Max-Planck-Institut für Bildungsforschung

50

50 Years of Research on Human Development

Structural Equation Models for Cognitive Neuroscience: A whirlwind tour of principles and applications

Andreas M. Brandmaier & Rogier A. Kievit

Ringberg 2018

Outline

- 1) SEM – What, why and how?
- 2) Cross-sectional models of brain and cognition
- 3) Multi-region multi-method models of brain integrity
- 4) Models of change over time
- 5) Models of reliability
- 6) Coupled change

What Do We Want to Achieve with SEM?

- **Multivariate** analysis: Understand the relations *between* multiple variables
- **Latent** variable analysis: Variables may not be directly observable
- Input: A **covariance** matrix and means vector (or raw data)
- Goal: develop **a *simpler explanation*** of that covariance matrix
- Test whether your data is compatible with your hypothesis (i.e. not rejected by it)

Definition: Structural Equation Modeling

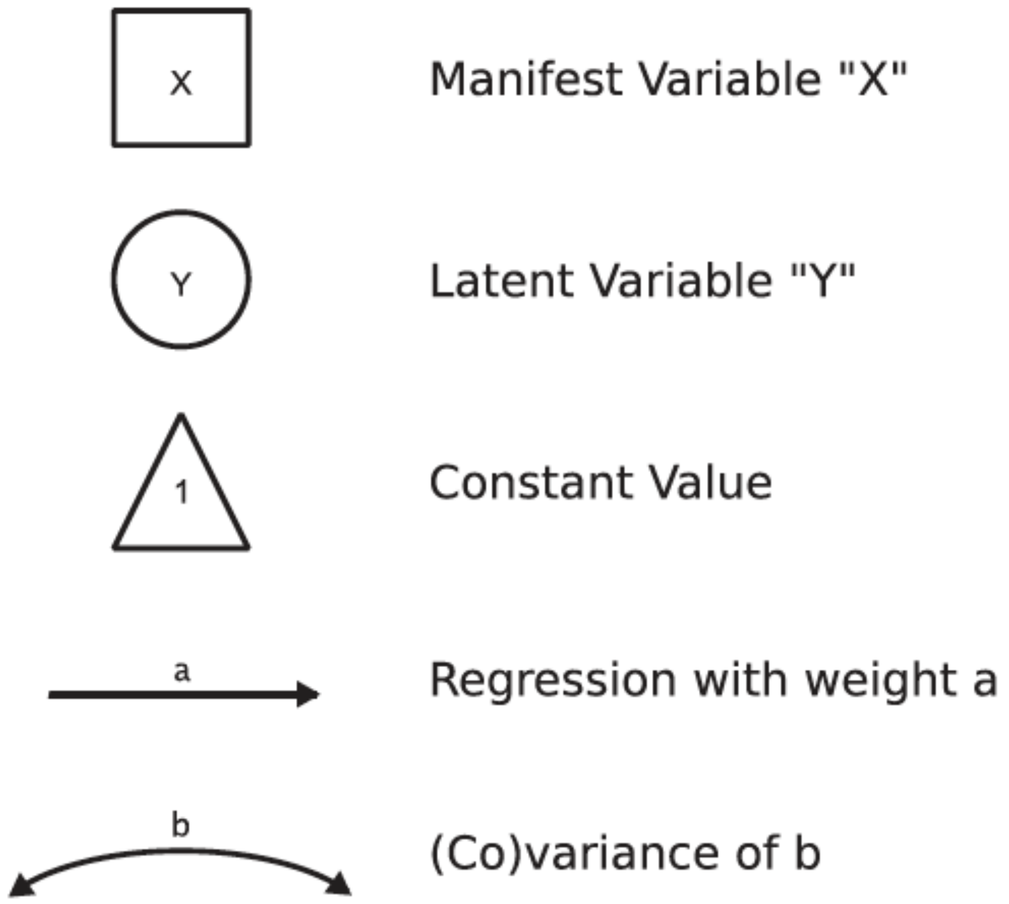
- A simple definition (1980s): any model of linear relationships between normally-distributed variables
- Modern SEM has various extensions to other types of distributions and non-linear relationships
- SEM generalize many general linear modeling techniques: t test, F test, regression, (repeated measures) ANOVA, mixed-effect models, mediation models, path analysis, growth curves

What Is SEM?

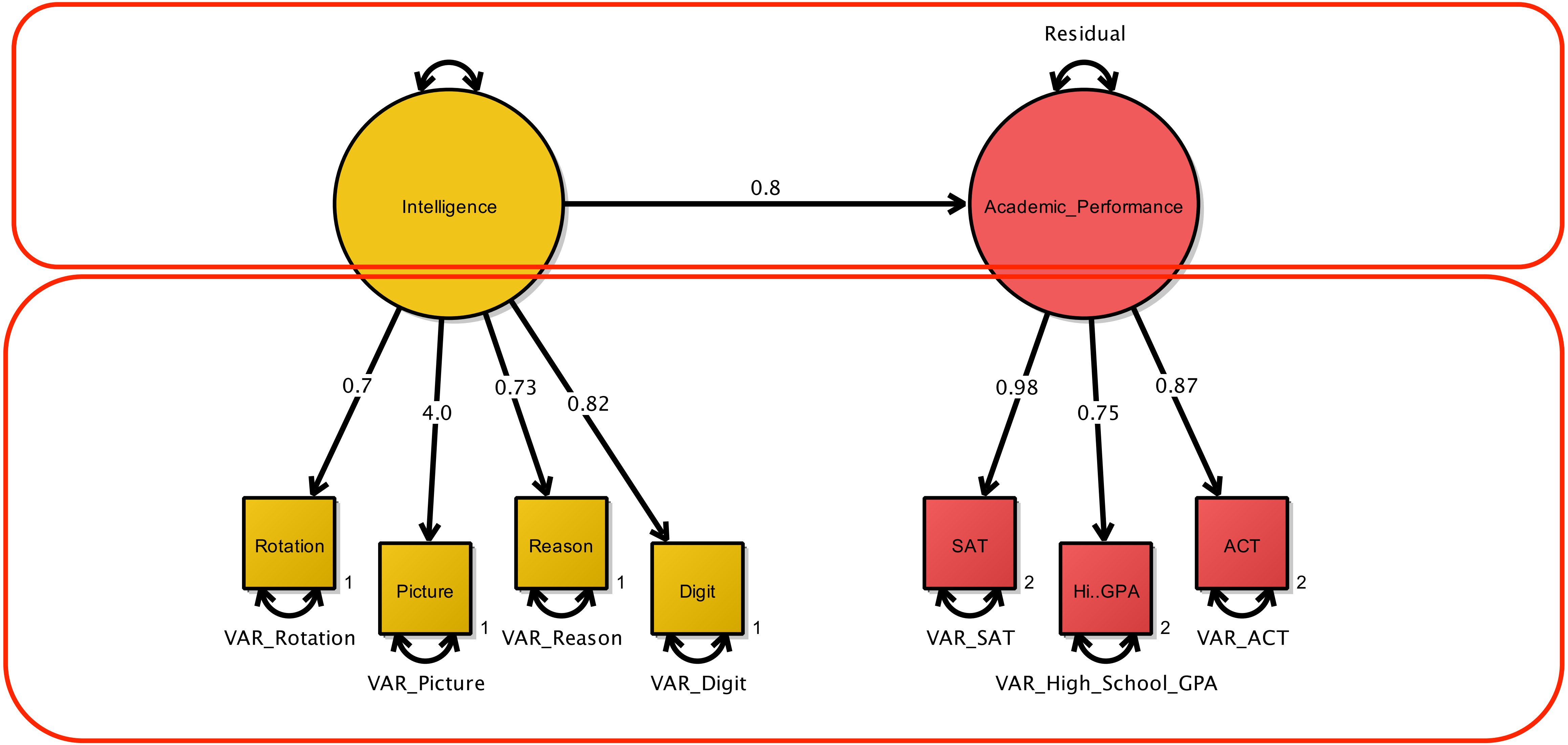
- A combination of two tools
  - Path analysis ("Structural Model")
    - Simultaneously estimating multiple pathways
  - Latent variable analysis ("Measurement Model")
    - Relating measured variables to hypothesized constructs

Path Diagrams

- Every SEM can be represented as a graphical model
- 1-to-1 mapping between graphical model and underlying mathematics

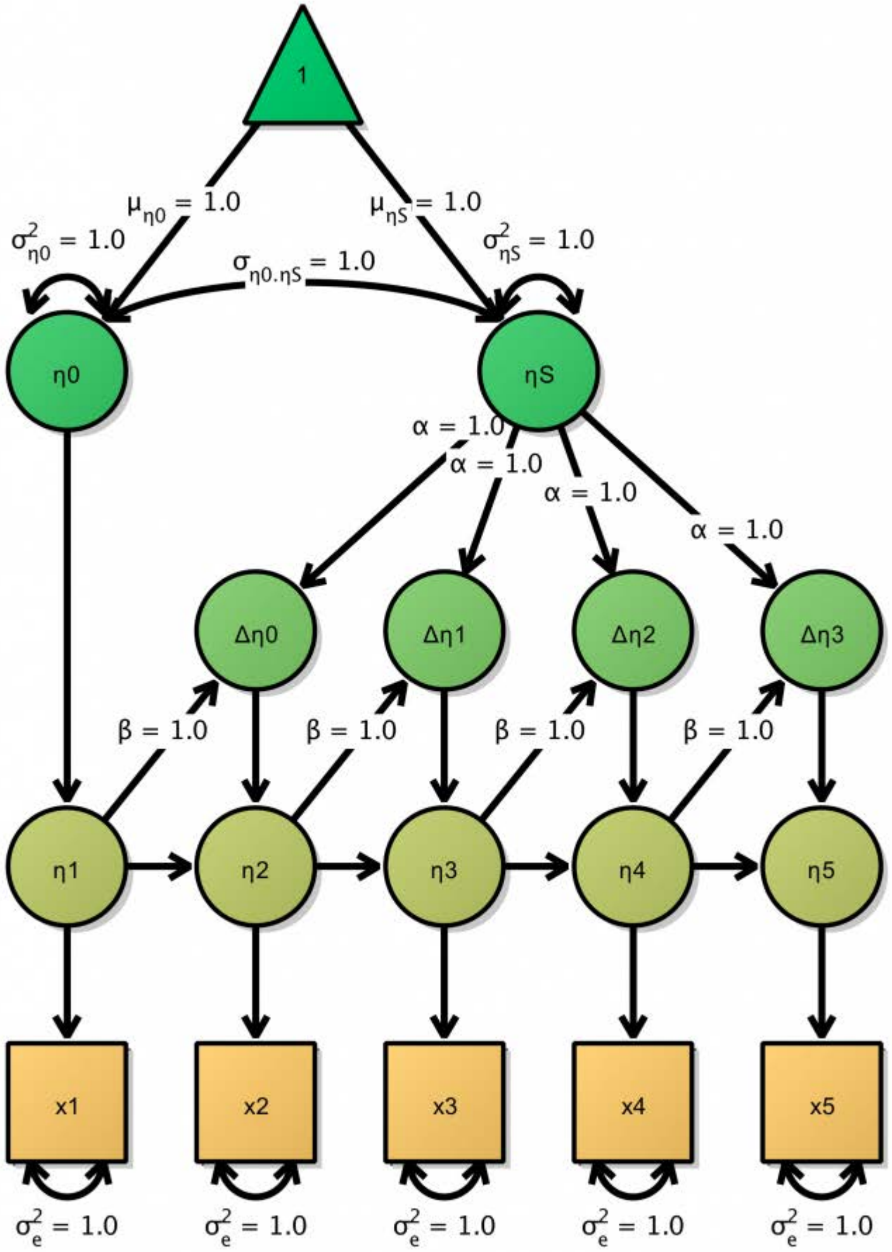


SEM = Measurement + Structure



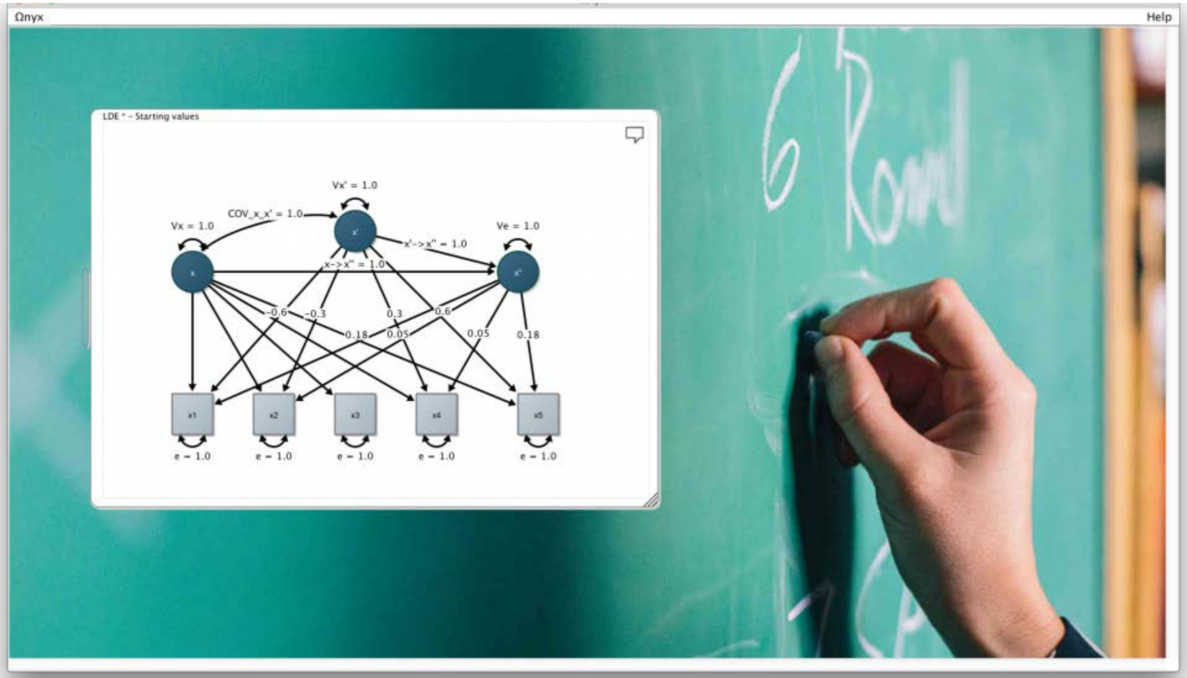
What Can SEM Do For You?

- **Saves a lot of trouble:** A universal language (to formalize and test your natural language hypotheses)
- More **valid, reliable,** and **sparse** models (moving your hypotheses from an item to a construct level)
- **Generative** models: great for simulations!
- **Pretty diagrams** (A one-to-one mapping between formal languages of SEM (matrix algebra, sets of equations, computer programs) to diagrams)

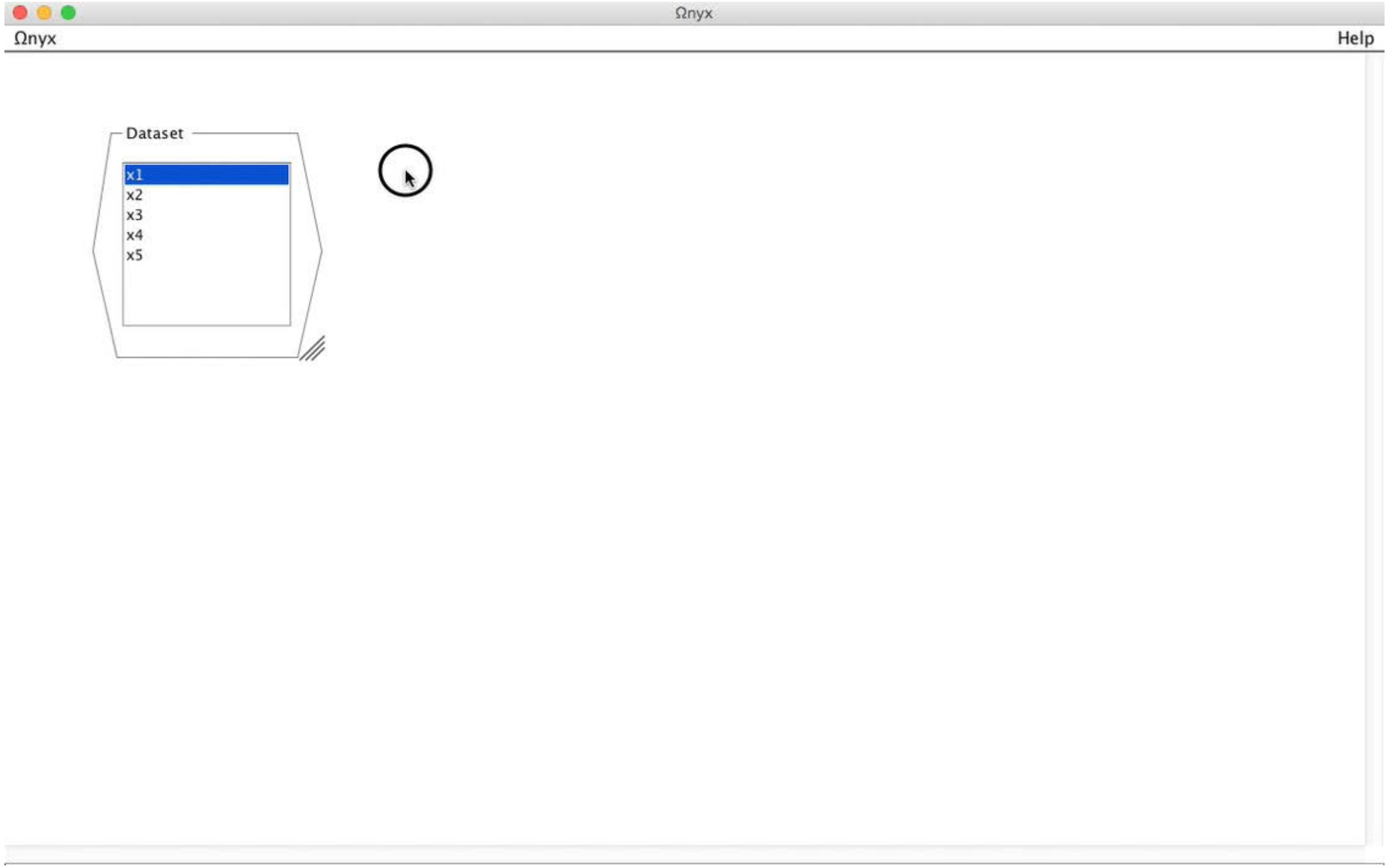


Commercial Break: Where To Start?

- A **graphical** interface for Structural Equation Modeling
- Free (as in beer) but not yet open source
- Platform independent
- Maximum Likelihood Estimator
- Import/Export to other formats (including specifications in other programs, such as lavaan)
- Developed at the MPIB, UniBW (and formerly UvA)



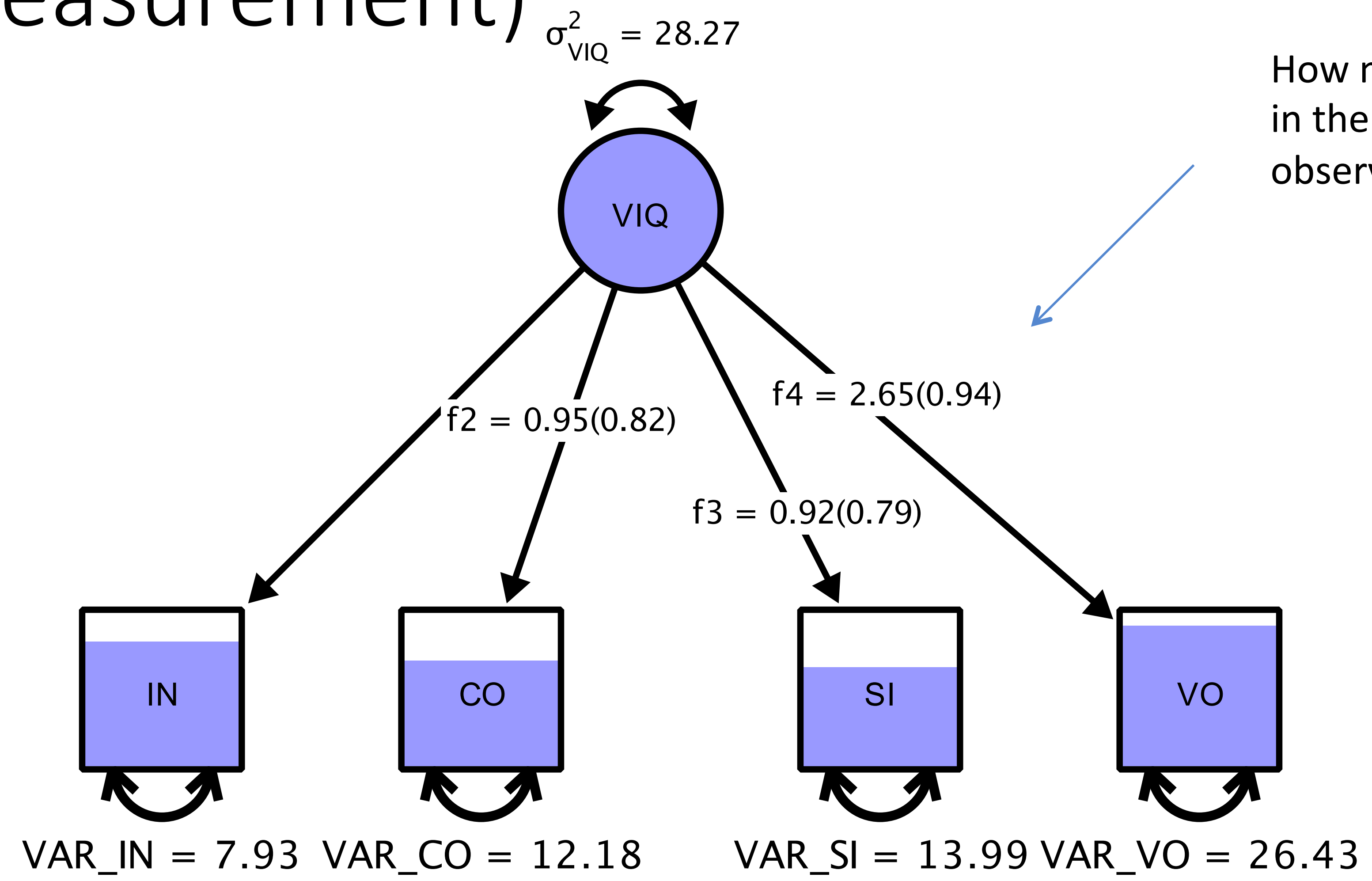
Ωnyx in Action



Reflective Latent Variables

- We rarely care about the actual variables we measure
  - ‘how many items can you recall’
- Rather, we care about hypothesized constructs
  - ‘Memory capacity’
- Observed scores *reflect* the latent variable
- Arrows from latent variable to a set of observed variables
- Spearman (1904): Scores on wide range of ability tests reflect underlying ability (‘general intelligence’)
- Other examples
  - Personality
  - Working memory
  - Attitudes

Basic Building Block: Factor models (Measurement)

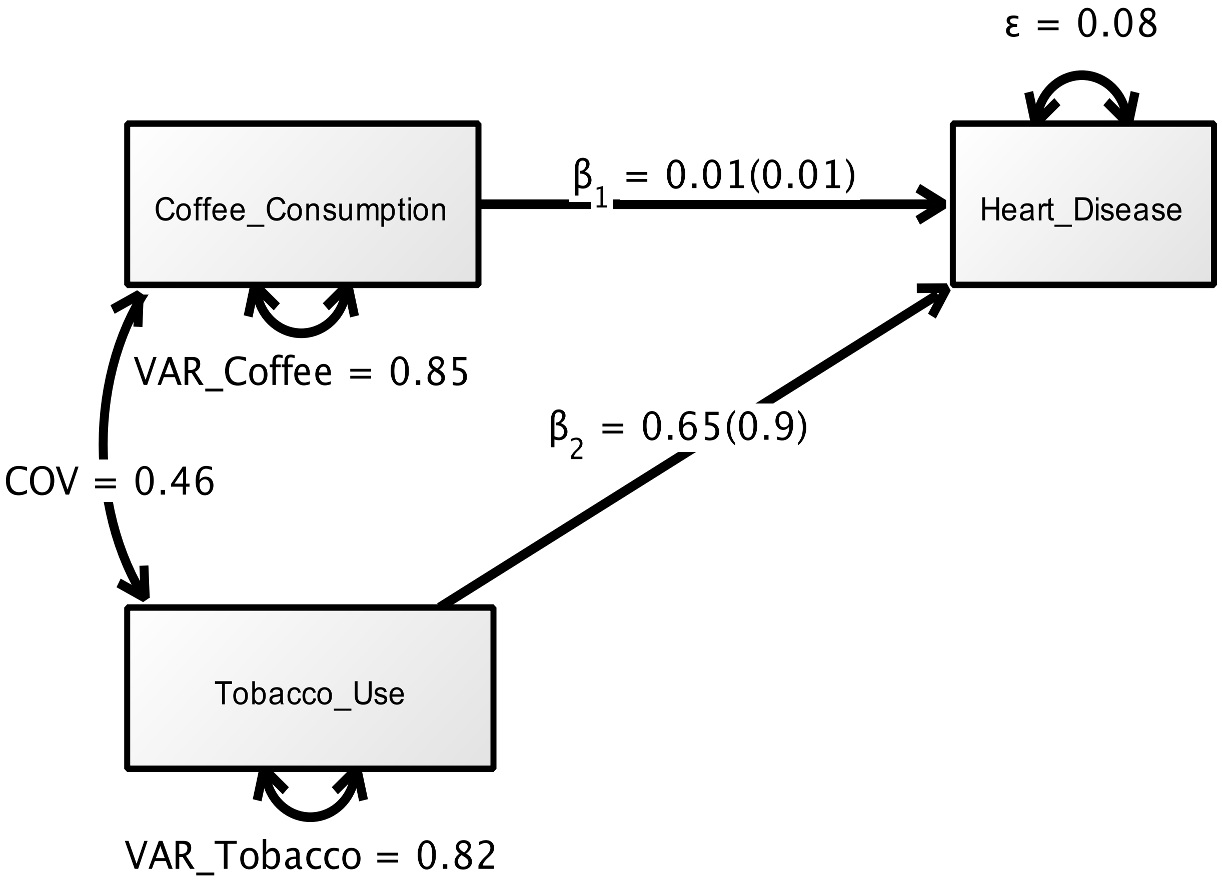


How much does one unit change in the latent factor change the observed variable?

