al., 2019)

- correlated with intelligence, but evidence is mixed

Conway, Kovacs, Hao, Rosales, & Snijder, 2021

Table 1. Glossary of common terms.

Cognitive control	A broad construct that refers to the regulation of information processing during goal-directed behavior. The execution of cognitive control requires executive attention processes, as defined below. The set of processes required depends on the goal, task, context, environment, and individual characteristics. Cognitive control is primarily, but not exclusively, dependent upon the prefrontal cortex and reflects the active maintenance of patterns of neural activity that represent goals and the means to achieve them (**Miller and Cohen 2001**).
Attentional control	A broad cognitive ability that refers to individual differences in cognitive control, as defined above (**Draheim et al. 2020**).
Executive function	A specific cognitive ability that refers to individual differences in cognitive control, as defined above. Functions are more specific than attentional control but more general than executive processes. Functions are defined at a level that is optimal for developmental/neuropsychological assessment, diagnosis, and treatment (**Friedman and Miyake 2017**).
Executive process	A low-level process involved in executive functions, attention control, and cognitive control. Processes are the most specific level in a cognitive model (**Oberauer 2009**).

COGNITIVE THEORIES

Executive functions (Diamond, 2013; Frischkorn et al., 2019)

- correlated with intelligence, but evidence is mixed