Unexplained variance

Verbal IQ from WAIS: Information; Vocabulary; Comprehension; Similarities

Basic Building Block: Regression (Structure)



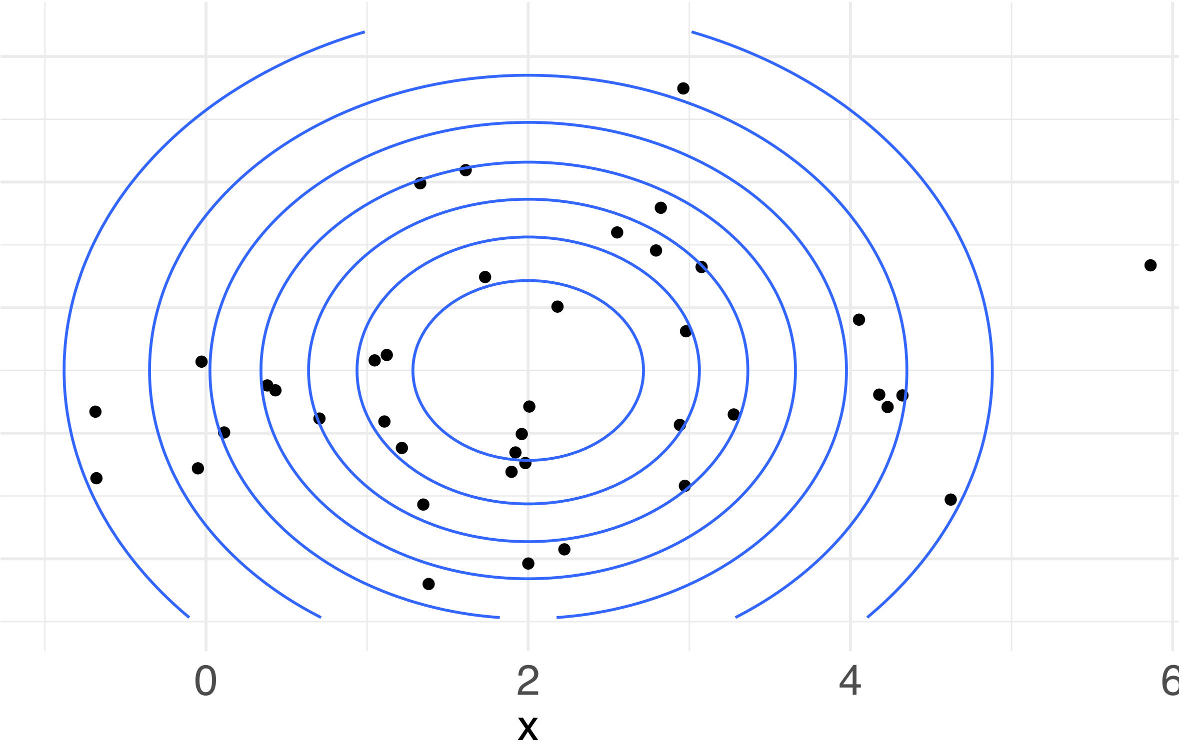
Generic Modeling Approach (see Quentin's talk)

- Propose a model that is *simpler* than the original dataset (i.e. fewer parameters than unique pieces of input)
- Estimate the model parameters
- Examine model fit (difference between observed and proposed covariance)
  - Provisionally accept the model
  - Reject the model
  - Refine the model (exploration; not confirmation)
  - Compare it to competing models
- Examine/Test model parameters

Evaluating Alternative Models in SEM

Fit indices:

- **χ^2**: Misfit against saturated model with test of perfect fit (but overpowered)
- **CFI**: Misfit per *df* in comparison to a null model (independence model)
- **RMSEA**: Misfit per *df and N* in comparison to a saturated model
- ***Likelihood-Ratio Test*** *(if models are nested) or information-criteria like AIC/BIC*



Crossectional Model of Brain-Cognition Relations

Application: Frontal Lobe Structure, Fluid Intelligence and Multi-Tasking

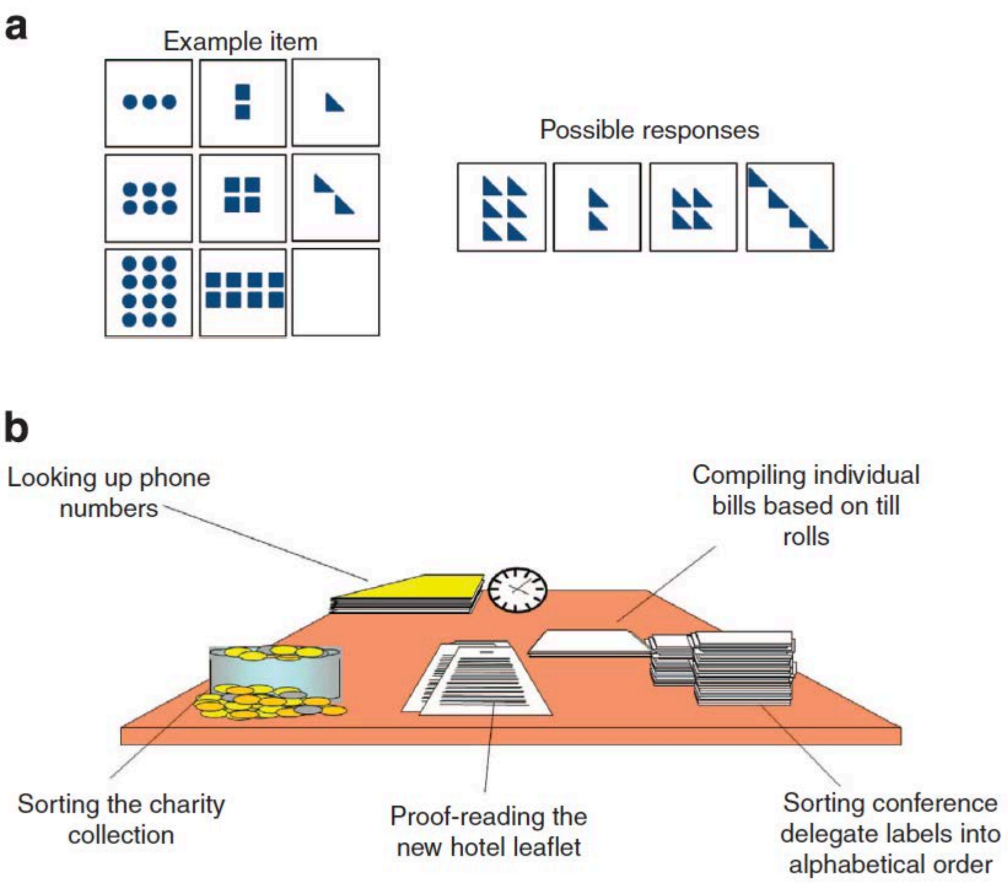
- How do frontal **gray matter** and **white matter** connecting those relate to two **executive functions**: fluid intelligence and multitasking.
- **Cam-CAN** data (N=567; mean age 54, range 18.46–88.9)
- Measurement model: Are multitasking and fluid intelligence separable cognitive factors? Single factor of pre-frontal integrity?
- Structural model: To what extent do differences in brain properties explain age-related differences in cognition?

Kievit et al., 2014 (DOI: 10.1038/ncomms6658)

Measurement Model of Cognition

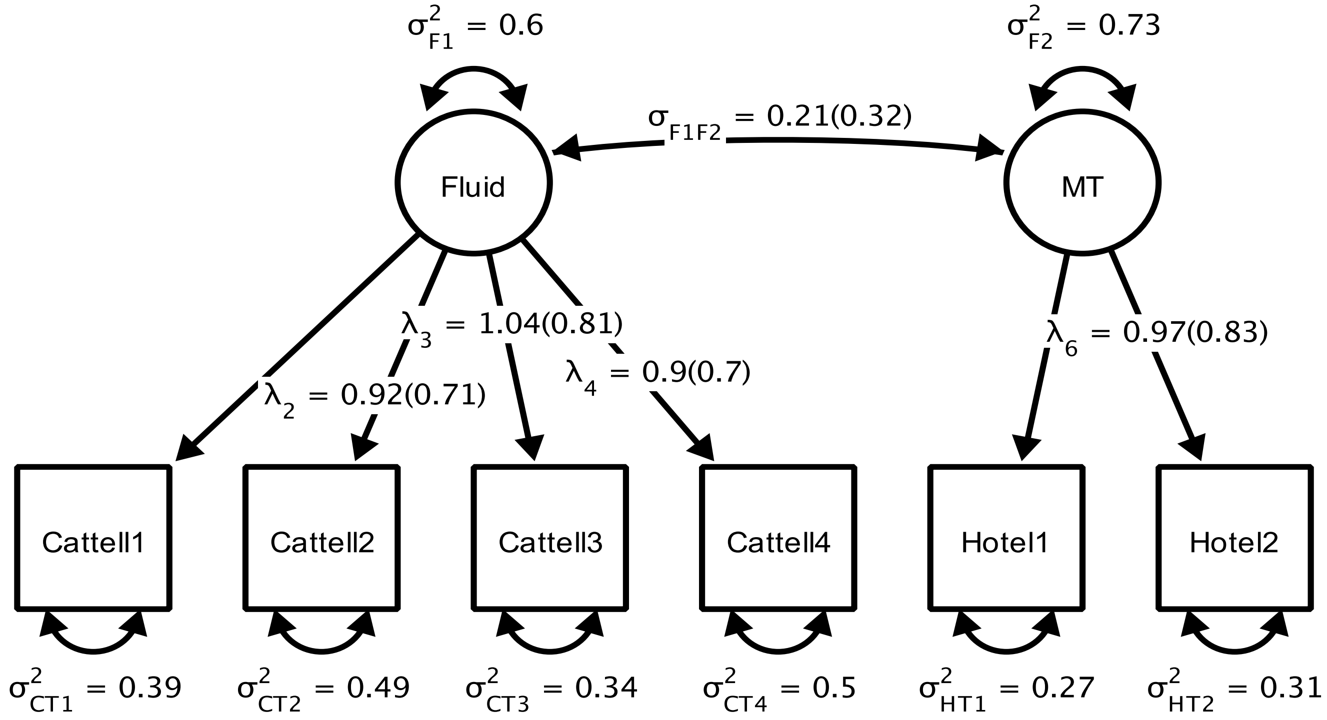
a) Cattell Culture-Fair pencil&paper test yields four scores on four subtests (series completions, odd-one-out, matrices and topology)

b) Simulated hotel work environment and measures the ability to distribute performance across multiple tasks (goal maintenance, task shifting, cognitive control) using number of tasks performed and time misallocated.



Kievit et al., 2014, Nat Commun

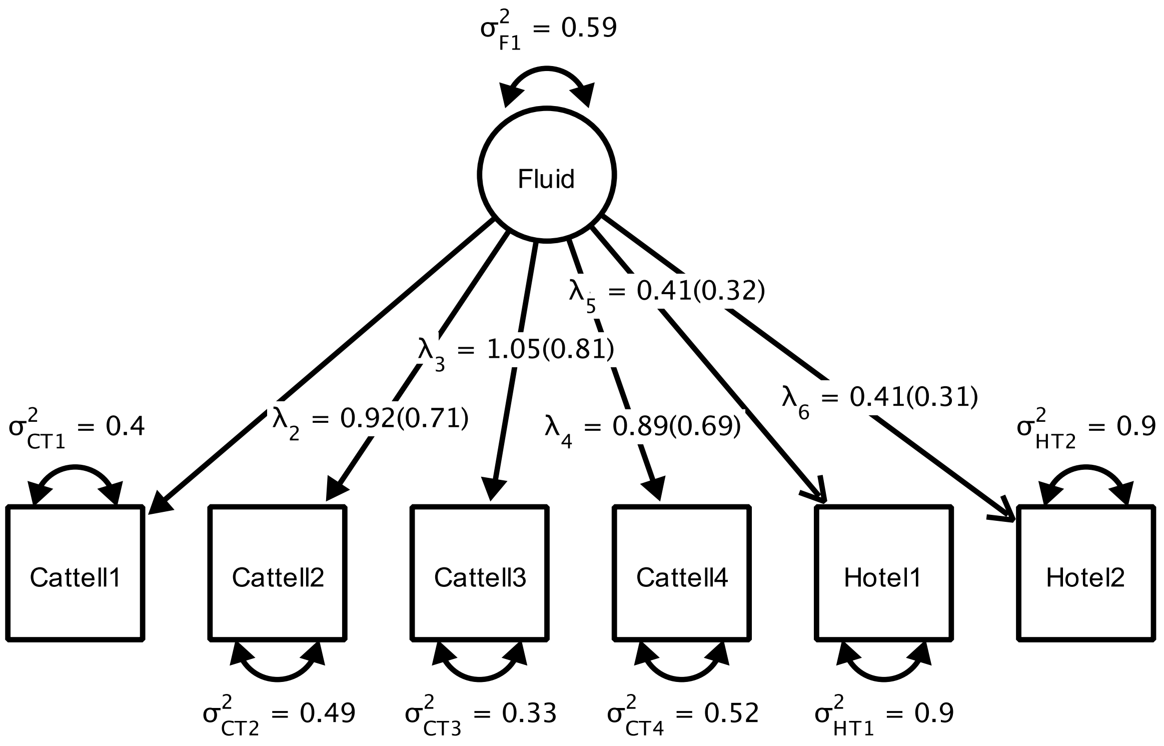
Separable Cognitive Factors



- ✔ RMSEA: 0.04
- ✔ CFI 0.99
- ✔ Item R^2 between 50% and 65%

Kievit et al., 2014, Nat Commun

Alternative Model of Unidimensional Cognition



- ✖ RMSEA: 0.27
- ✖ CFI 0.72
- ✖ Item R^2 down to 10%

Kievit et al., 2014, Nat Commun

Likelihood Ration Model Comparison

- Null hypothesis: There is no difference in fit
- Interpretation of significant result: The loss of fit by the more restricted, more parsimonious model (1 Factor) is significantly worse than what we expect by chance so that we can reject it

Here: $\chi^2(1)=355.01$, $p=0.00000...$

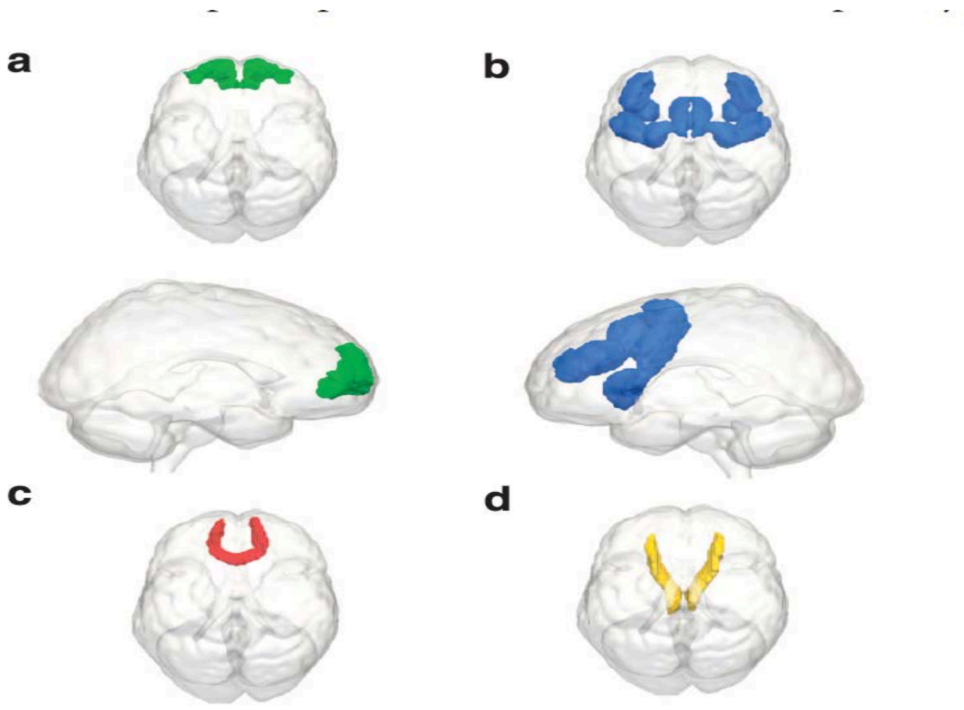
Indicators of Brain Integrity

a) GMV in frontopolar BA10

b) GMV within the frontal section of the multiple demand system

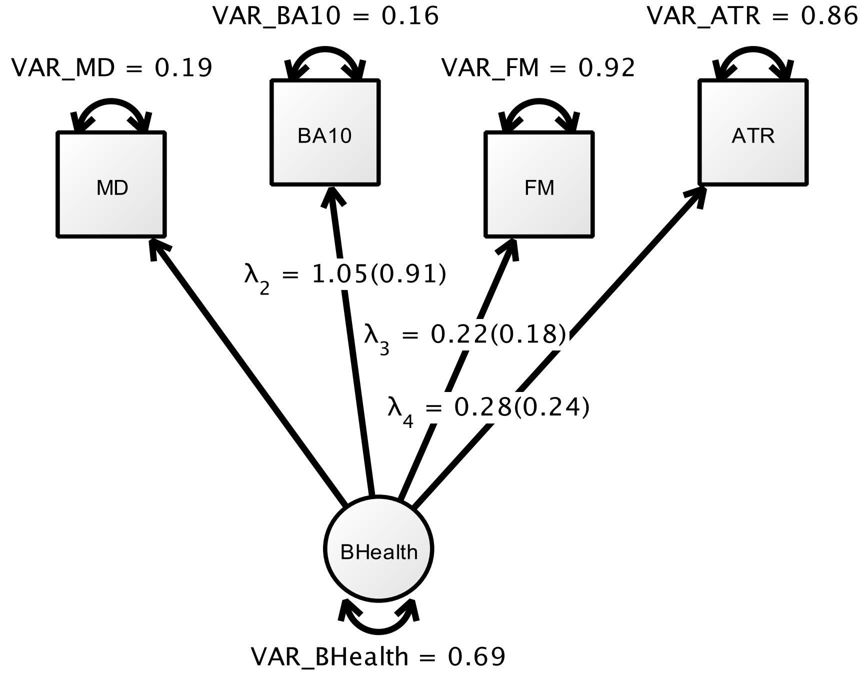
c) WM Forceps Minor (connecting left and right BA10)

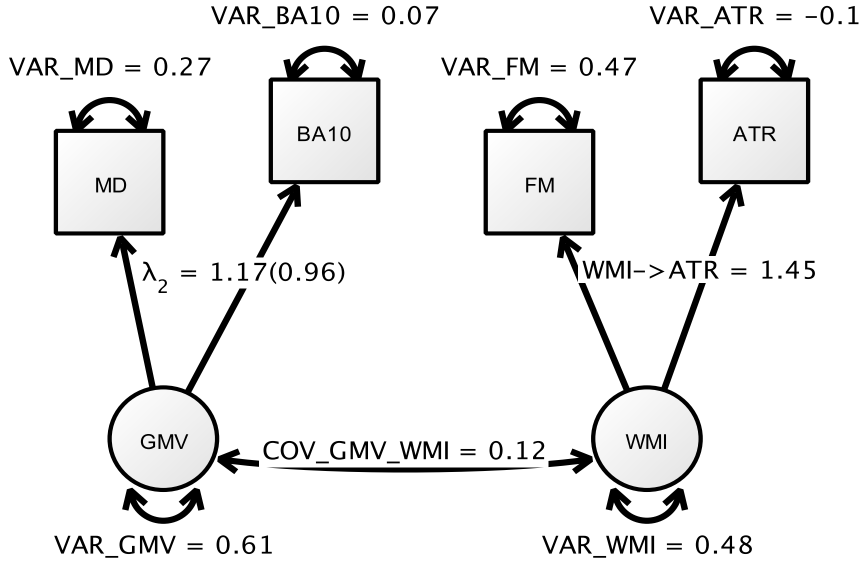
d) WM Anterior Thamalic Radiations (considerable connectivity with frontal MD system)



Kievit et al., 2014, Nat Commun

Measurement Model of the Brain





„General PFC“

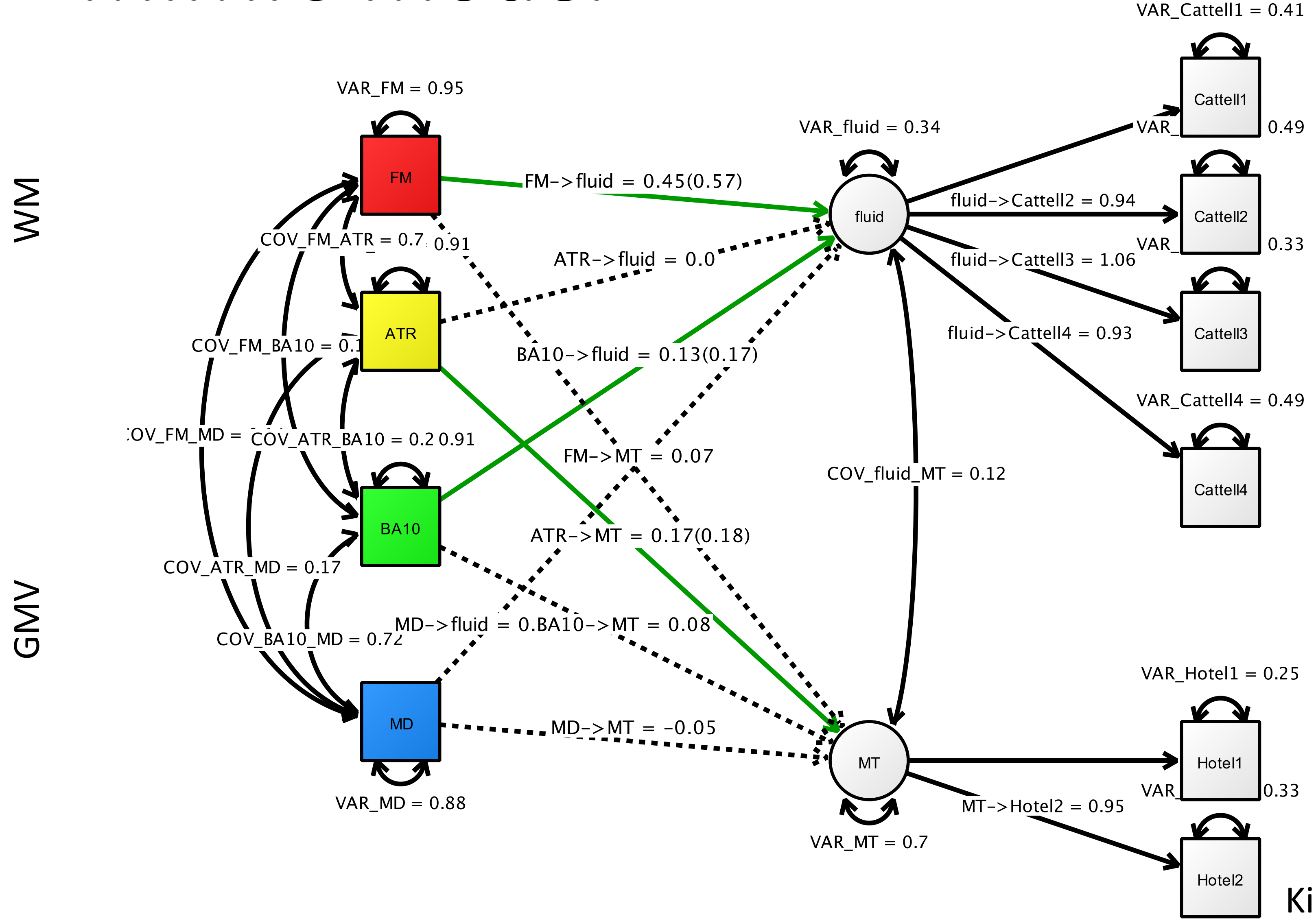
- RMSEA 0.63 ✖
- CFI 0.58 ✖
- R^2 down to 3% ✖

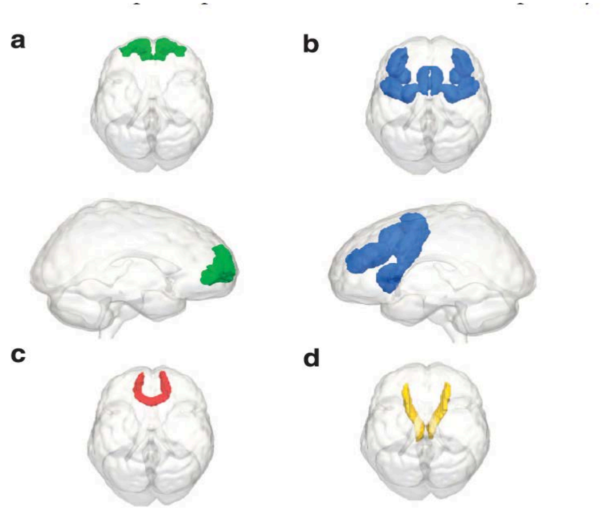
„WM and GM“

- RMSEA 0.05
- CFI 0.99
- But negative variance! ✖

Kievit et al., 2014, Nat Commun

MIMIC Model



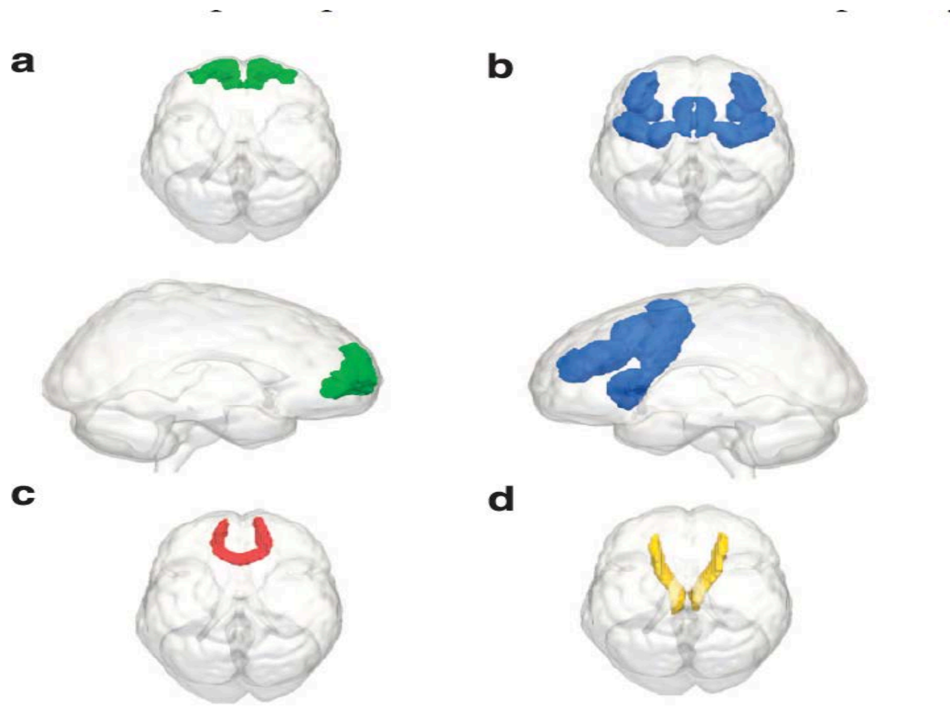


Kievit et al., 2014, Nat Commun

Results

Differential Predictions of the GM and WM variables-of-interest:

- BA10 and FM jointly predict (age-related) individual differences in fluid intelligence: $R^2=0.35$
- ATR predicts multitasking, even though with very small effect size, $R^2=0.03$



A multiregion-multimethod model of brain structure

In search of a biologically plausible model from multimodal imaging data

"Brain Integrity"
Or
"Brain Health"

Magnetization transfer ratio

Mean diffusivity

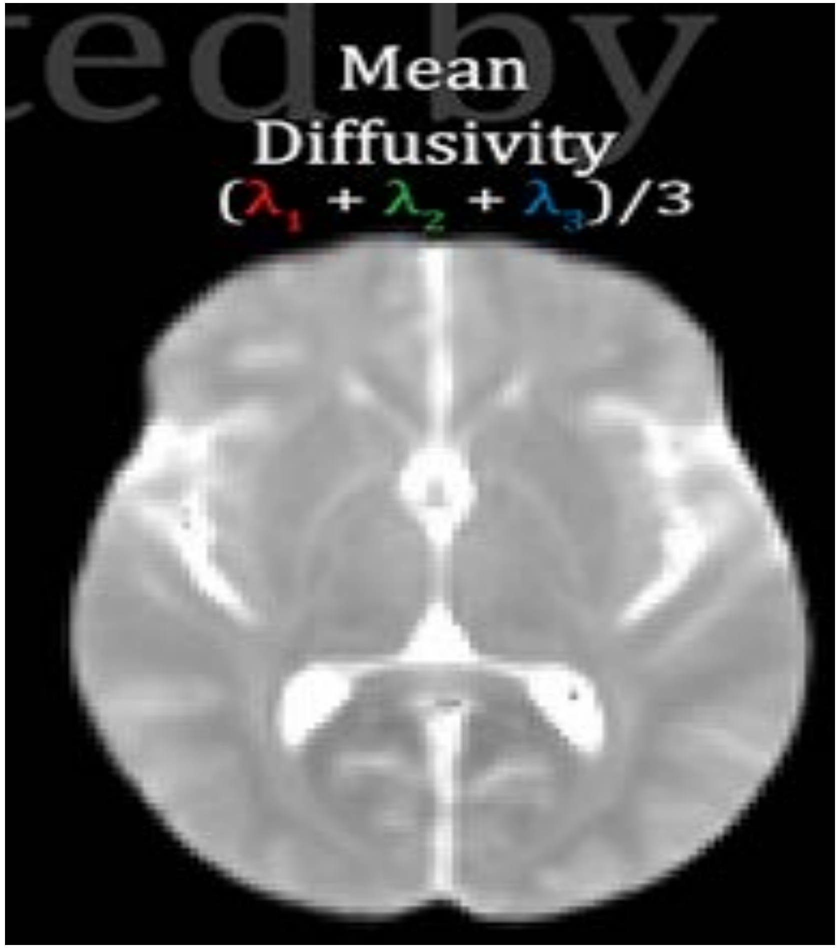
Grey matter probabilities

...

Köhncke, Brandmaier, et al. (in prep)

Mean diffusivity in grey matter

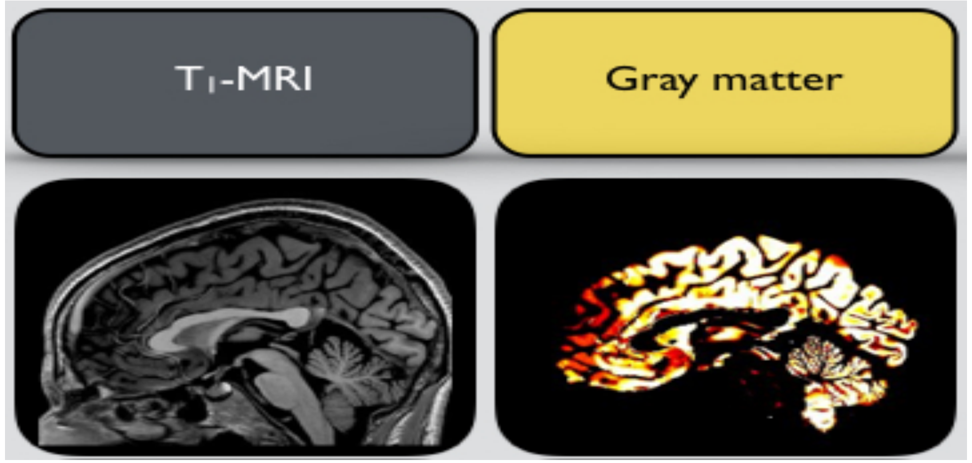
- Water will diffuse more rapidly in the direction aligned with the internal structure
- Diffusion-tensor imaging
  - Uses pulsed gradients to excite water protons along directions
  - Uses vector/tensor math to get parameters for each voxel: rate and direction (3D) of diffusion
- Mean Diffusivity (MD) is defined as the mean of the diffusion tensor eigenvalues)
- MD = rotationally invariant magnitude of water *diffusion* within brain tissue;
- Interpretation: Higher values ~ less "dense" structure



Grey matter probability

- Volume-based morphometry (VBM)
  - T1 images -> spatial registration to a reference brain
  - Tissue classification (GM/WM/CSF)
- Interpretation: Higher values ~ “larger” structure

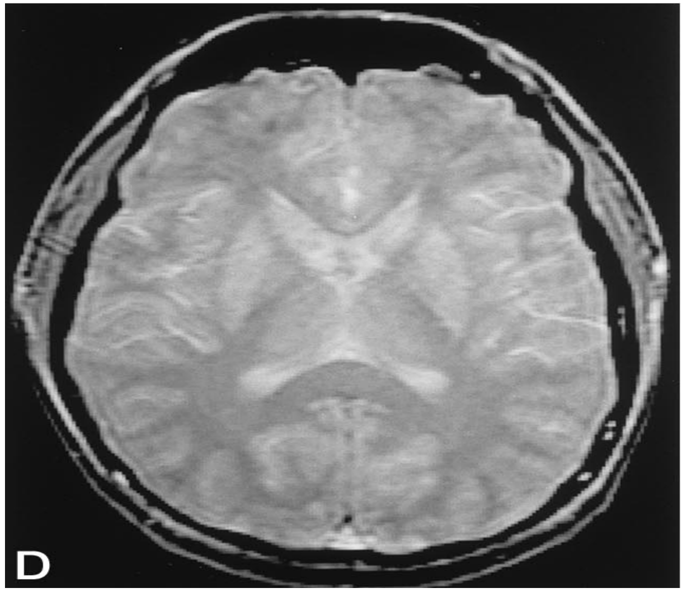
- GMV captures volume shrinkage, pathological changes on a coarse level
- GMV changes previously linked to cognition (Becker et al. 2015; Cabeza et al., 2008)



http://www.neuro.uni-jena.de/cat

Magnetization transfer in grey matter

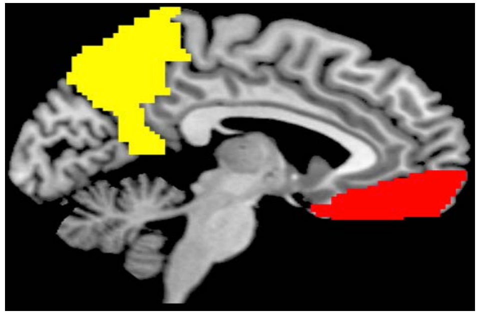
- Magnetization transfer (MT)
  - MT imaging uses specially designed MT pulse which capitalises on energy exchange between excited molecules and on differences in relaxation time between water bound to macromolecules and free water *[*
  - After MT pulse, free water is partly saturated
- MT Ratio:
  - ratio between image without MT pulse and image with MT pulse
  - depends on water content, concentration & chemistry of macromolecules
- Interpretation: Higher ratio ~ “denser” tissue

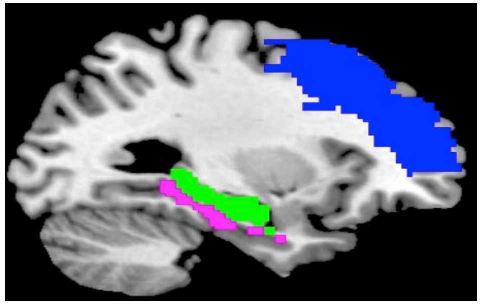


MT saturation (Ge et al., 2002)

Focus on Episodic Memory

- ROI-approach: average across voxels in a region
  - Selection of ROIs based on their relevance for episodic memory
    - Functional networks
      - see e.g. Geib et al. 2017; Benoit et al., 2015
    - Volumetric ROIs
      - e.g. Becker et al. 2015; Cabeza et al., 2008
  - MT&MD and cognitive ability
    - MT as structural indicator in relation to cognition
      - e.g. Schmidt et al., 2014, Düzel et al. 2008, 2010, Eckert et al., 2004; Tambasco et al., 2011
    - MD as structural indicator in relation to cognition
      - e.g. Bhagat and Beaulieu 2004; Ni et al. 2010; Pfefferbaum et al. 2010, Grydeland et al., 2015
  - ROIs (from AAL atlas): medio-orbitofrontal PFC, precuneus, parahippocampal gyrus, hippocampus
  - Grey matter volume values adjusted for total intracranial volume

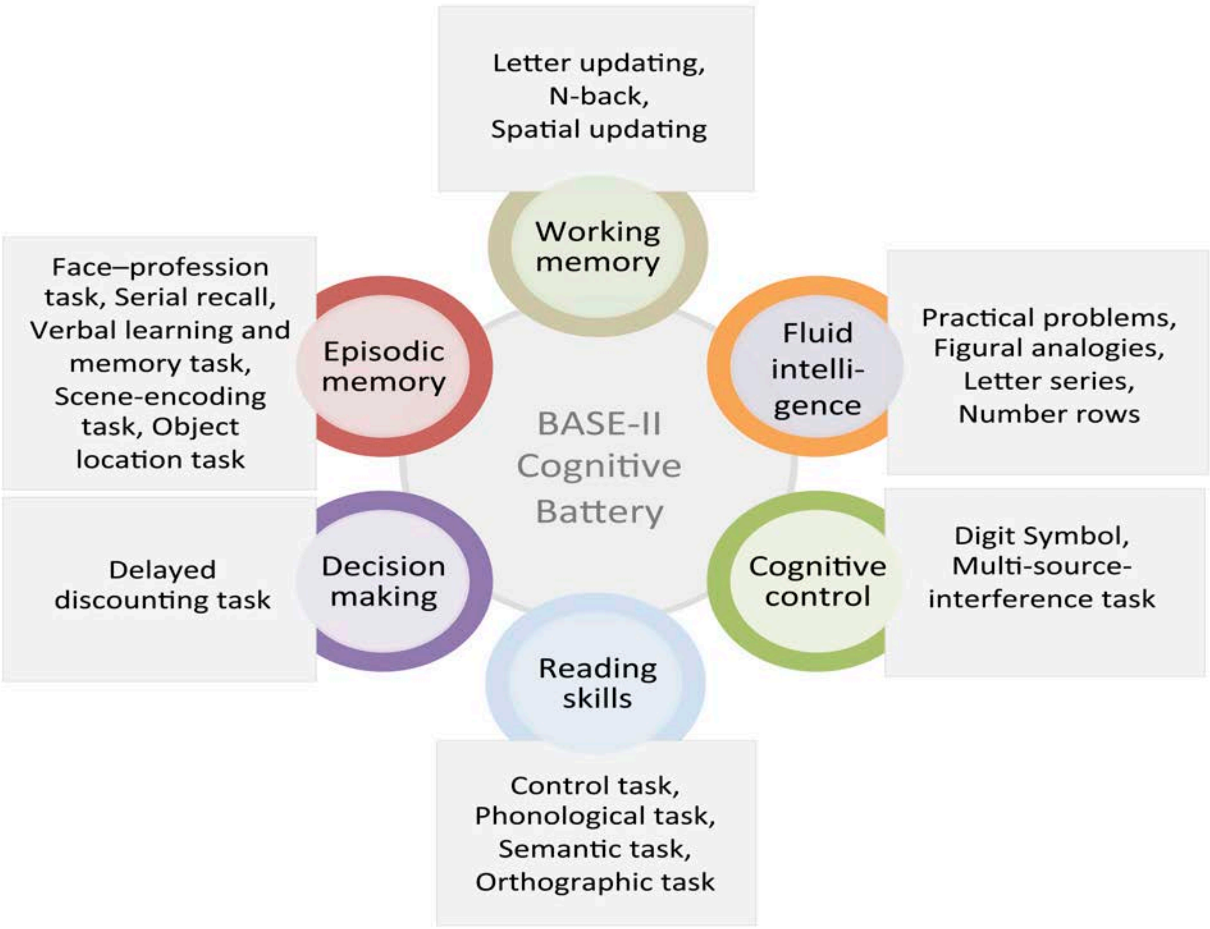


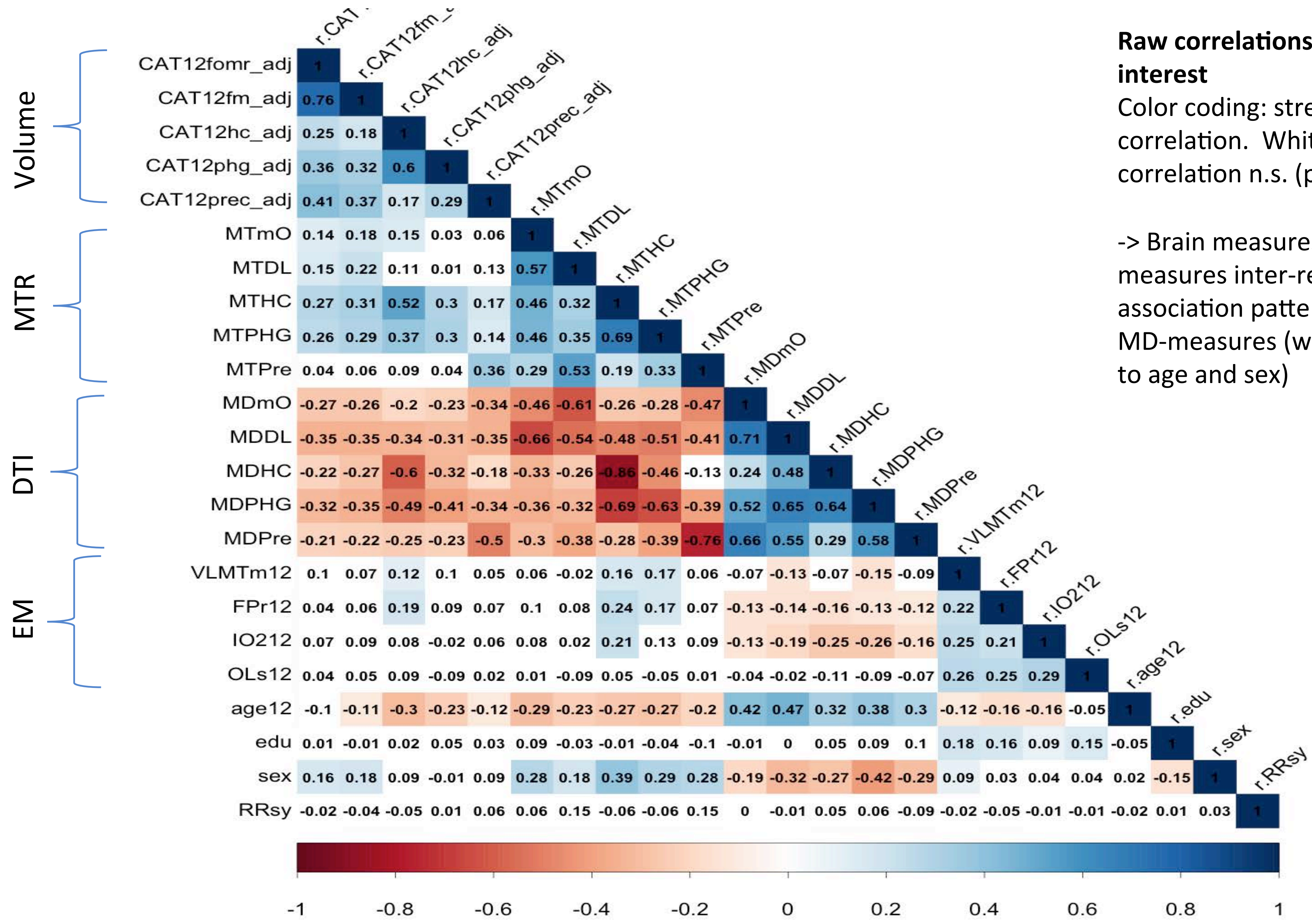


BASE-II

Berlin Aging Study II

- Comprehensive cognitive test battery
  - 1.600 older adults (60-80 years)
  - Effective sample: n=1532
- Imaging
  - 344 older adults
  - (172 cases have complete data across MR modalities and EM)





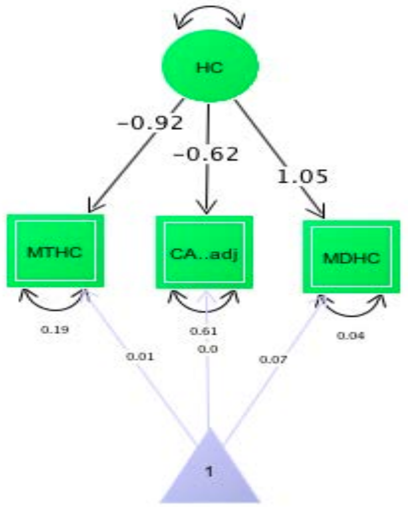
Raw correlations among all variables of interest

Color coding: strength and direction of correlation. White background: correlation n.s. (p>.05).

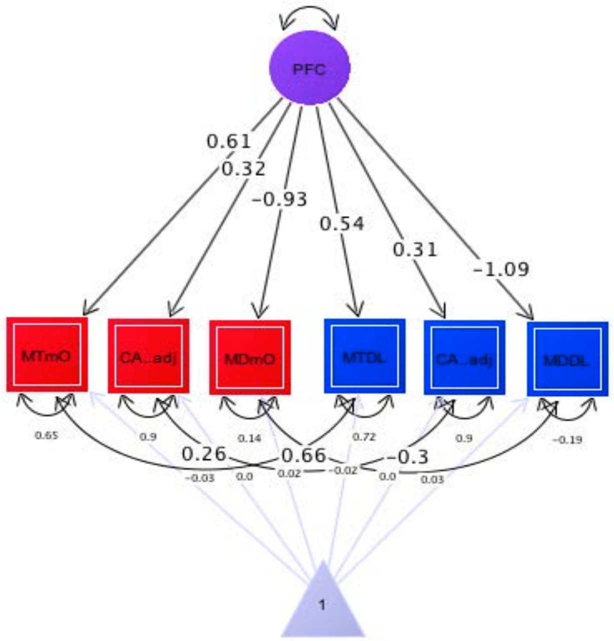
-> Brain measures inter-related, EM measures inter-related, but no strong association pattern brain-EM, except for MD-measures (which also strongly relate to age and sex)

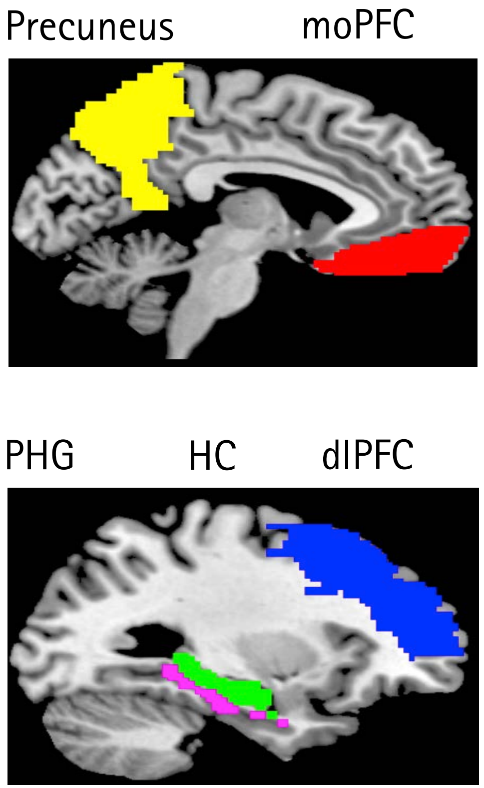
Region-wise measurement models

- Indicators: Mean across left&right hemi in each measure, **z-standardized**
- For example: Factor for HC and for PFC
  - moPFC and dlPFC were so highly correlated that they are better represented by one factor



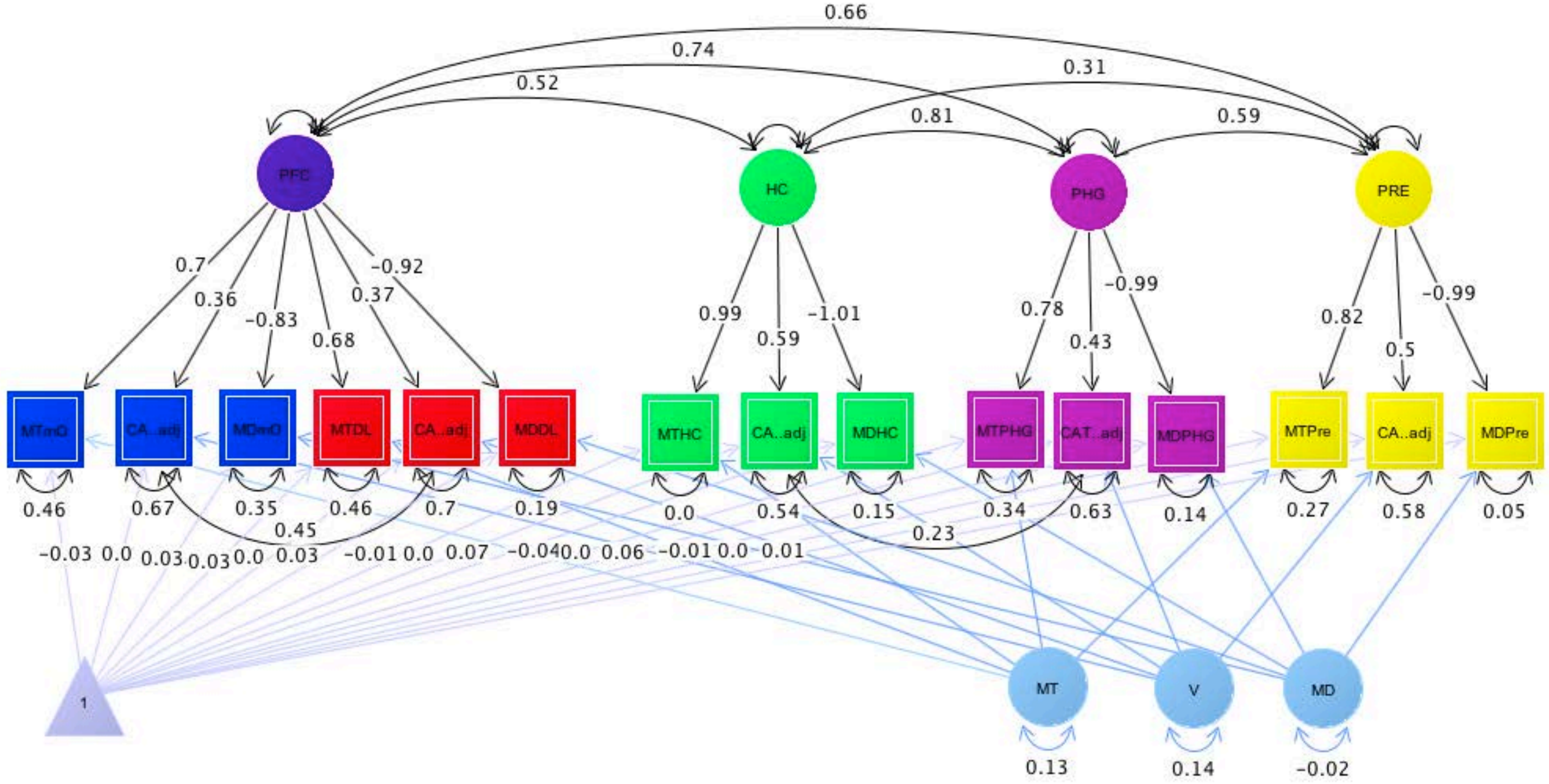
(Standardized estimates in parentheses)





…some loadings are rather low but we keep the items for the full model

Multiregion-multimethod model

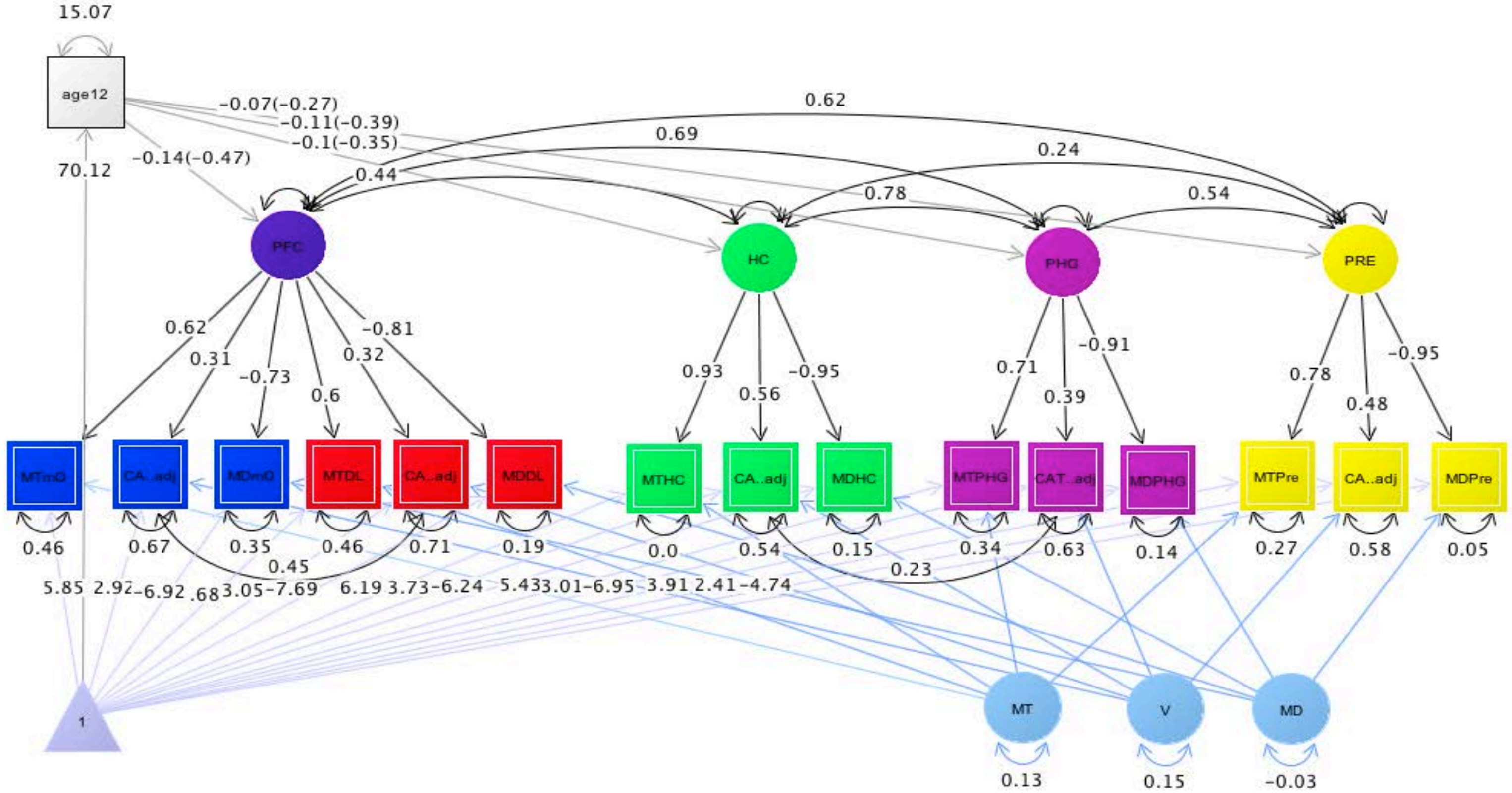


(RMSEA = .08 CFI = .93; SRMR = .07)

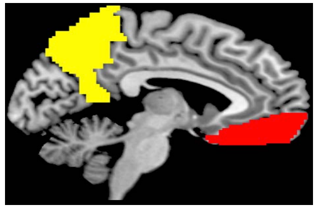
Multiregion-multimethod model

Age differences in all latent regional factors significant

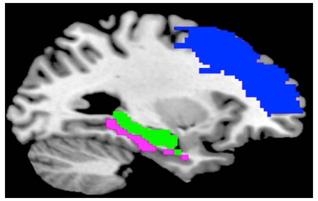
(Standardized estimates in parentheses)



Precuneus moPFC



PHG HC dlPFC



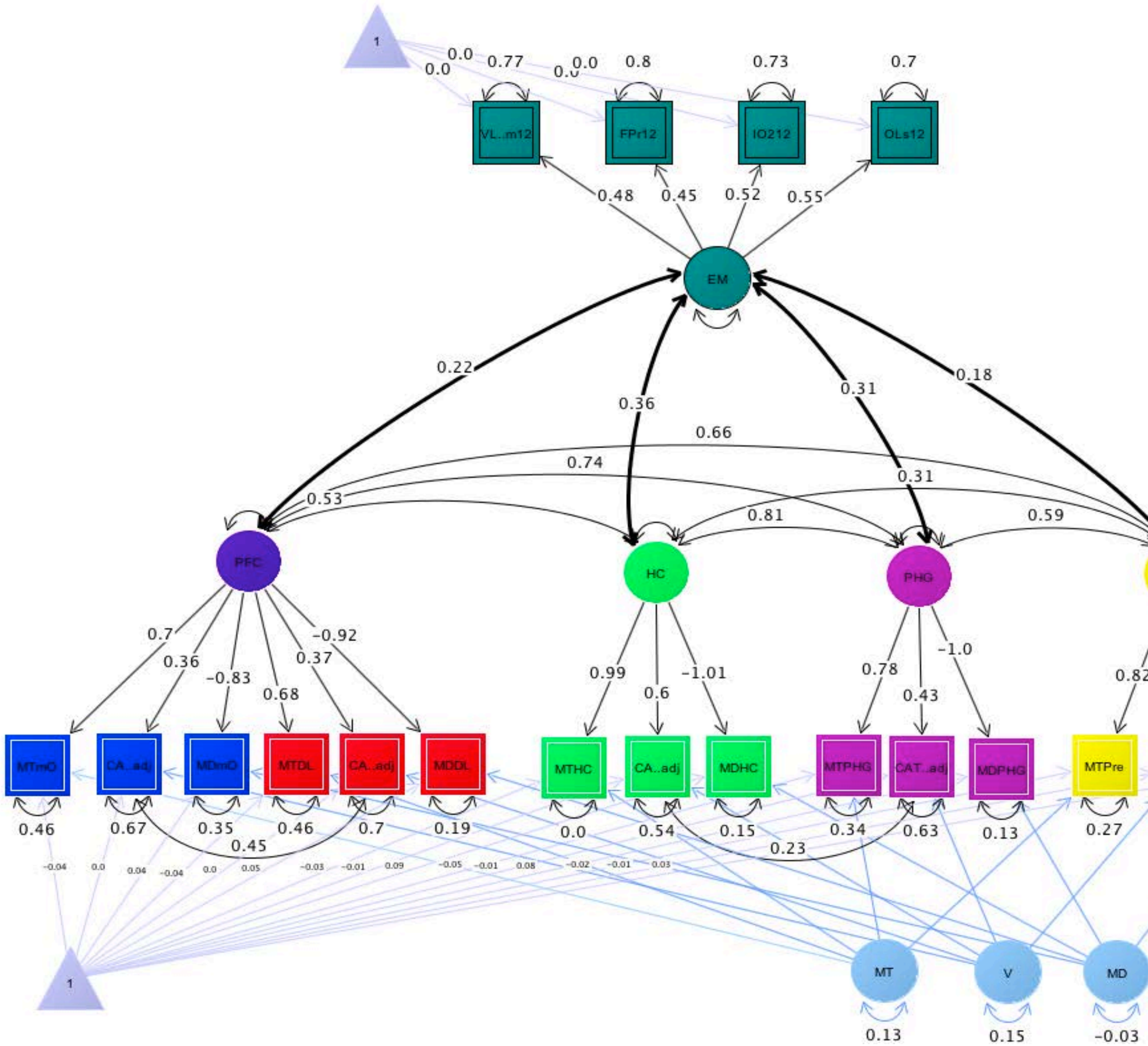
Associations with Episodic Memory

Episodic Memory in BASE-II:

- Auditory verbal learning. The sum of items recalled over five trials
- Face profession task assessing associative binding on the basis of recognition of incidentally encoded faceprofession pairs. Corrected hit rates for rearranged face-profession pairs
- Object location memory task assessing object-location memory with 12 colored photographs arranged on a 6x6 grid. The sum of correct placements
- Scene encoding task assessing nthe ability of incidental scene encoding. Delayed recognition hit-rate.

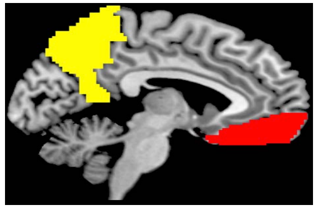
Region-wise associations with EM: all significant

EM with	
PFC	.22
HC	.36
PHG	.31
PRE	.18

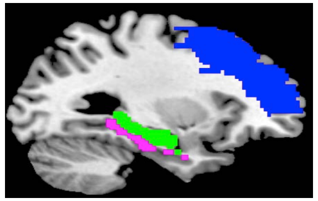


(RMSEA = .03 CFI = .93; SRMR = .07)

Precuneus moPFC

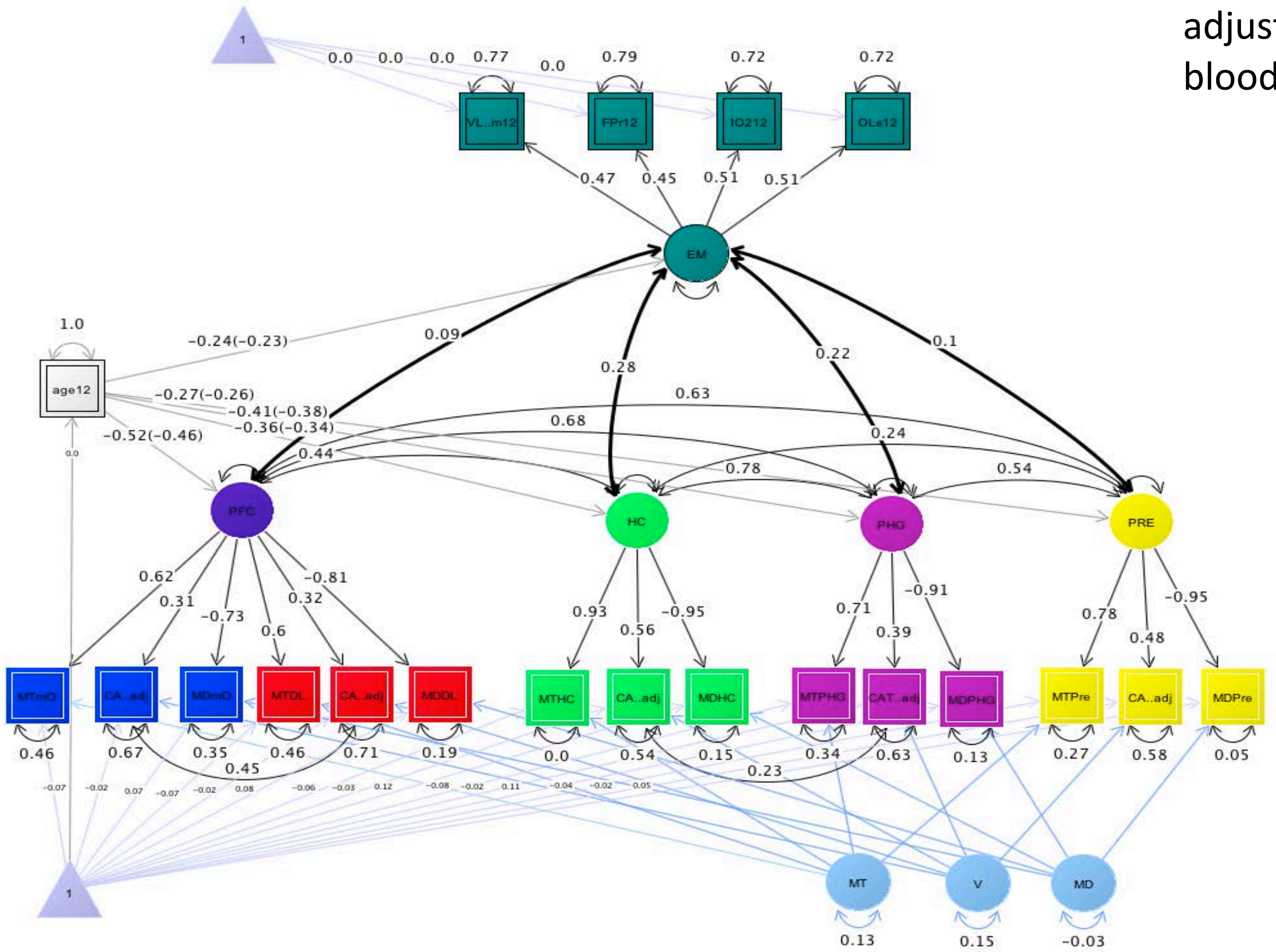


PHG HC dlPFC



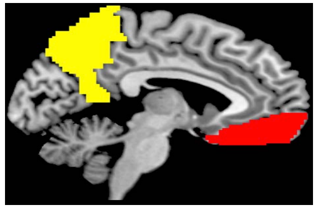
Age-adjusted: only EM<-> HC significant.

Covariances can be read as residual correlations accounting for age differences

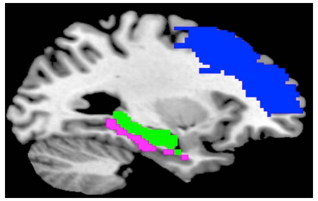


HC <-> EM is also robust to adjusting for education, sex, and blood pressure

Precuneus moPFC

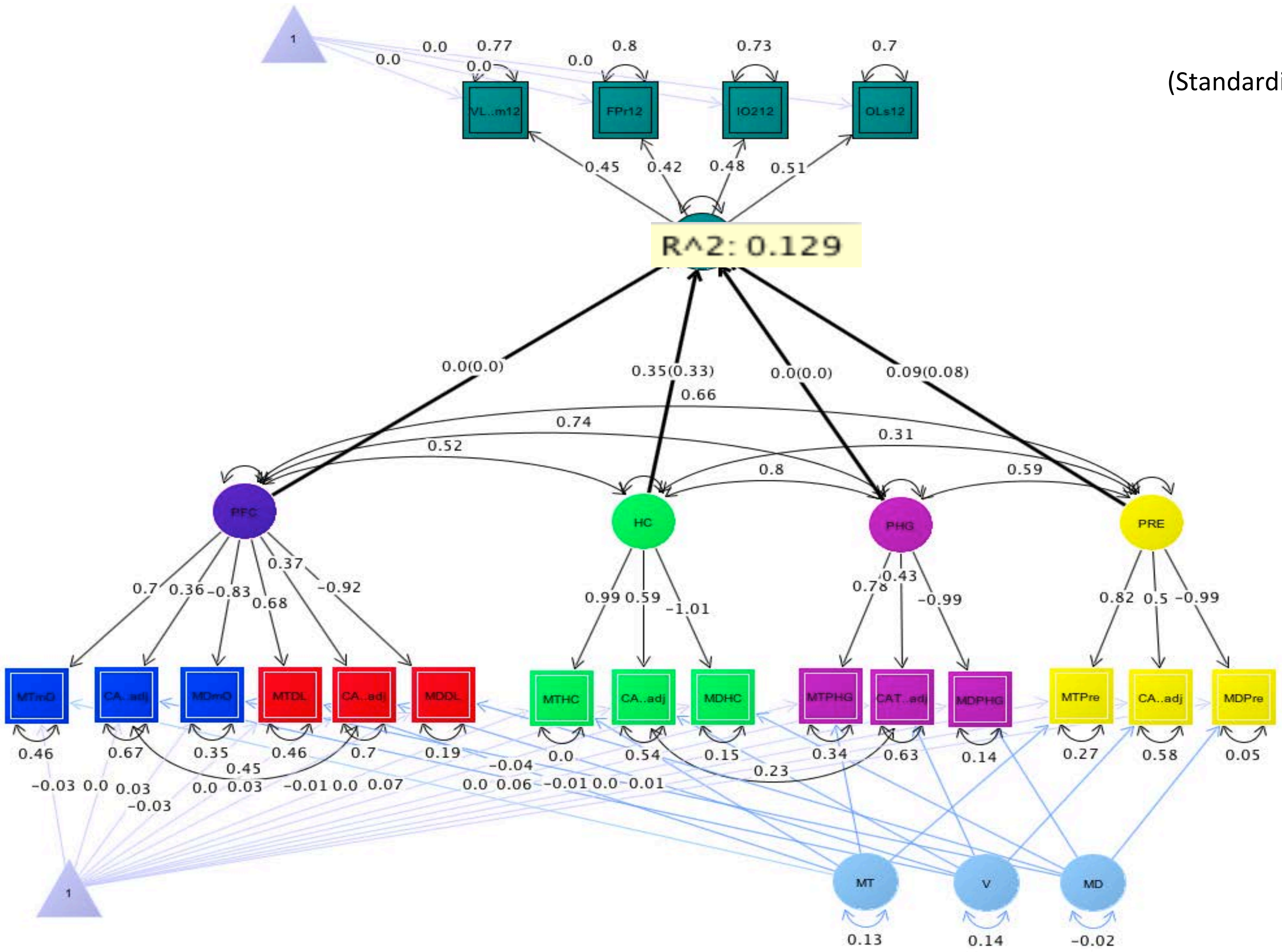


PHG HC dlPFC



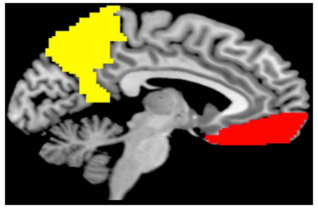
-> thus, the covariance of HC and EM is largely independent of age, education, sex, and blood pressure

Region-unique (MIMIC model): only hc->EM

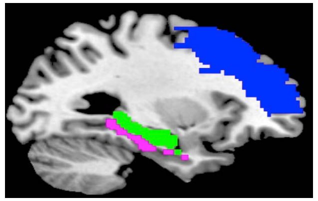


(Standardized estimates in parentheses)

Precuneus moPFC



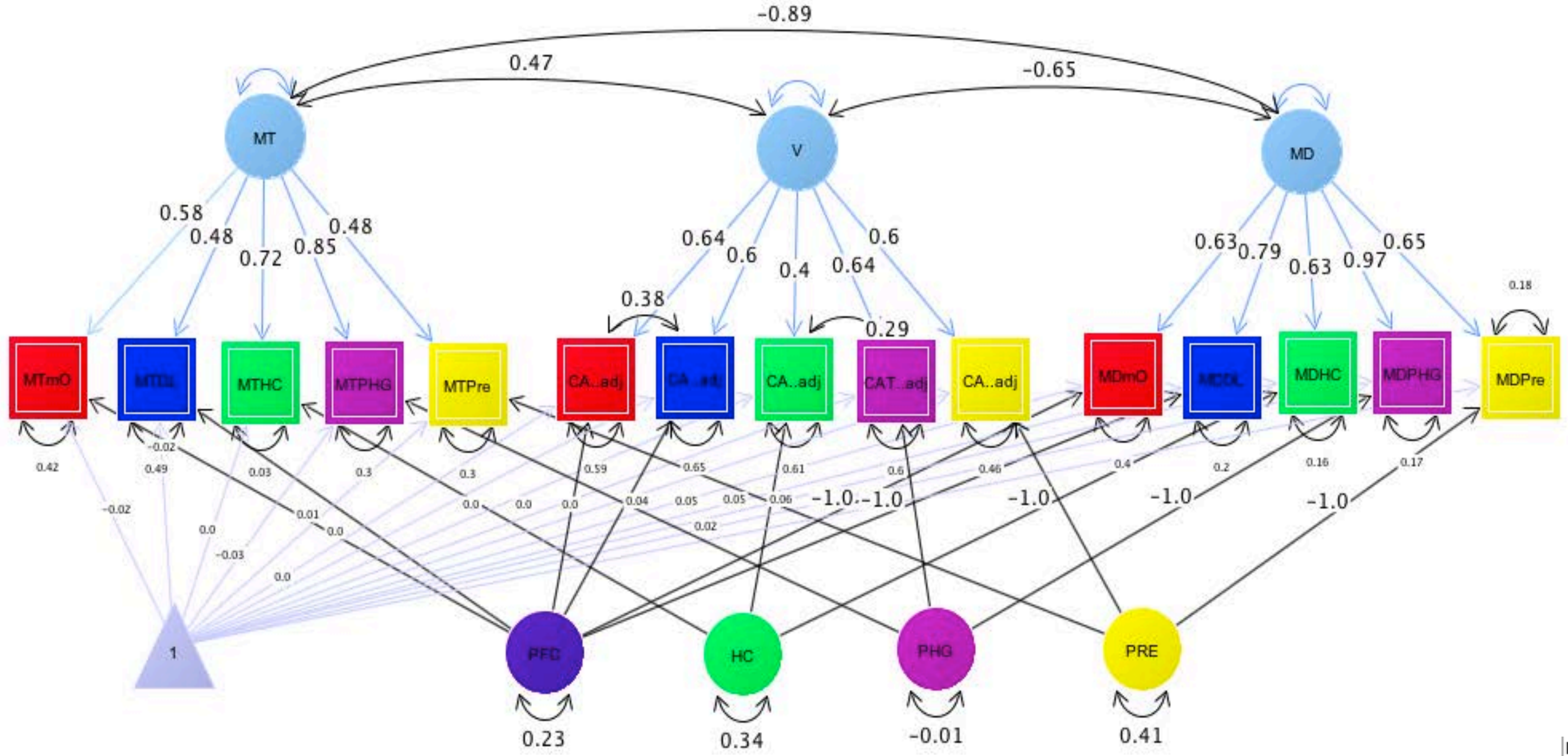
PHG HC dlPFC



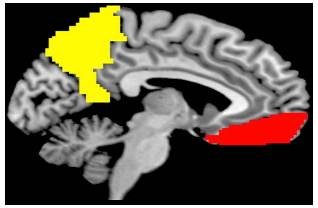
-> thus, considering overlapping variance with the other regions, HC contributes with unique association

Interested in method factors? Flip!

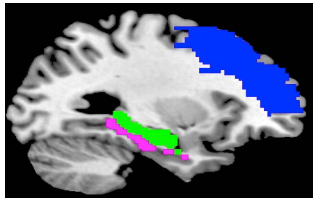
Covariances can be read as correlations because latent variances are @1
Loadings are standardized by z-transforming indicator variables



Precuneus moPFC



PHG HC dlPFC



RMSEA (df corrected)	: 0.128
RMSEA (Kulback Leibler)	: 0.141
RMSEA (classic)	: 0.113
SRMR (covariances only)	: 0.096
CFI (to independent model)	: 0.857
TLI (to independent model)	: 0.815

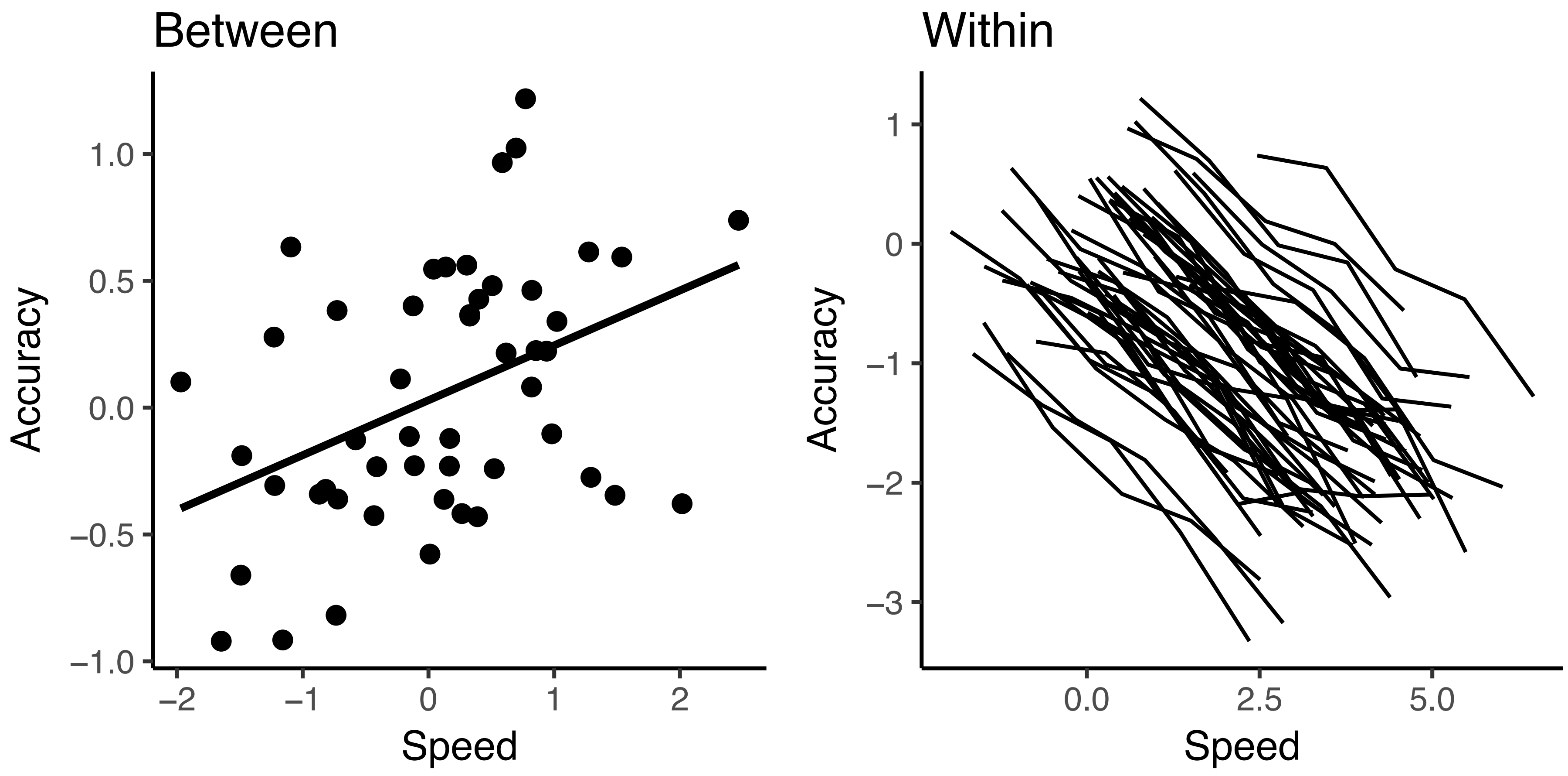
Advantages of the Multitrait-multimethod (MTMM) - modelling approach

- A theory-based integration / dimensionality reduction is applied to ROI-wise data from 3 imaging modalities
  - Represents a theoretical model of grey matter structural “integrity” or (micro-) “damage”
  - reduces multiple testing problem
- Region-specific variance, method-specific variance, and residual variance are separated (orthogonal to each other) and can each be investigated with respect to their associations with other variables of interest.

What‘s change got to do with it?

Longitudinal Modeling

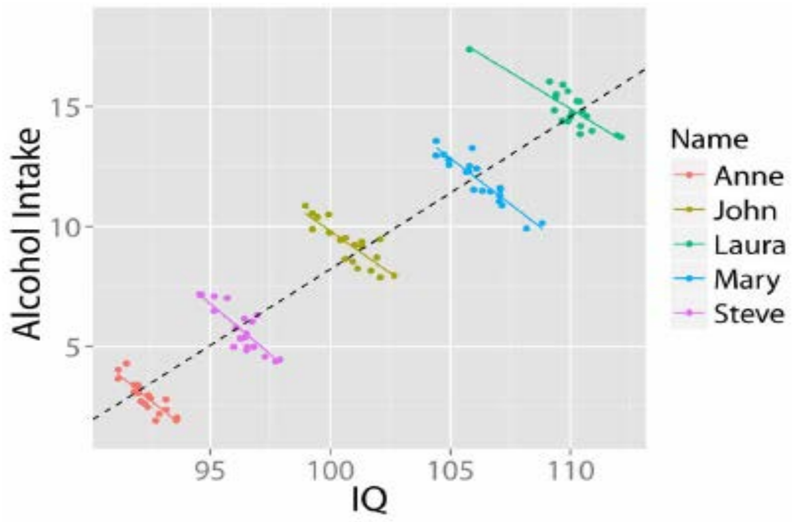
Within ≠ Between
(Or Simpson‘s Paradox)

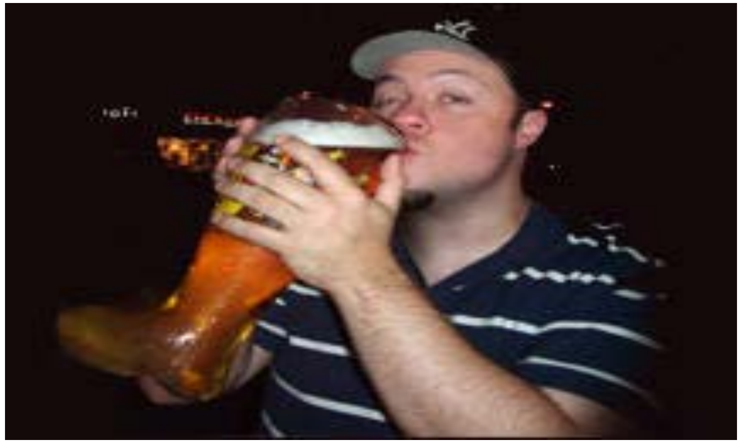


Modelled after Hamaker (2012)

Why Study Change?

- Alcohol and IQ
  - Positively correlated in the (UK) population (Batty et al., 2008)
  - Neither is *untrue:* they are claims about *qualitatively different phenomena*
  - Inter: Sociological mechanism
  - Intra: Pharmocological mechanism





Only Time will Tell: Cross-sectional Studies Offer no Solution to the Age-Brain-Cognition Triangle—Comment on Salthouse (2011)

Naftali Raz[1] and Ulman Lindenberger[2]

How to Study Change?

- 1) The identification of intra-individual change
- 2) The identification of individual differences in change
- 3) The identification of relationships among change
- 4) The identification of determinants of change
- 5) The identification of determinants of individual differences in change

Baltes, P. B., & Nesselroade, J. R. (Eds.). (1979). *Longitudinal research in the study of behavior and development*. Academic Press.

Latent Change Score Model

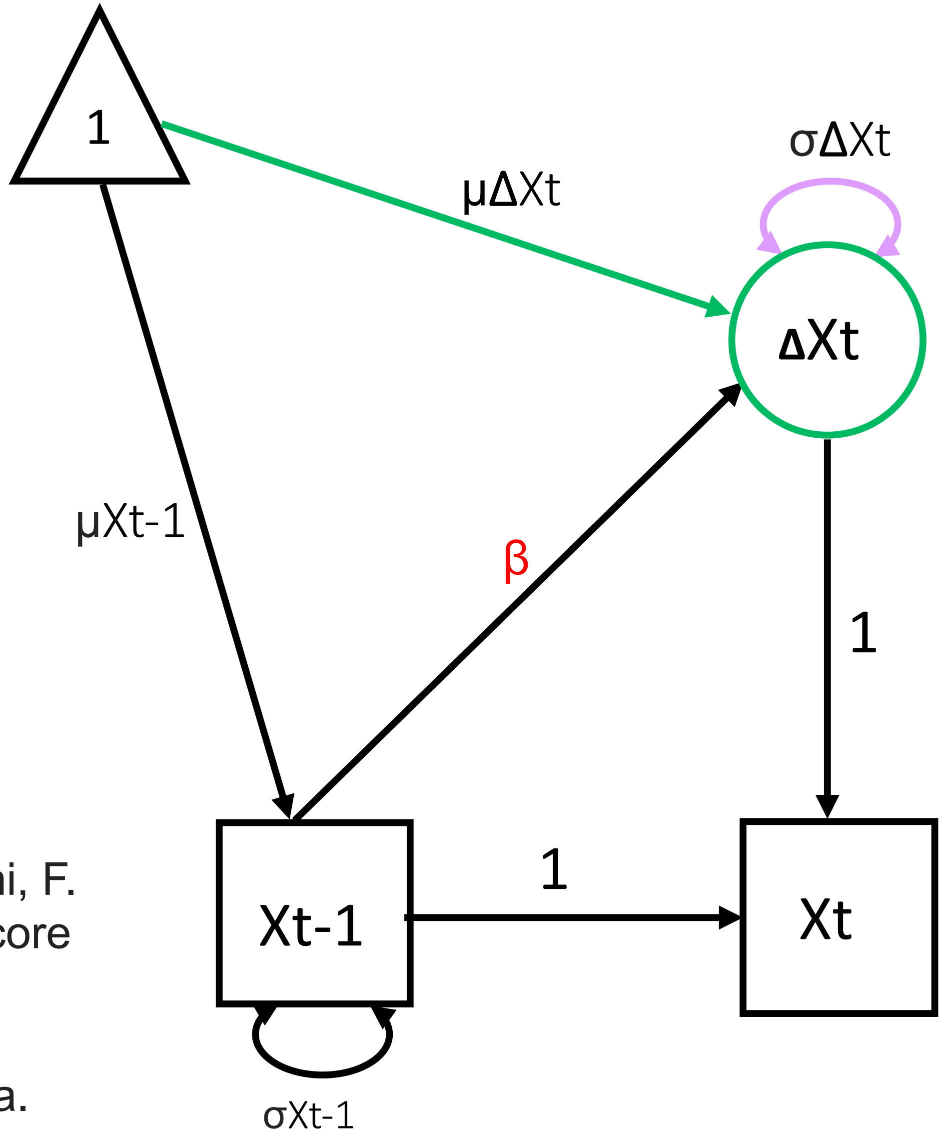
- Scores on at least 2 occasions
- Specify autoregressive path
- Create latent change factor which captures residual
- Mean ΔX == paired t-test
- But: 2 parameters for free
  - Change variance
  - Proportional change, or self-feedback parameter

$$1.\ X_{it} = X_{i,t-1} + \Delta X_{i,t}$$

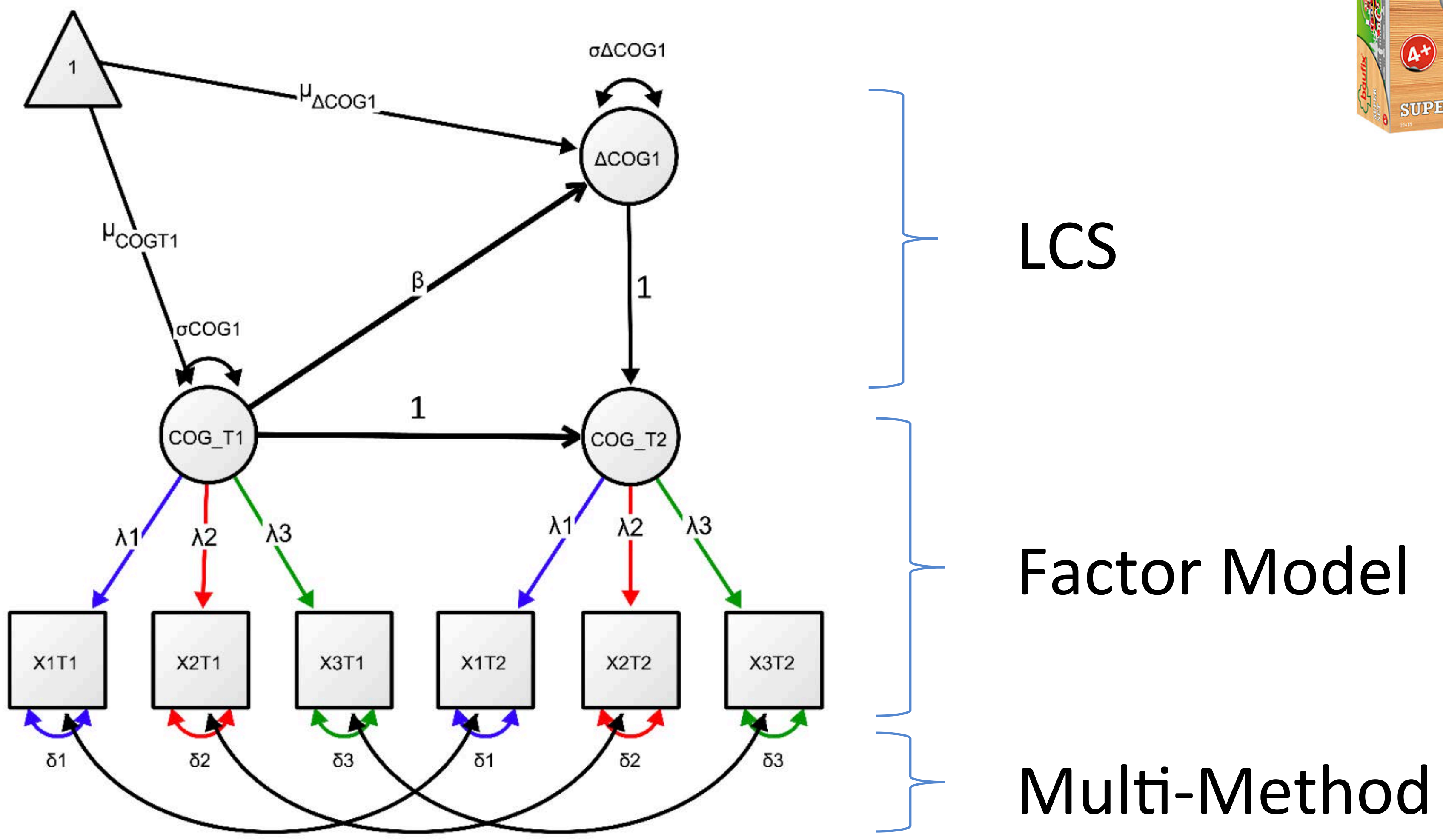
$$2.\ \Delta X_{i,t} = X_{it} - X_{i,t-1}$$

$$3.\ \Delta X_{i,1} = \beta * X_{i,t-1}$$

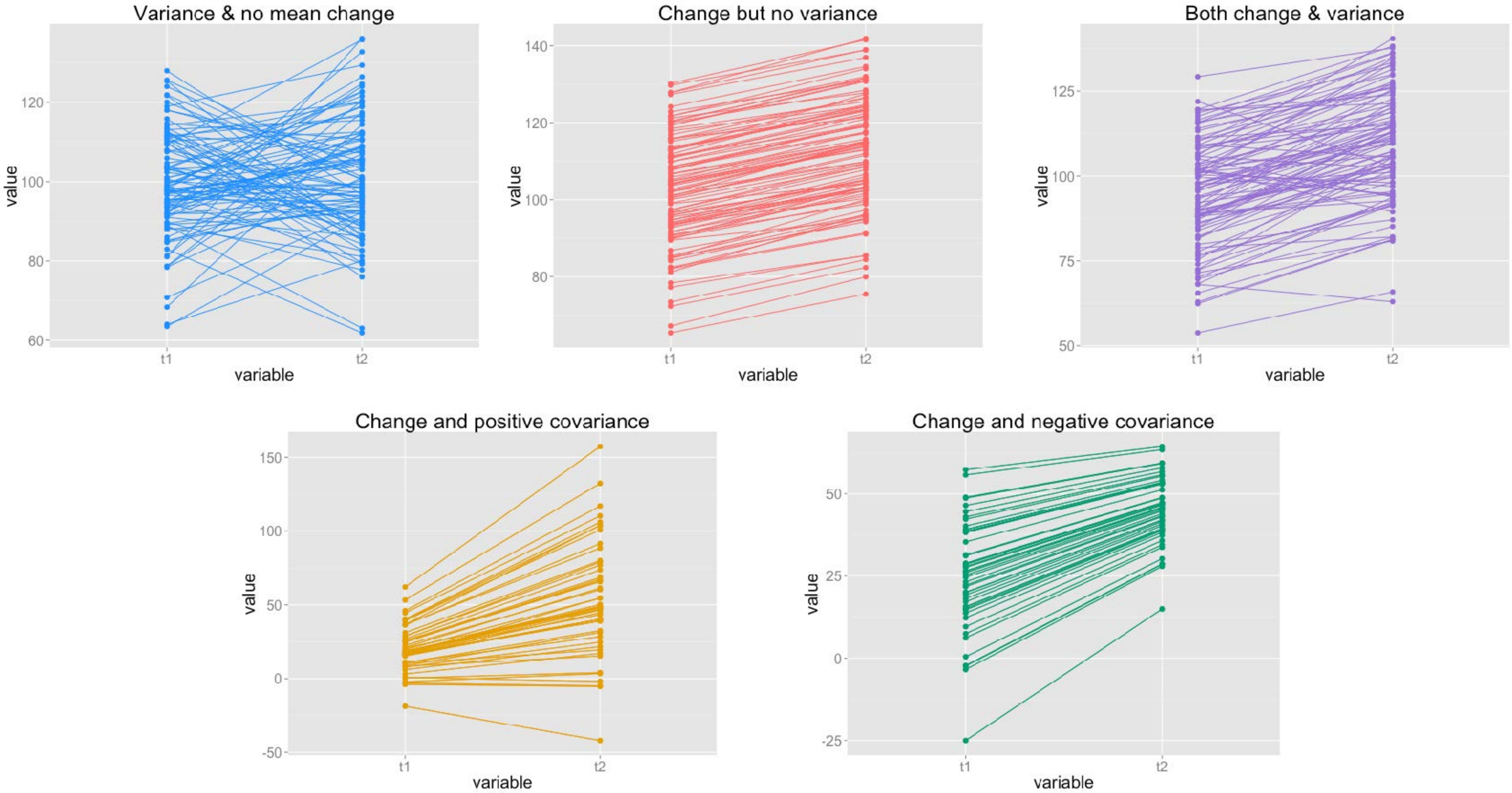
McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data.



A Multiple-Indicator Latent Change Score Model



Powerful tool to distinguish various mechanisms



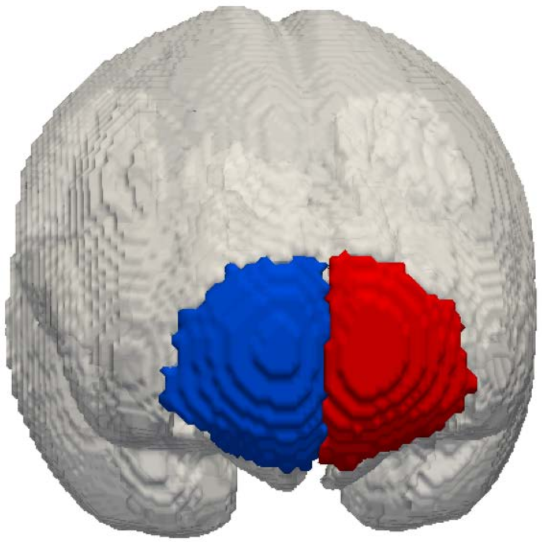
Application: Frontal lobe development in adolescence

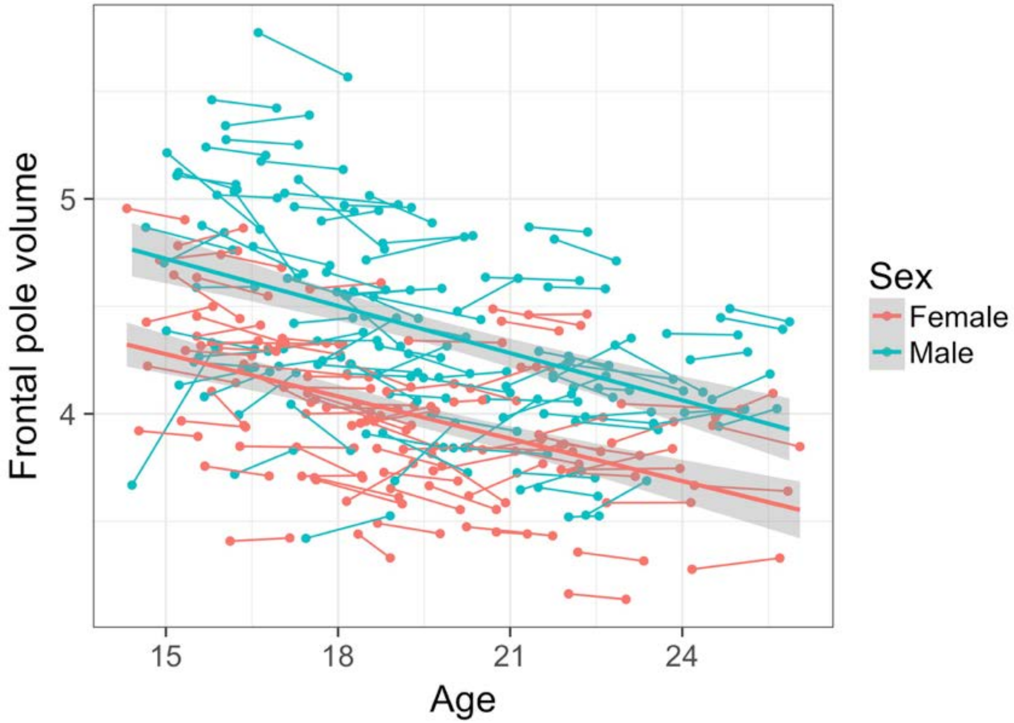
- Goal: Modeling volumetric changes in frontal pole
- Data from Neuroscience in Psychiatry Network (NSPN), 176 individuals, mean age = 18.84, range 14.3-24.9, 82 girls, scanned on two occasions (average interval: M=1.24 years, SD= 0.33 years).
- Assess degree of volumetric changes in the frontal pole, which is often discussed w.r.t speed of maturational changes and its purported role in controlling higher cognitive functions and risk taking behaviour

Kievit et al., 2017, Dev Cog Neurosci

Application: Frontal lobe development in adolescence

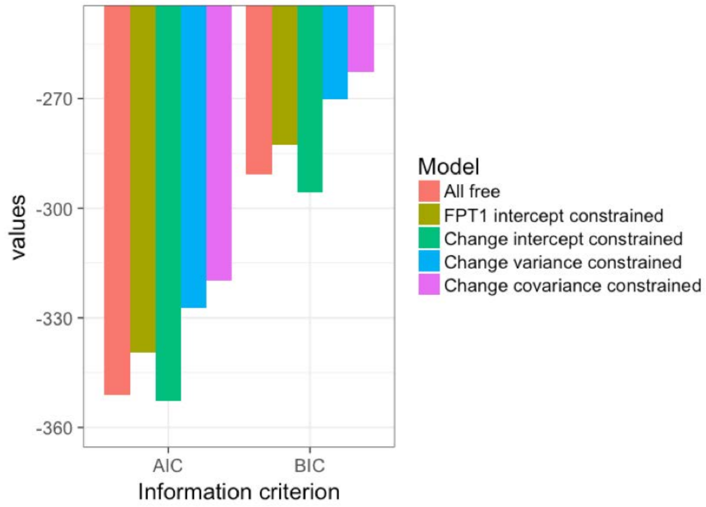
- Frontal lobe as latent variable measured by left and right volume
- Question: Are there sex differences in cortical development?

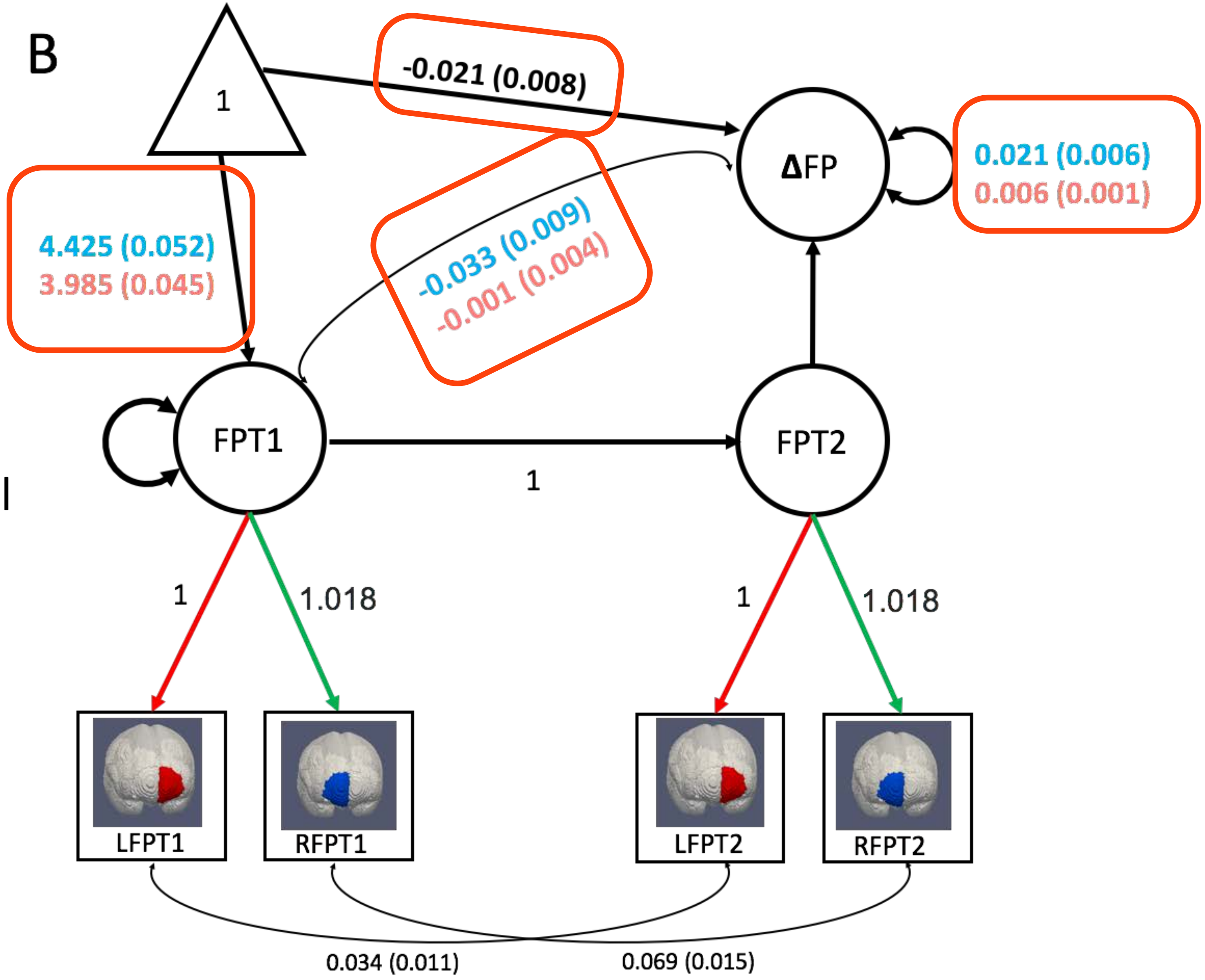




Results

- Sex difference in:
- Intercept
- Change variance
- Intercept-Change covariance
- No difference in mean change (so 'standard methods' would have missed interesting patterns)
- Pattern compatible with delayed cortical thinning in boys





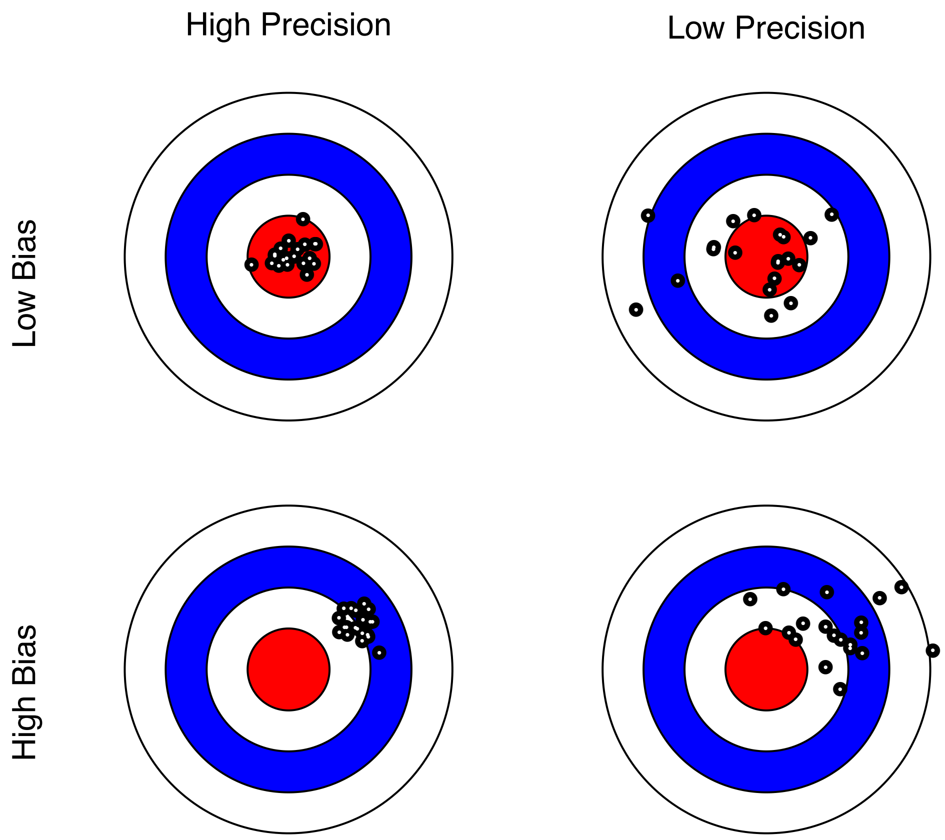
Reliability

ICED: Intra-Class Effect Decomposition

Reliability Is Necessary for Measuring Change

<u>Definition: </u>*The term reliability in psychological research refers to the consistency (or repeatability) of scores across repeated measures*

A measure is considered reliable if it would give us the same result over and over again (assuming that what we are measuring isn't changing!).

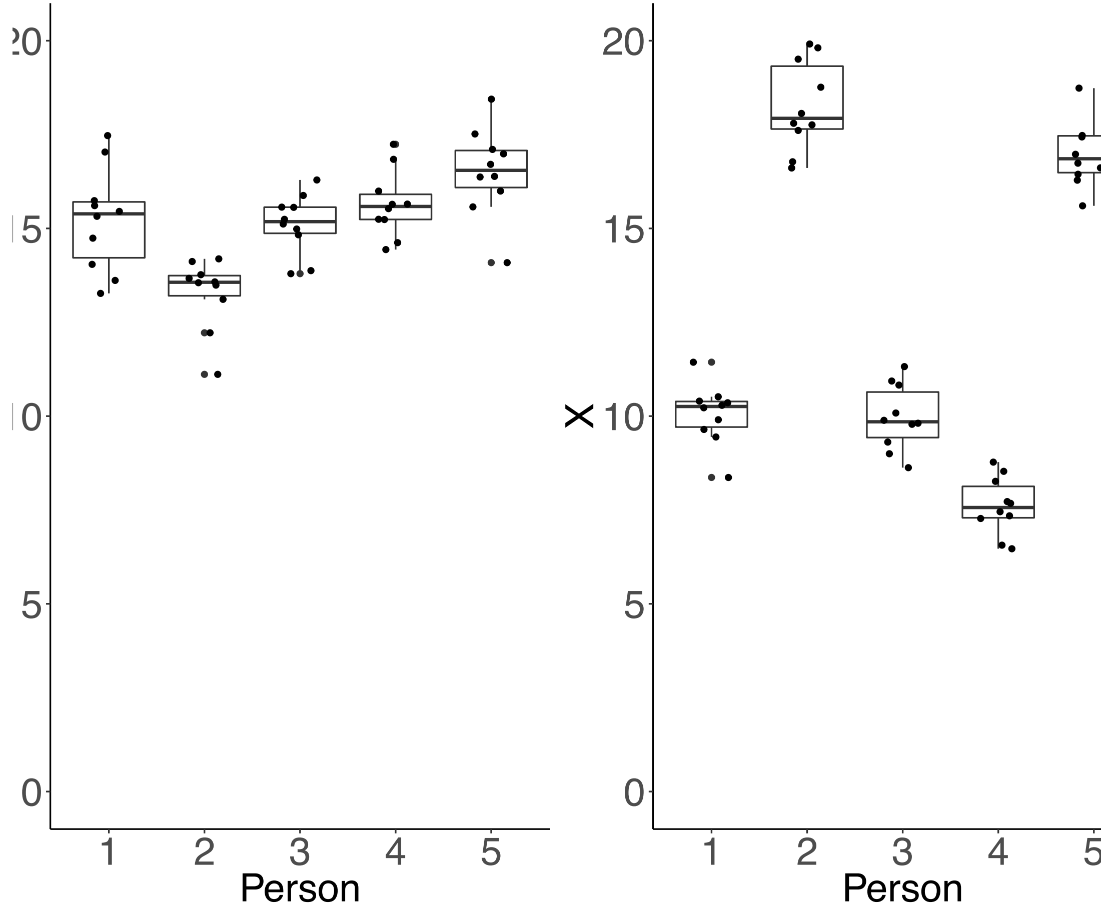


ISO 3103: The Standardized Method For Brewing Tea

Abstract: *The method consists in extracting of soluble substances in dried tea leaf, containing in a porcelain or earthenware pot, by means of freshly boiling water, pouring of the liquor into a white porcelain or earthenware bowl, examination of the organoleptic properties of the infused leaf, and of the liquor with or without milk, or both.*

=> Consistency!

Reliability: Coefficient of Variation vs Intra-Class Correlation

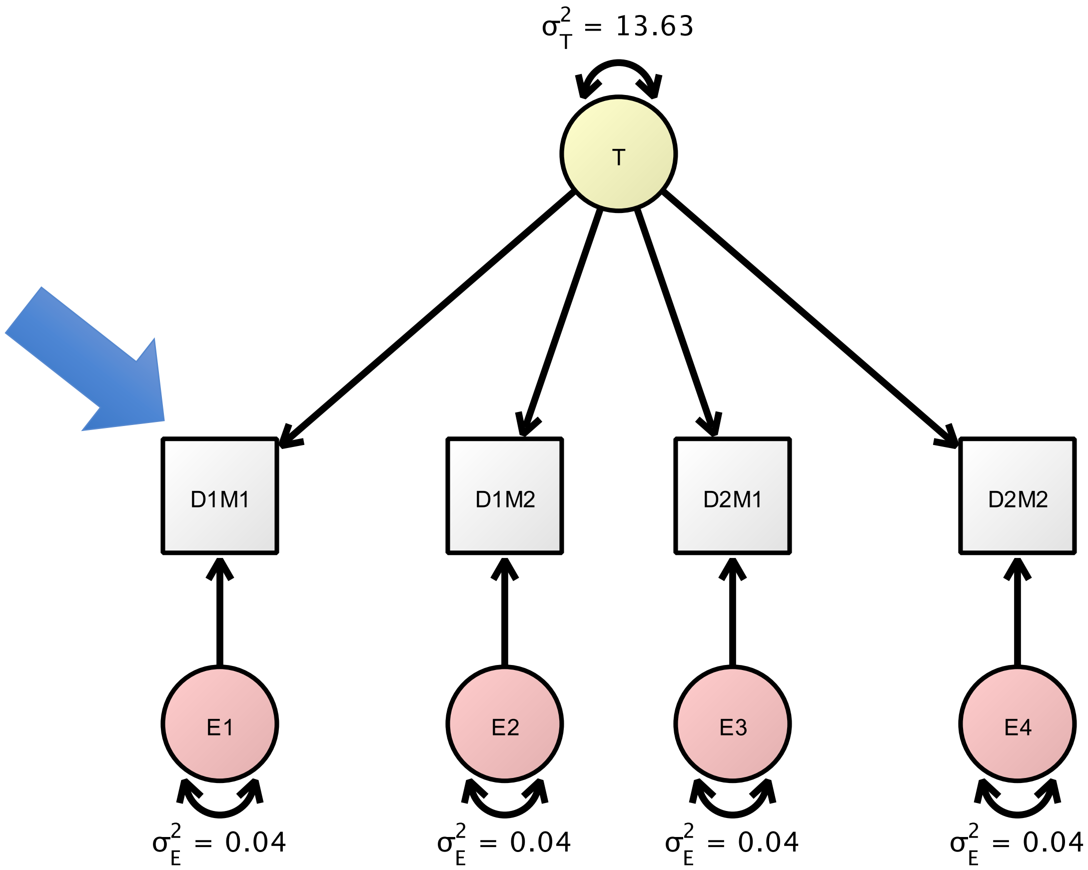


- **Coefficient of variation** relates average to within-person SD => reliability of averages
- **Intra-class correlation** relates between-person SD to within-person SD => reliability of individual differences

Brandmaier et al., 2018, eLife

Item-Level Reliability

- Reliability is the ***proportion of systematic variance*** in total observed variance
- Estimable from a repeated measures design
- In parallel forms models, ICC = Explained variance at any one occasion

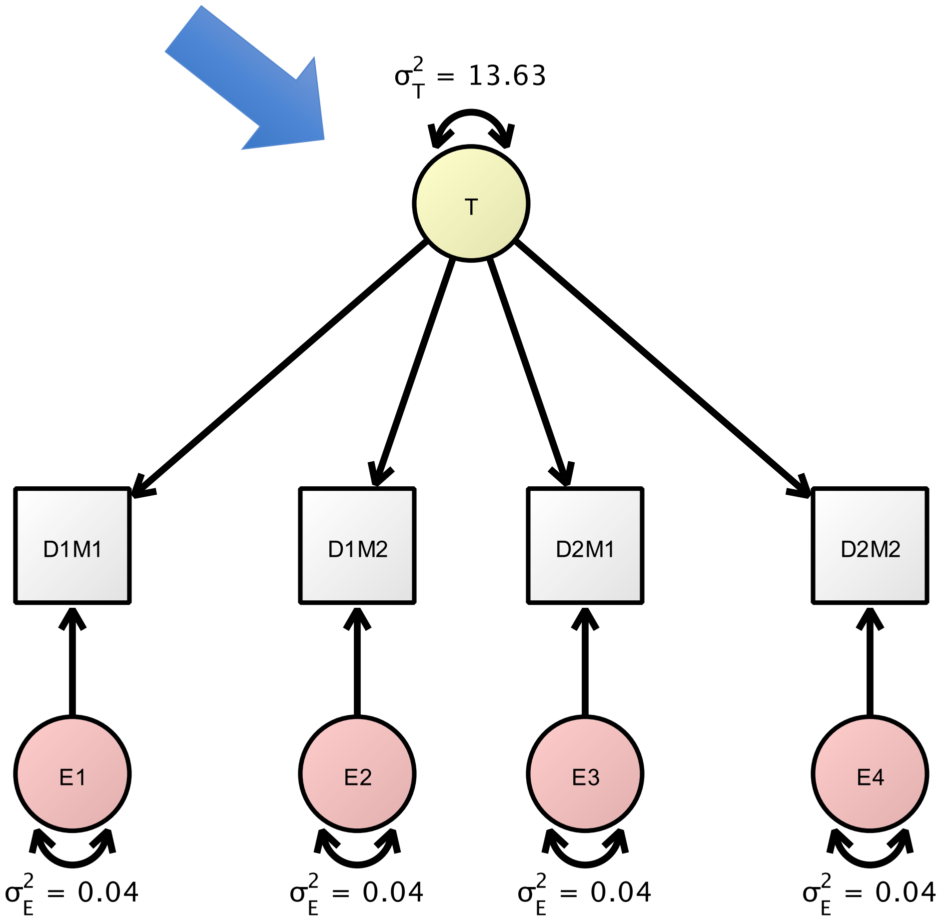


$$ICC = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}$$

Construct-Level Reliability

- ICC2 is the effective measurement error of the entire design measuring the latent score of interest
- In parallel forms models, this corresponds to dividing the measurement error by number of measurements

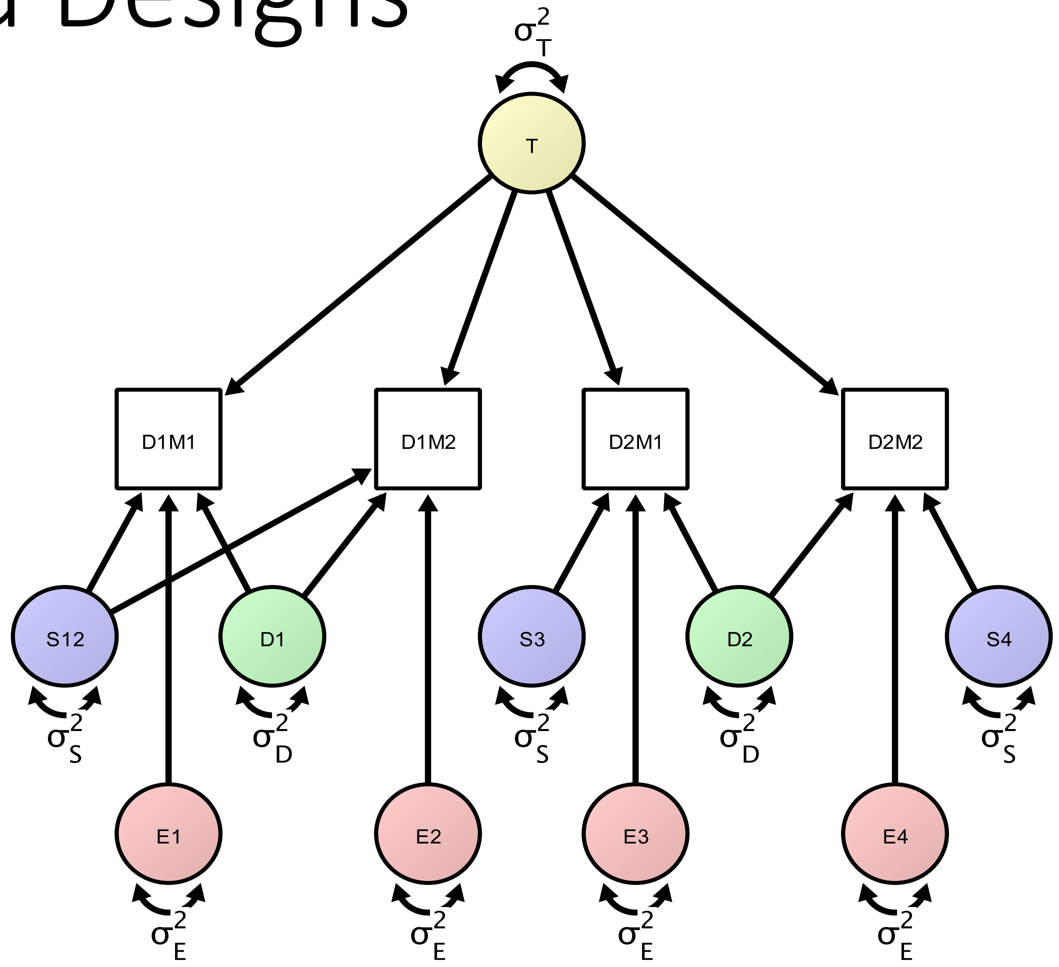
$$ICC2 = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2 / 4}$$



ICED: Reliability in Nested Designs

<table>
<tr><td colspan="2">Day #1</td><td colspan="2">Day #2</td></tr>
<tr><td colspan="2">Session #1</td><td>Session #2</td><td>Session #3</td></tr>
<tr><td>Scan #1</td><td>Scan #2</td><td>Scan #3</td><td>Scan #4</td></tr>
</table>

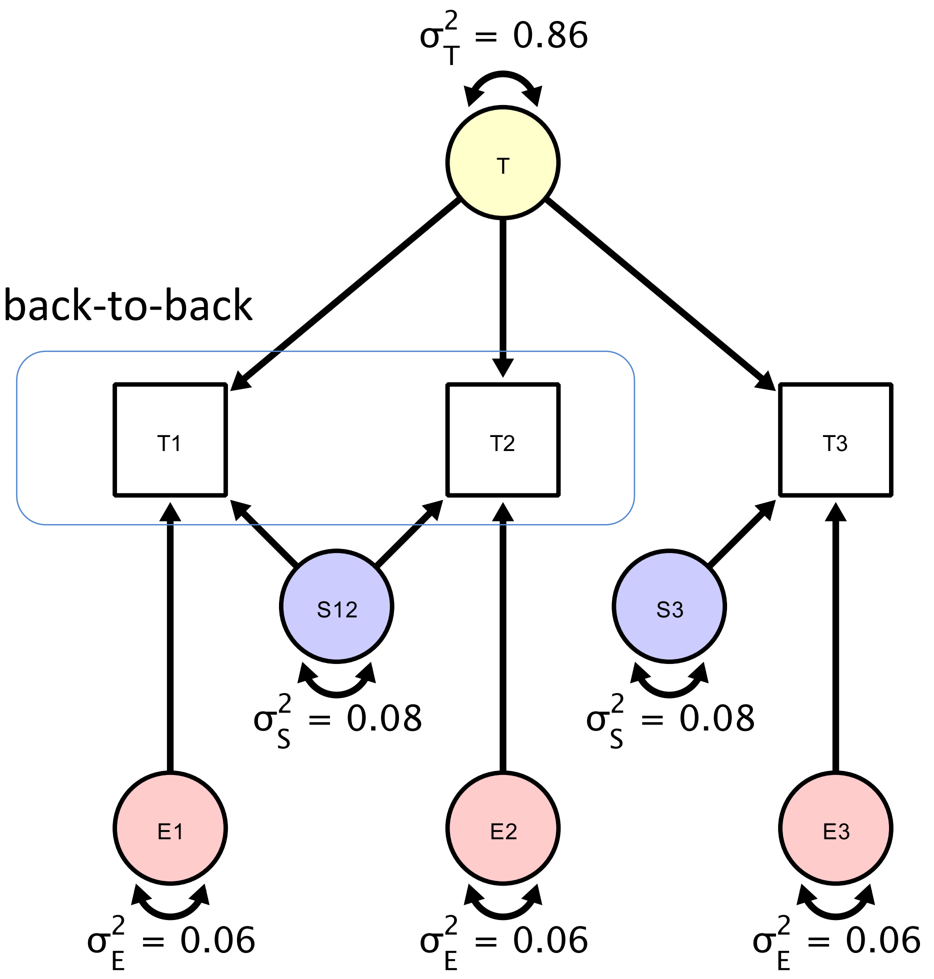
Time



Brandmaier et al., 2018, eLife

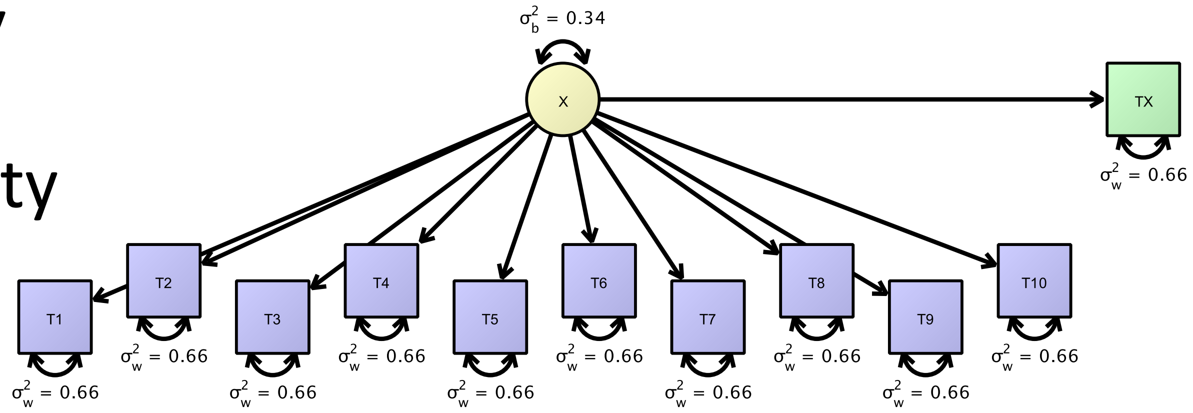
Application: Myelin

- Changes of myelin structure and quantity have been proposed as neuroanatomical substrates of cognitive decline
- 3 repeated Myelin-Water Fraction measurements in 20 healthy adults (24.4–69.5 years)
- Result: 86% true score, 8% session, 6% residual variance
- ICC=.86, ICC2=.94 (back-to-back ICC=.94)



Brandmaier et al., 2018, eLife

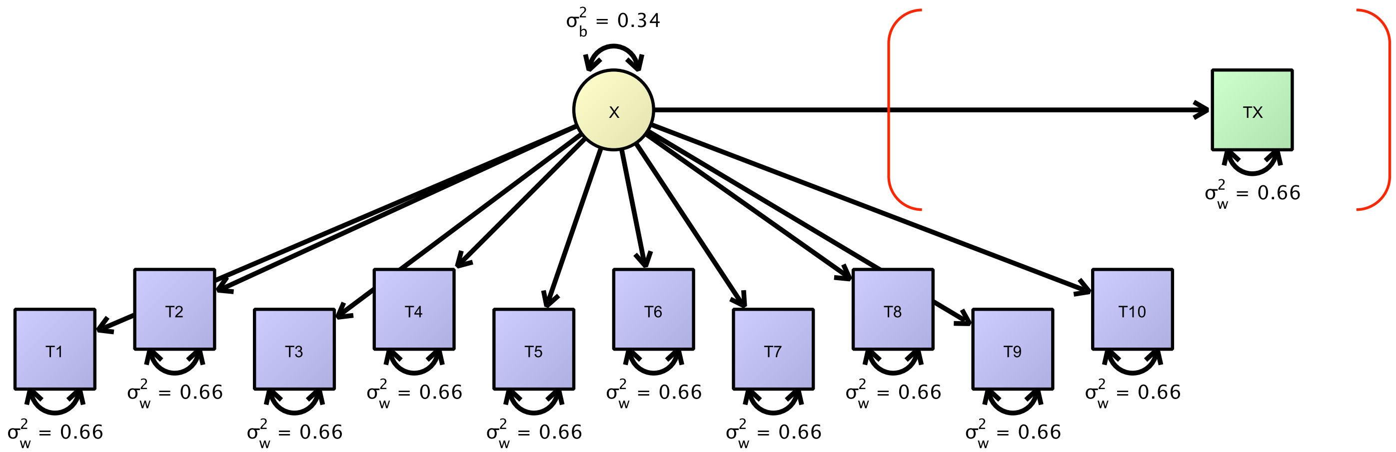
Application: Reliability of Resting-State Functional Connectivity



- Resting-state functional connectivity was proposed as a promising index of age-related or pathology-induced changes in the brain
- 5-minutes rsfc data from Pannunzi et al., 2018, which is based on the publicly available raw data from the Day2day study (Filevich et al., 2017) with up to 50 scans in 6 individuals + 1 scan each of 50 individuals
- pairs of ROI including pre-frontal, sensor-motor, parietal, temporal, limbic, occipital cortices, cerebellum and subcortical structures.

Result

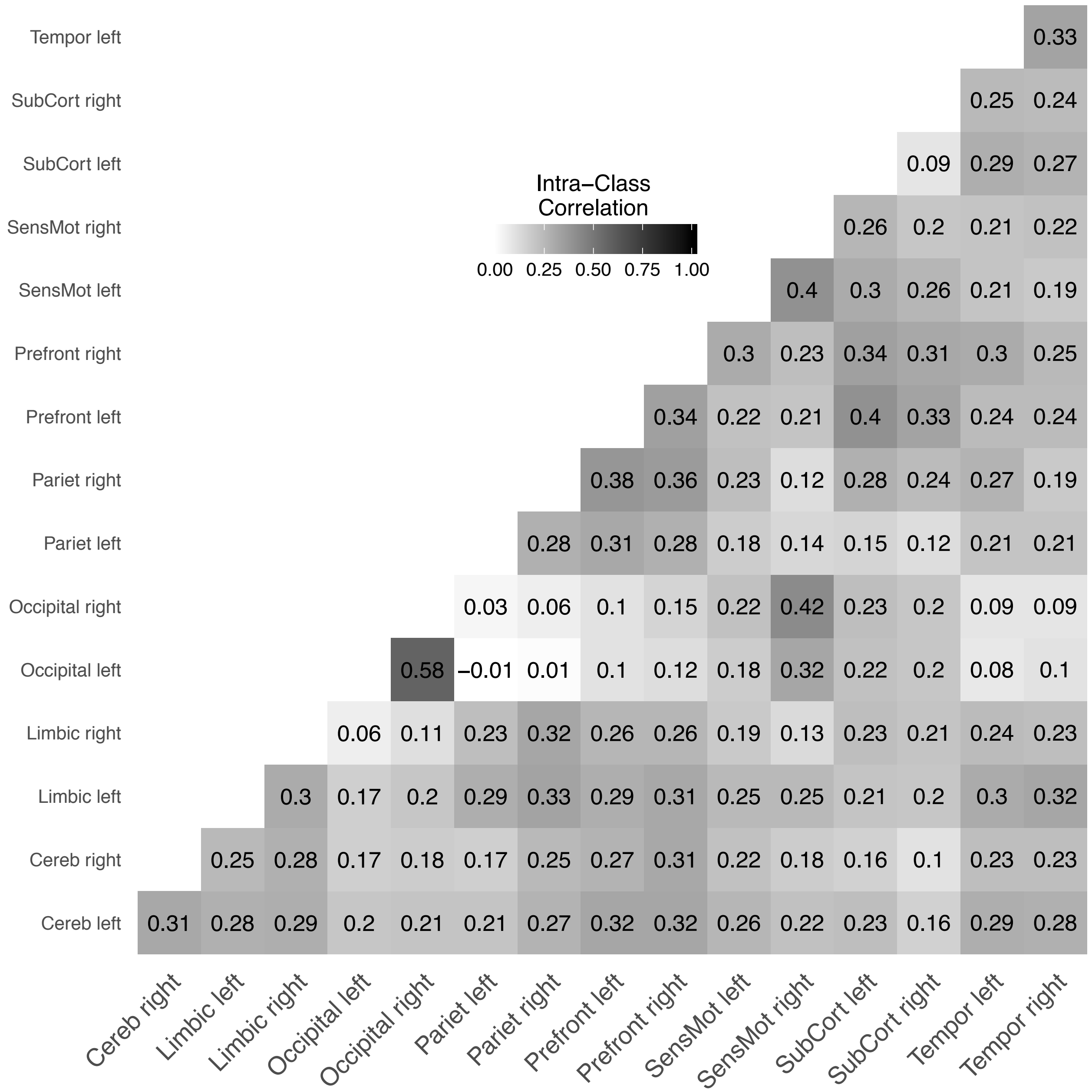
- Example rsfc: left prefrontal cortex and right prefrontal cortex
- Longitudinal only (n=6): The true score variance was estimated to account for 49% of the total variance (est = 0.013; W = 2.46; df = 1; p=0.117) and the error variance contributed 51% of the total variance (est = 0.014; W = 27.00; df = 1; p<0.0001)
- ICC is 0.49



Result II

Longitudinal + Cross-Sectional:

- ICC between left and right pfc dropped to 0.39.
- ICC Range 0.0 to 0.7
- Average ICC of .22
- Signal outweighted by noise by a factor of about 4 on average



Summary

- ICED framework is an extension of G-theory (Cronbach et al., 1972)
- ICED allows identification and estimation of measurement characteristics to precision of measurement / reliability
- Characteristics such as run, session, day, or scanning site (in multi-site studies).
- Knowing the sources of error will hopefully lead to better study designs (higher power, more resources saved)
- ICED can be extended to reliability of individual differences in change (Brandmaier et al., 2018, Frontiers in Psychology)

Does it all go together when it goes?

Coupled Changes in Cognition

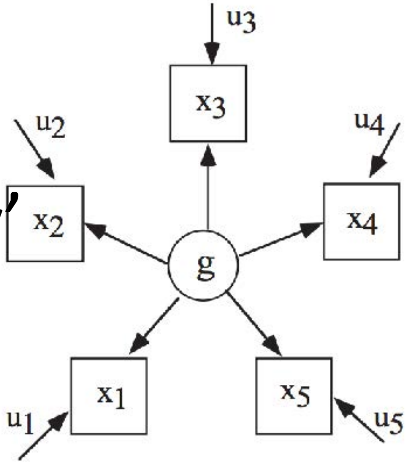
The Positive Manifold

- Test of cognitive abilities are universally positively correlated
- Among the most robust findings in all of psychology
- Famously: g model (1927)
- g scores predict
    - Health (Morbidity/mortality)
    - Income
    - Education
    - Psychopathology
    - Etc.

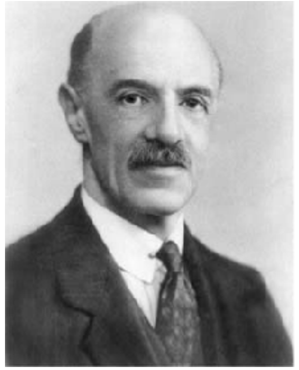
Table 1.4: Pearsonian intercorrelation matrix, combined kindergarten to adult sample (decimals omitted). 29 variables from the Woodcock-Johnson psycho-educational battery – revised, $N = 1425$ (correlations corrected for age).

Variable:		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Memory for Names	1	1000																												
Memory for Sentences	2	279	1000																											
Visual Matching	3	213	254	1000																										
Incomplete Words	4	167	255	191	1000																									
Visual Closure	5	148	103	178	176	1000																								
Picture Vocabulary	6	404	403	202	267	229	1000																							
Analysis-Synthesis	7	275	324	280	205	161	323	1000																						
Visual-Auditory Learning	8	342	343	267	192	205	382	376	1000																					
Memory for Words	9	208	559	221	245	066	225	215	246	1000																				
Cross Out	10	170	241	621	168	241	242	291	265	203	1000																			
Sound Blending	11	245	323	245	367	153	323	265	332	335	246	1000																		
Picture Recognition	12	293	216	212	123	234	256	233	299	155	257	213	1000																	
Oral Vocabulary	13	388	534	310	319	234	632	419	405	364	315	589	304	1000																
Concept Formation	14	306	382	306	236	206	325	481	376	237	305	275	255	458	1000															
Memory for Names (Delayed Recall)	15	721	236	155	168	129	383	269	460	173	123	242	236	359	284	1000														
Visual-Auditory Learning (Delayed Recall)	16	345	164	162	120	192	255	269	450	110	168	192	275	271	323	446	1000													
Numbers Reversed	17	259	416	384	227	129	255	368	321	401	309	316	206	396	354	225	182	1000												
Sound Patterns	18	233	257	204	221	109	269	271	259	243	229	291	168	331	299	222	214	262	1000											
Spatial Relations	19	280	266	278	158	265	317	389	369	169	343	225	238	388	404	240	289	311	294	1000										
Listening Comprehension	20	331	469	266	334	204	576	349	344	279	263	351	256	642	375	294	221	308	274	320	1000									
Verbal Analogies	21	379	454	334	228	242	522	455	445	310	344	355	322	639	496	377	330	403	304	465	526	1000								
Calculation	22	256	331	435	142	132	299	423	347	252	358	293	208	471	401	249	242	413	257	376	374	483	1000							
Applied Problems	23	337	416	419	206	175	439	470	388	312	388	360	273	603	489	313	268	438	315	486	524	651	655	1000						
Science	24	380	437	260	285	233	633	368	364	246	280	323	246	658	389	362	270	336	260	385	619	544	440	570	1000					
Social Studies	25	371	477	298	262	200	626	386	374	270	278	323	255	693	411	348	245	332	256	344	638	595	508	617	702	1000				
Humanities	26	390	447	308	281	252	622	343	414	297	284	355	283	665	359	368	285	326	282	340	572	598	427	536	633	672	1000			
Word Attack	27	281	370	356	263	119	316	303	366	322	255	484	202	468	329	269	228	398	316	313	326	415	422	450	346	354	398	1000		
Quantitative Concepts	28	342	427	408	205	162	497	437	416	309	361	320	244	615	413	337	280	433	299	440	513	624	636	728	602	637	576	471	1000	
Writing Fluency	29	225	350	494	193	123	260	309	347	266	410	358	196	398	335	197	194	365	229	276	285	394	420	426	293	336	409	488	434	1000

a.



‘Mental energy’



Deary, I. J. (2012) Intelligence. Annual Review of Psychology, Vol. 63, pp. 453-482, 2012

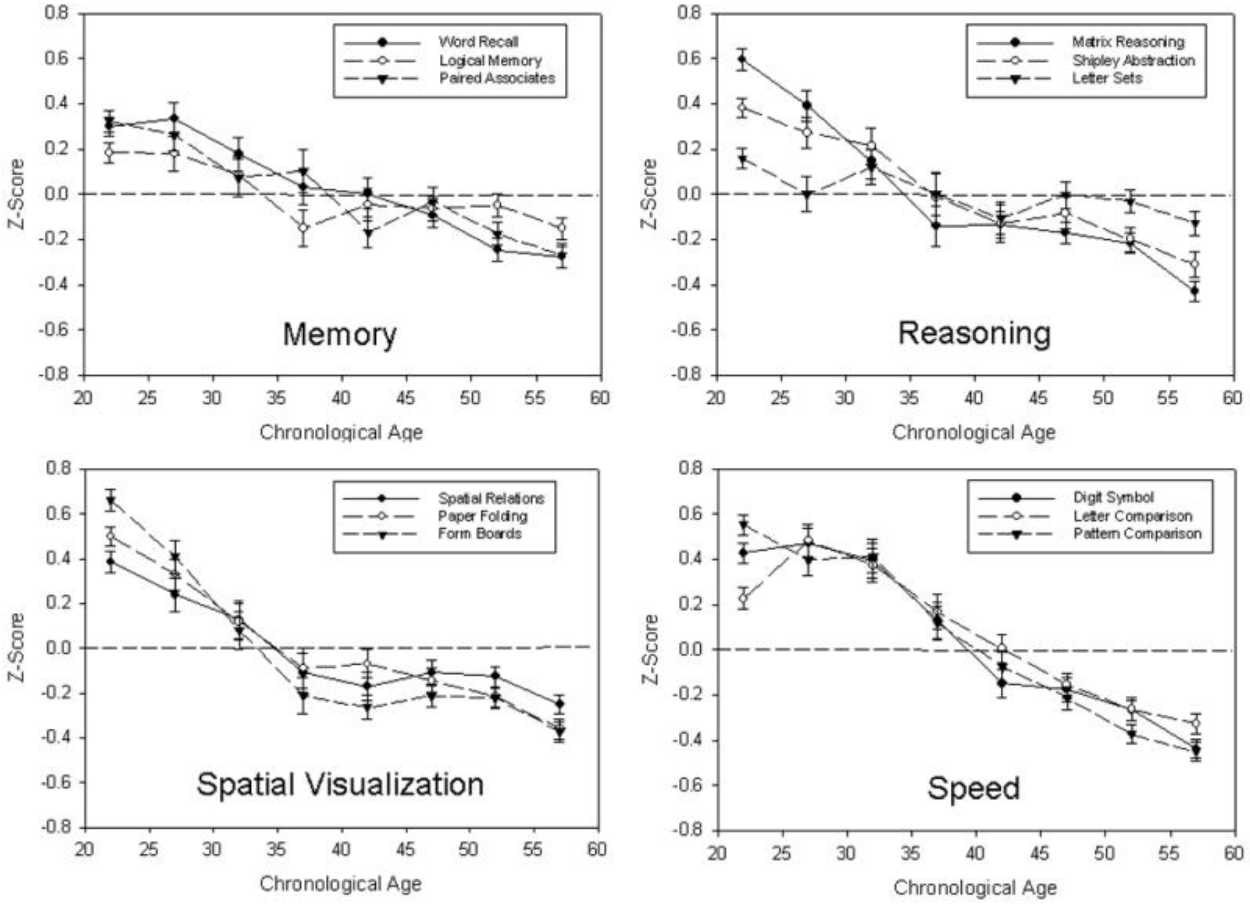
Gow, A. J., Johnson, W., Pattie, A., Brett, C. E., Roberts, B., Starr, J. M., & Deary, I. J. (2011). Stability and change in intelligence from age 11 to ages 70, 79, and 87: the Lothian Birth Cohorts of 1921 and 1936. *Psychology and aging*, *26*(1), 232.

Tucker-Drob, E. M. (2011). Neurocognitive functions and everyday functions change together in old age. *Neuropsychology*, *25*(3), 368.

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. Journal of Experimental Psychology: General, 132(4), 566.

Coupled Cognitive Change in Ageing

- Is individual cognitive decline a general process, or is it differentially manifested across different domains?
- Within-person level of analysis: Does decline within a person tend to occur simultaneously across different cognitive abilities?



Salthouse, 2009, Neurobiol Aging

Cross-Sectional vs. Longitudinal „g“

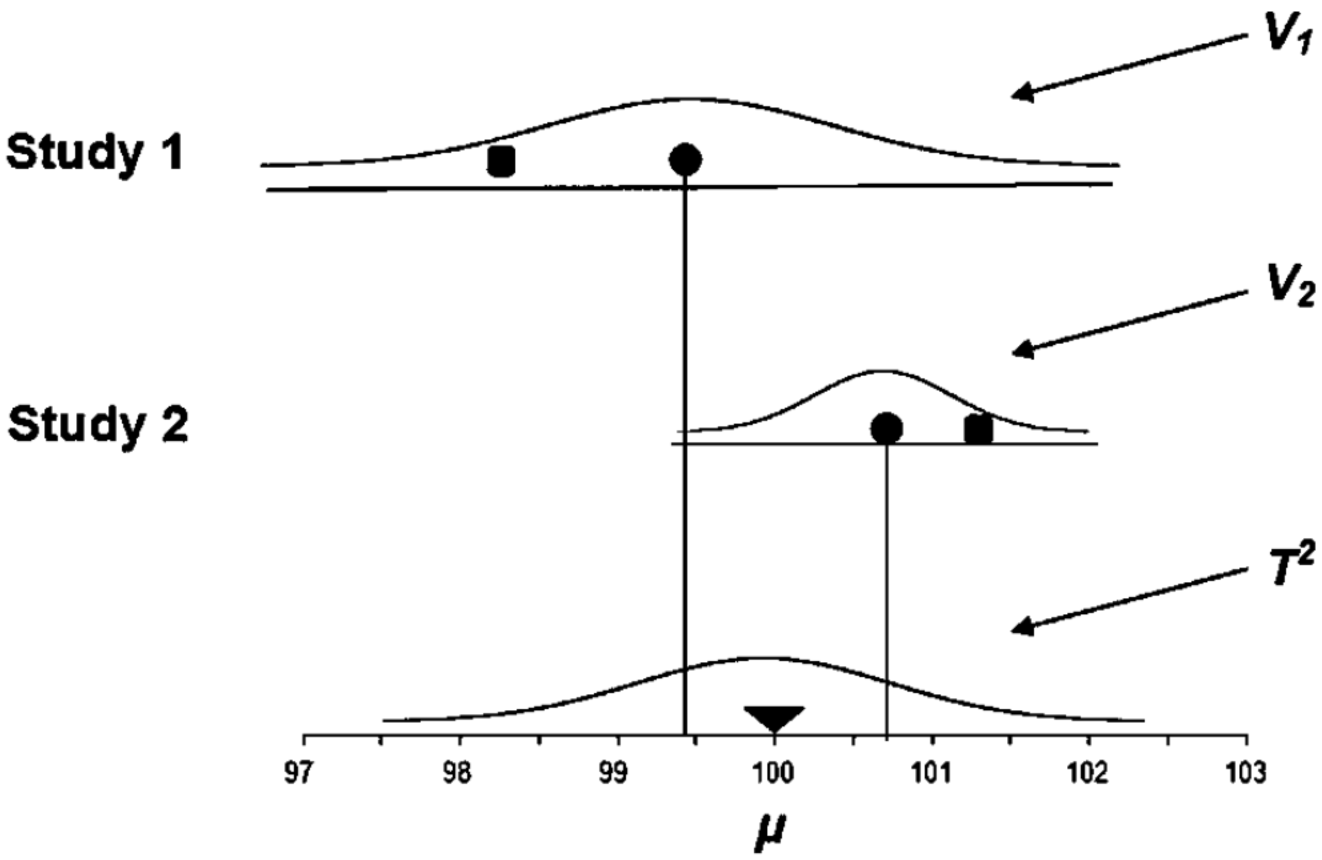
- Longitudinal factor structure may barely resemble longitudinal structure
- Development: FS reflective of how heterogeneity in environmental experience is structured over childhood (e.g., experiences that foster growth in one ability tend to co-occur with other experiences that foster growth in other abilities, or broad effects of intellectual engagement and achievement motivation on many different cognitive abilities),
- Ageing: specific neurodegenerative processes in different neural structures and functions that each subserves a different ability

Meta-Analysis: Coupled Change in Cognition

- Approach: Multi-level meta-analysis (effect sizes may randomly vary among studies) of longitudinal studies of coupled change in ageing (LCS and LGCM)
- Effect size: „shared variance", communality from a factor model fit to longitudinal changes in indicators of two or more ability domains
- Weight: asymptotic standard errors based on effective error (Brandmaier et al., 2018)
- Data: 89 effect sizes from 98 cognitive outcomes from 22 unique datasets composed of over 30,000 unique individuals in total, number of waves ranged from 2 to 12, with a median of 5.00 with the average age at baseline wave ranged from 35.42 years to 84.92 years, with a median of 64.90,

Tucker-Drob, Brandmaier, Lindenberger, submitted

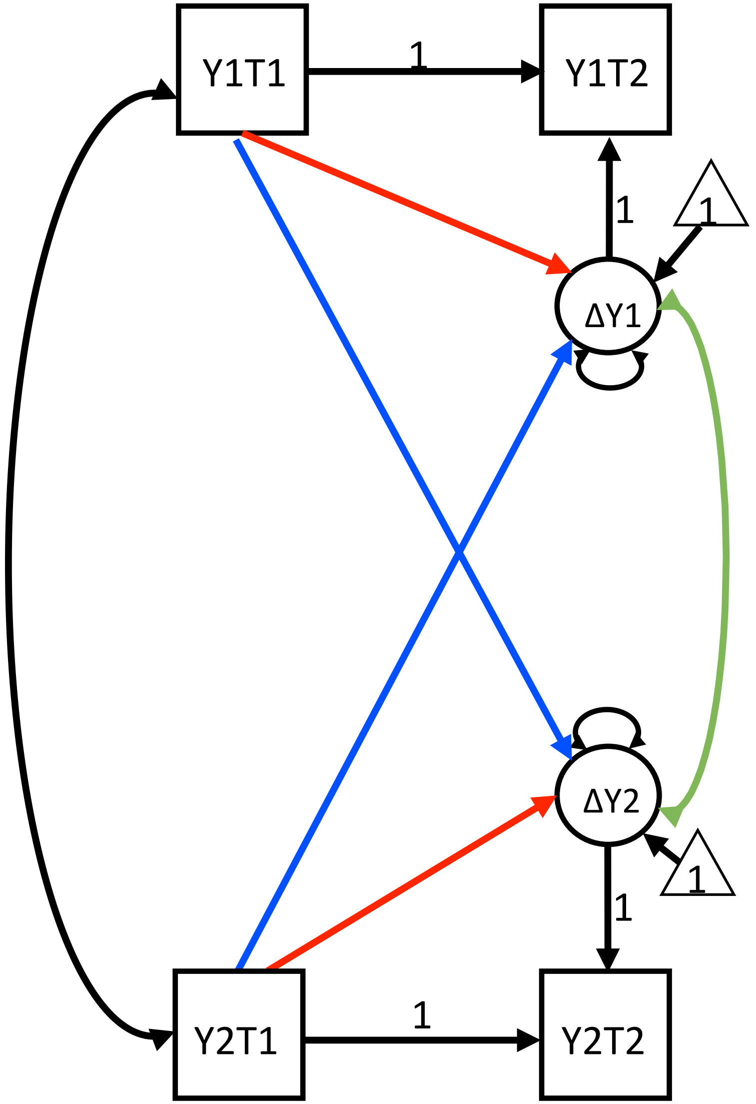
Random-Effects Meta Analysis



Bivariate Latent Change Score Model

- Simple extension
- Assume two domains (Y1 and Y2)
- Allows for investigation of
    - Self-feedback parameters
    - Cross-lagged parameters
    - Coupling parameter

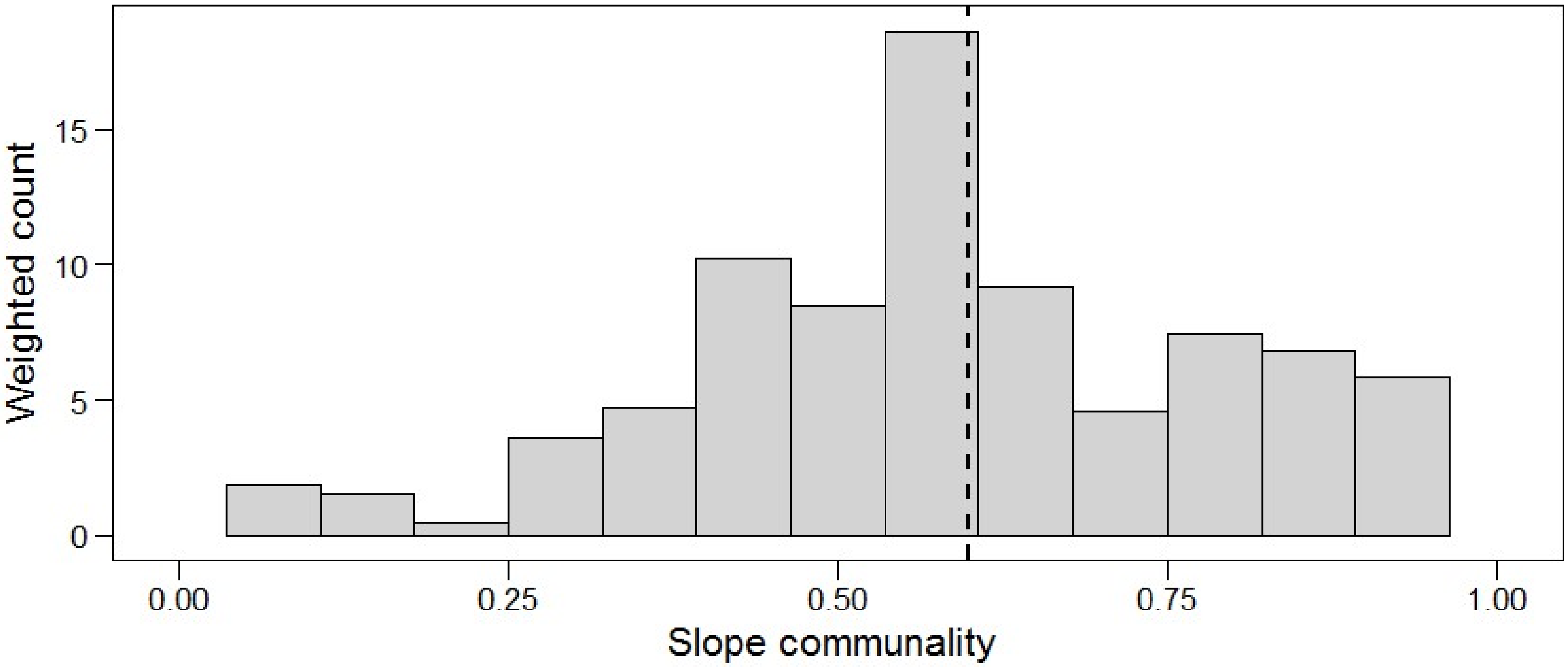
$$\Delta Y1_{ti} = \boxed{\beta 1} * Y1_{it-1} + \boxed{\gamma 12} * Y2_{it-1}$$



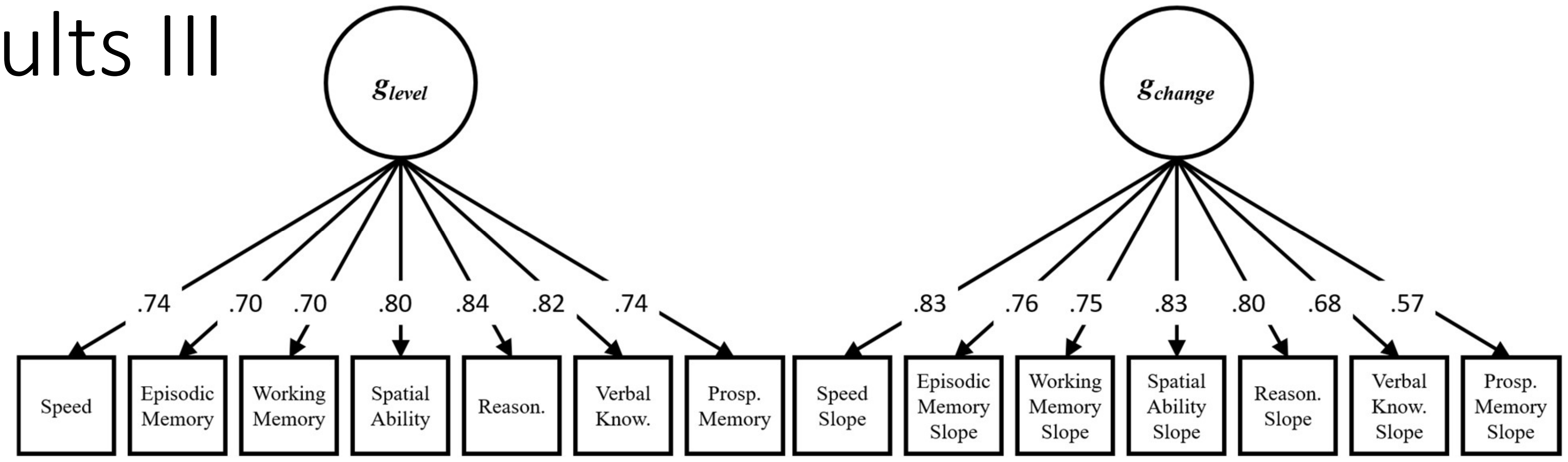
Results

- Mean change across domains was –.051 (SE = .007, $p < .0005$), that is, 1/20 SD change per year, or ½ SD per decade
- Mean rates of longitudinal change varied substantially across domains and across samples.
- Individual contrasts indicated that processing speed, spatial ability, and reasoning displayed significantly more decline (more negative) than the grand mean estimate across domains, and verbal knowledge displayed significantly less (less negative) decline than the grand mean estimate.

Results II - Slope Communality Estimates



Results III



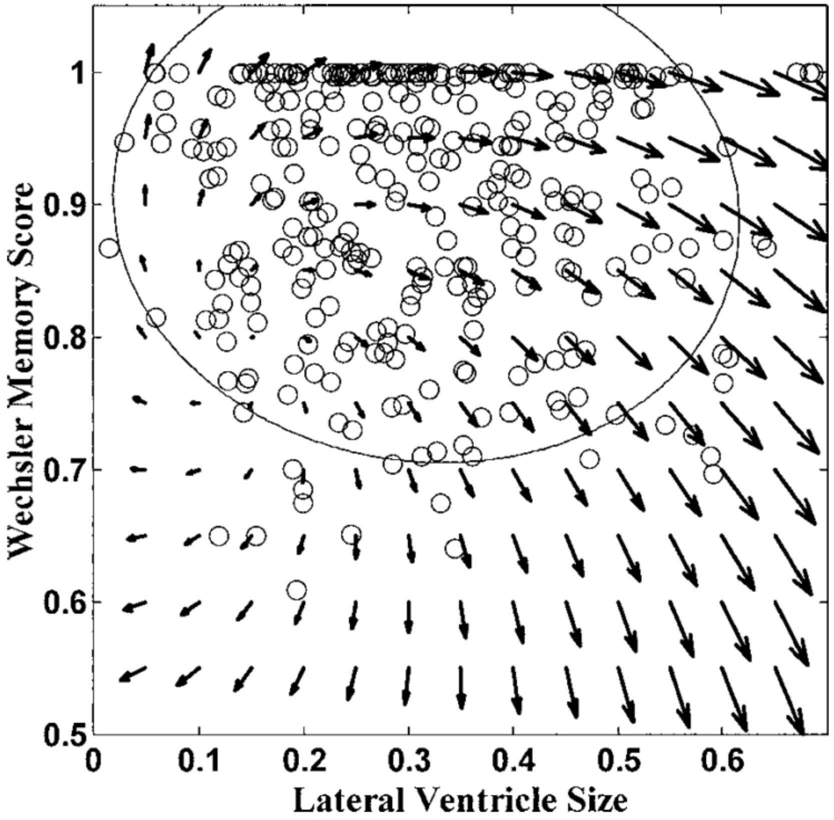
- An average of 60% of the variance in aging-related cognitive changes is explained by the common factor
- Longitudinal changes in different cognitive abilities changes are moderately-to-strongly correlated with one another.
- Despite pronounced differences in fluid and crystallized mean patterns of change, individual differences are coupled

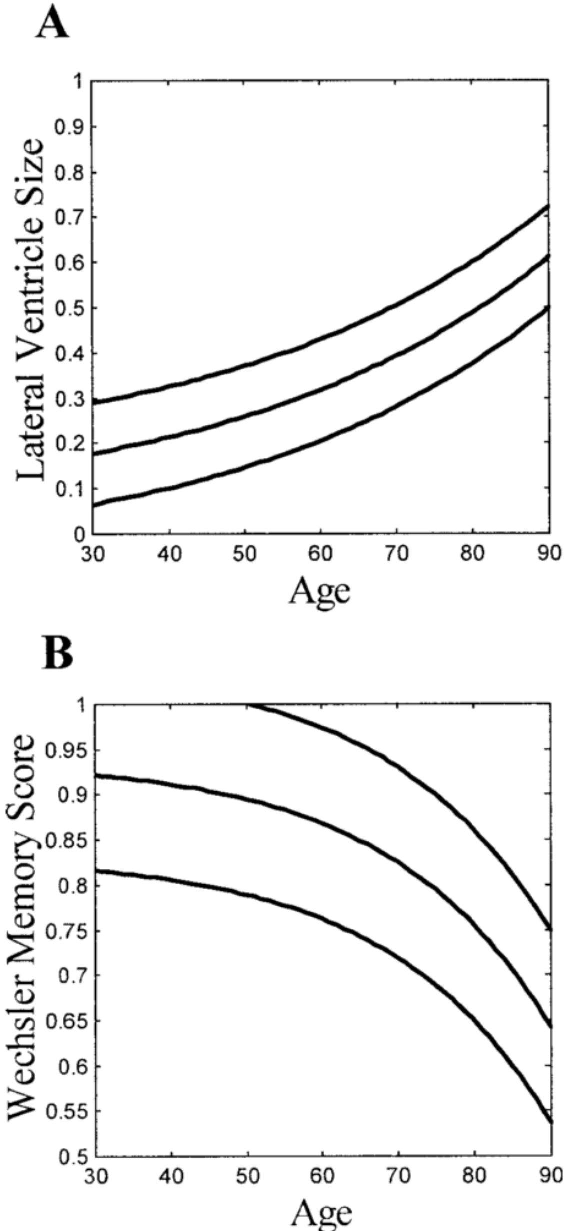
Results IV

- Abilities become increasingly correlated with adult age (see Baltes et al., 1980)
- A common g factor should account for increasing variance in abilities with age
- Mean age at baseline was positively related to slope communalities ($b = .005$, $SE = .002$, $p = .001$)
- Evidence for Dynamic Dedifferentiation

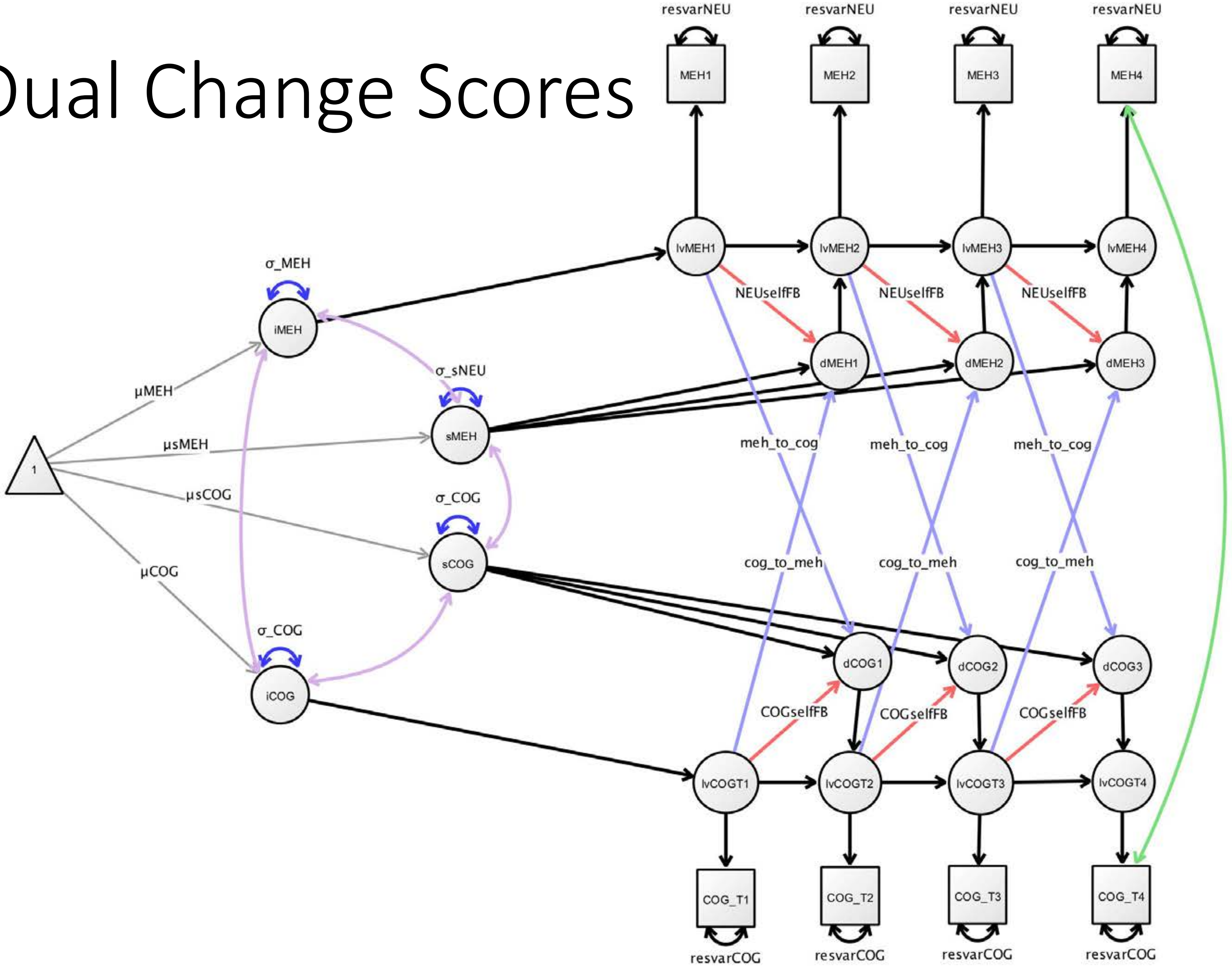
Brain and Cognition: Coupled Latent Change Scores

- McArdle et al. 2004: Larger ventricles -> faster memory decline
- Grimm et al. 2012: *Increase* in ventricle size further explains memory decline





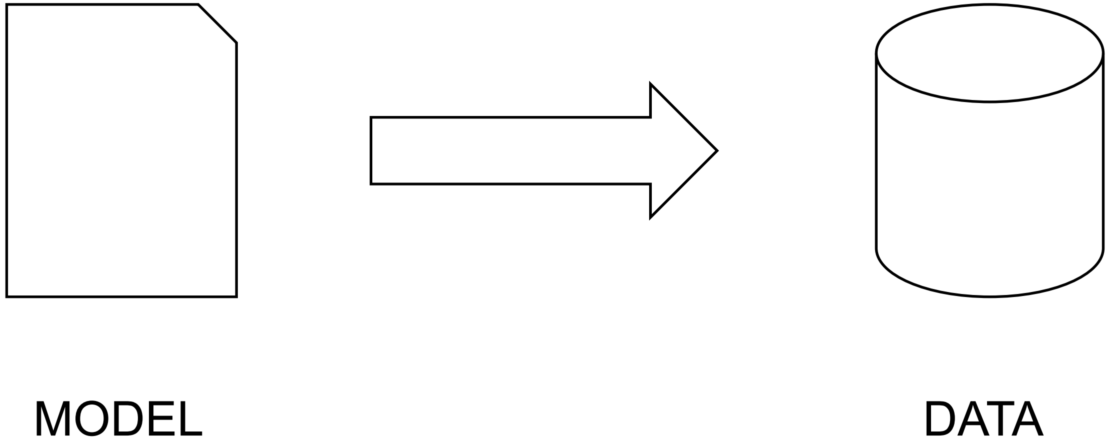
Coupled Dual Change Scores



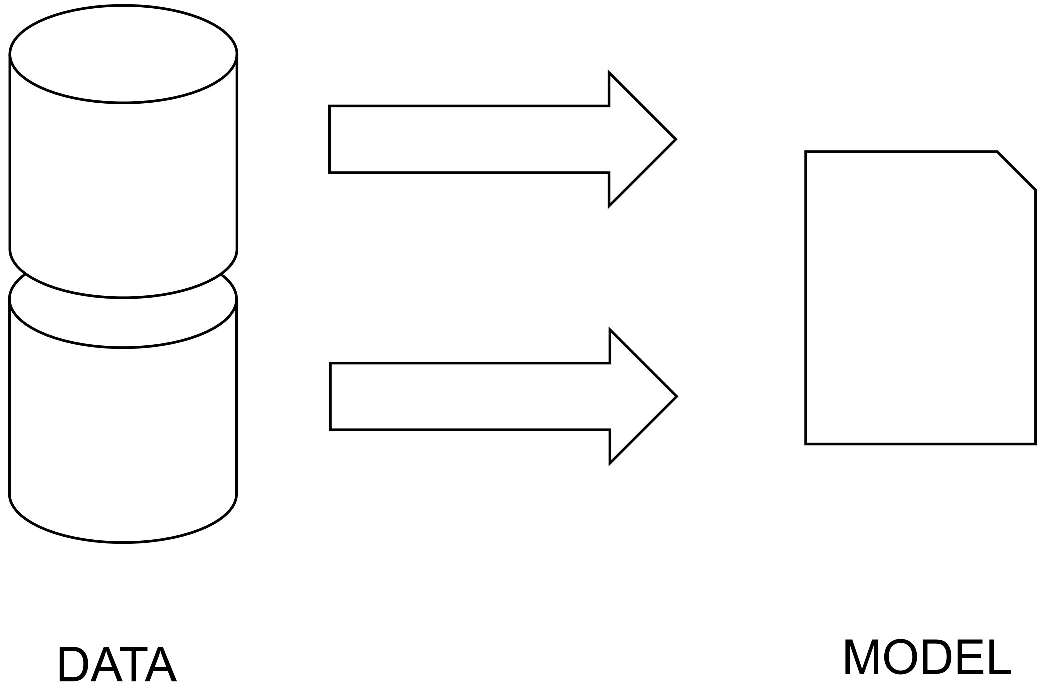
Suggested level of excitement

You survived! Questions?

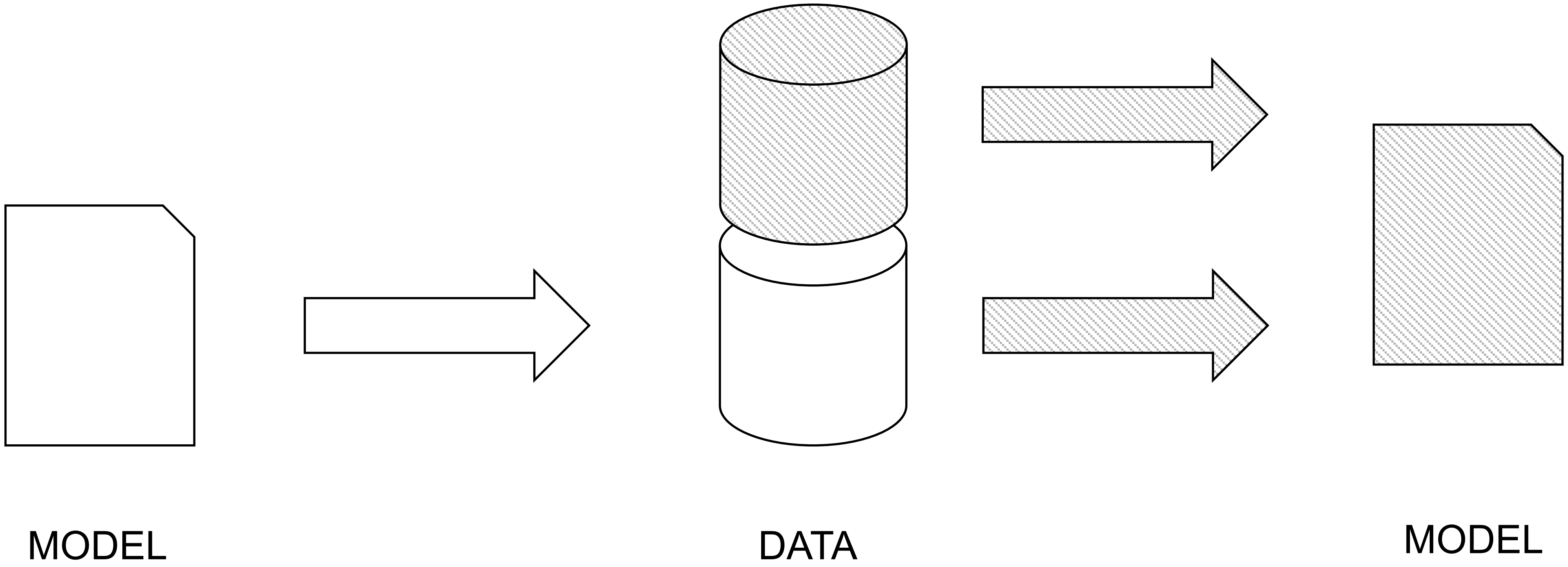
Theory-Based/Explanatory Modeling



Data-Driven/Predictive Modeling



Theory-Driven Exploration

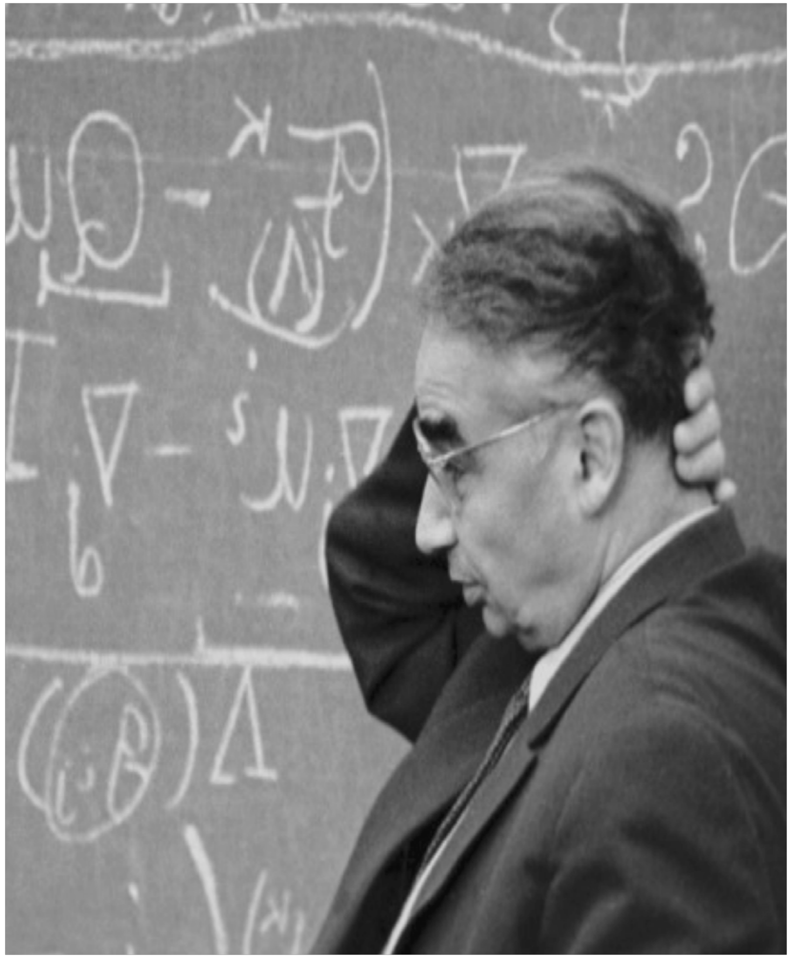


Typical Questions Asked

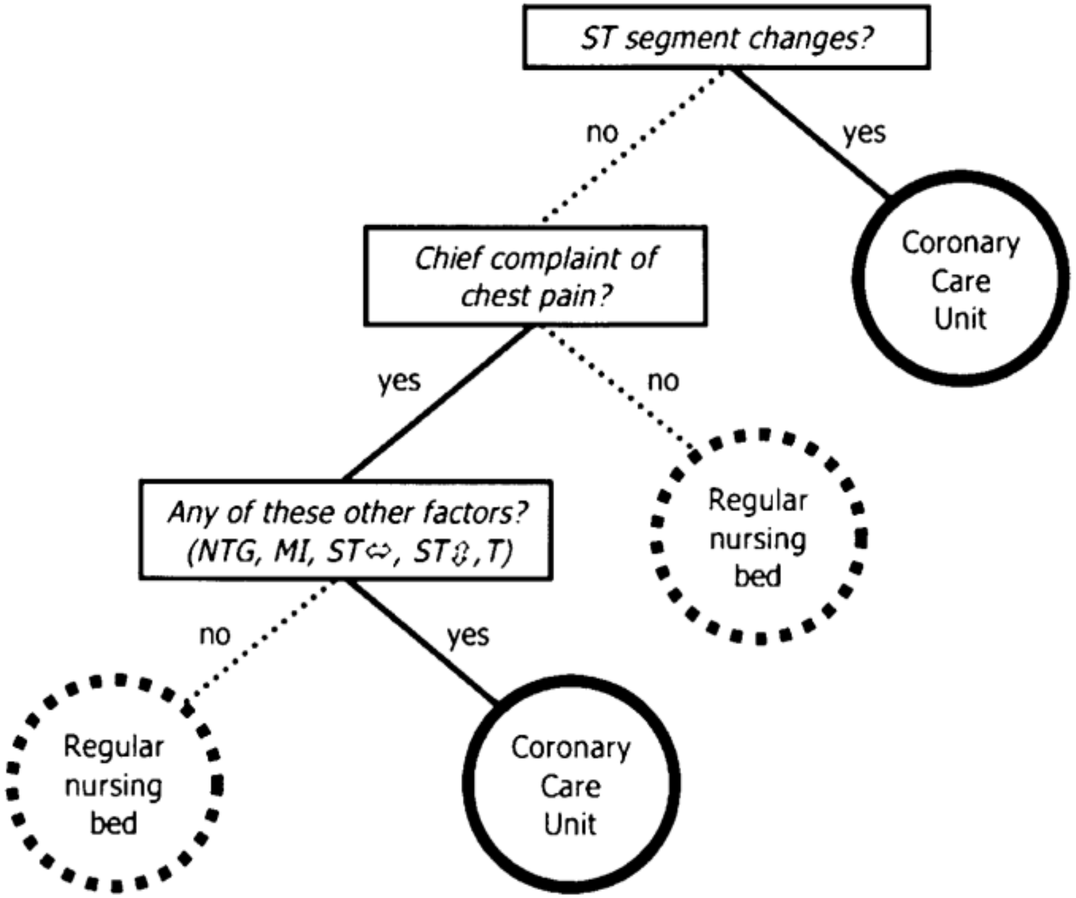
Given a theory/model:

- “How can we best explain the observed heterogeneity/ uncertainty?”

- “What subset of variables is most predictive about my outcome(s)?”



The Decision Tree

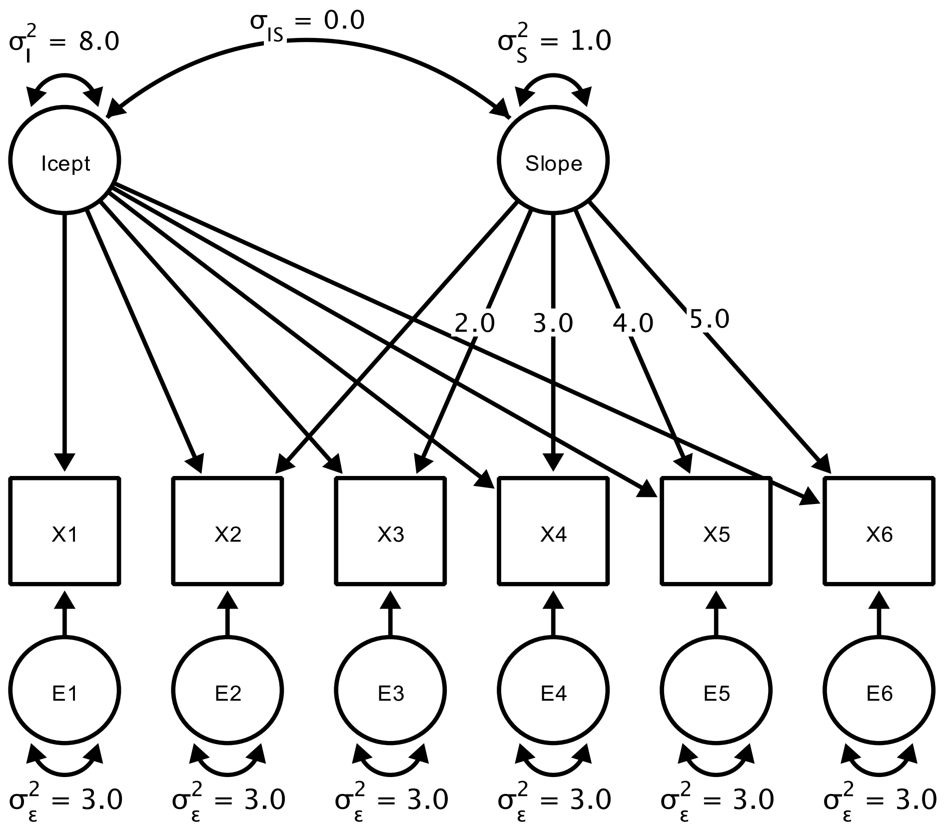


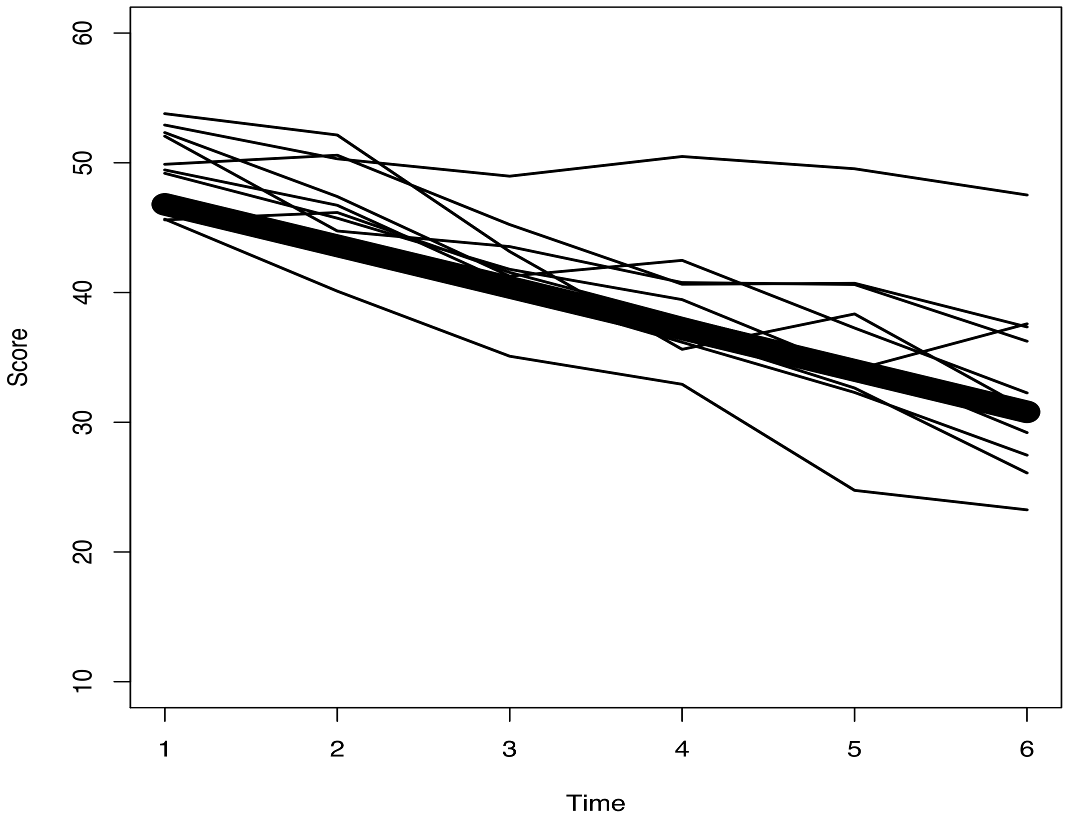
Gigerenzer & Kurzenhäuser 2005, Green & Mehr 1997

What if…

…we combined SEM and decision trees?

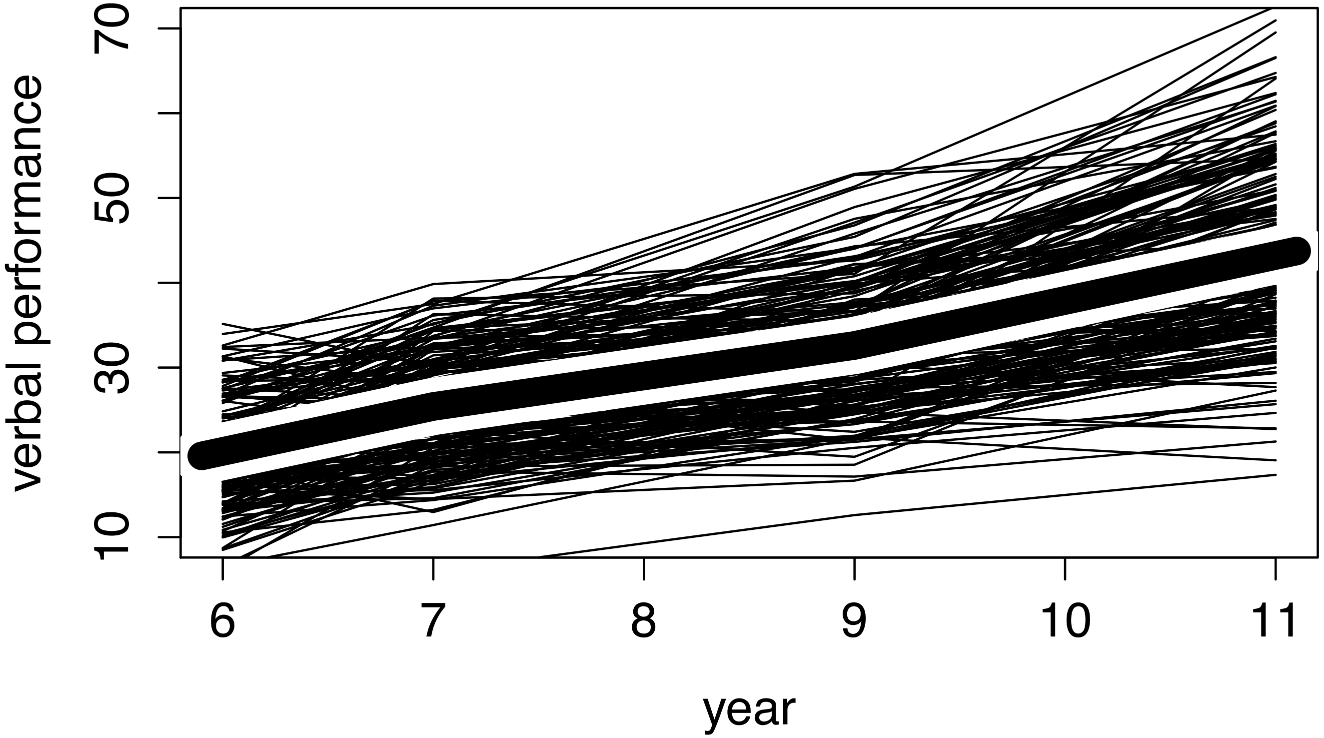
A Simple Example: Wechsler Intelligence Scale for Children





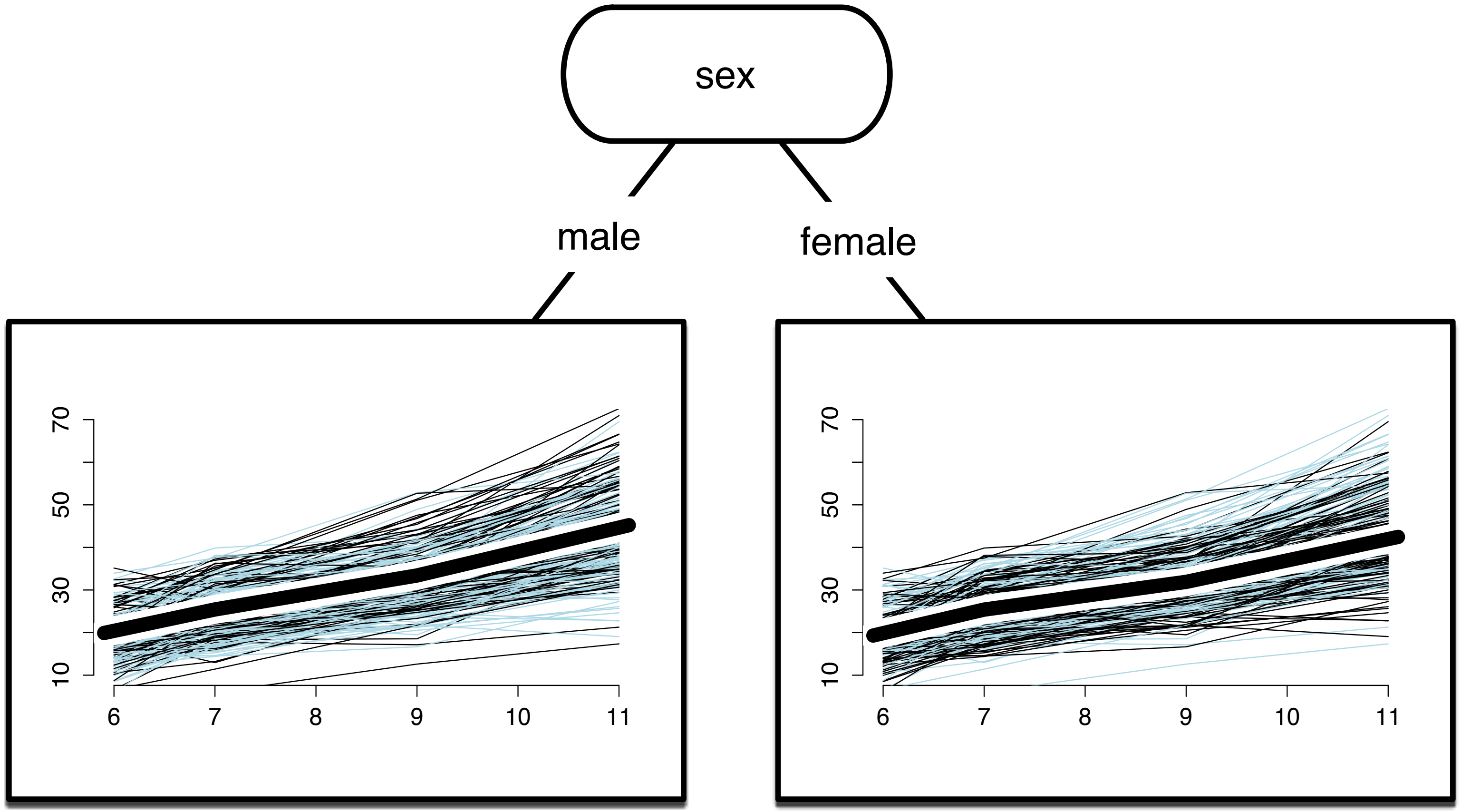
Brandmaier et al., Psychol Methods, 2013

A Simple Example: Wechsler Intelligence Scale for Children



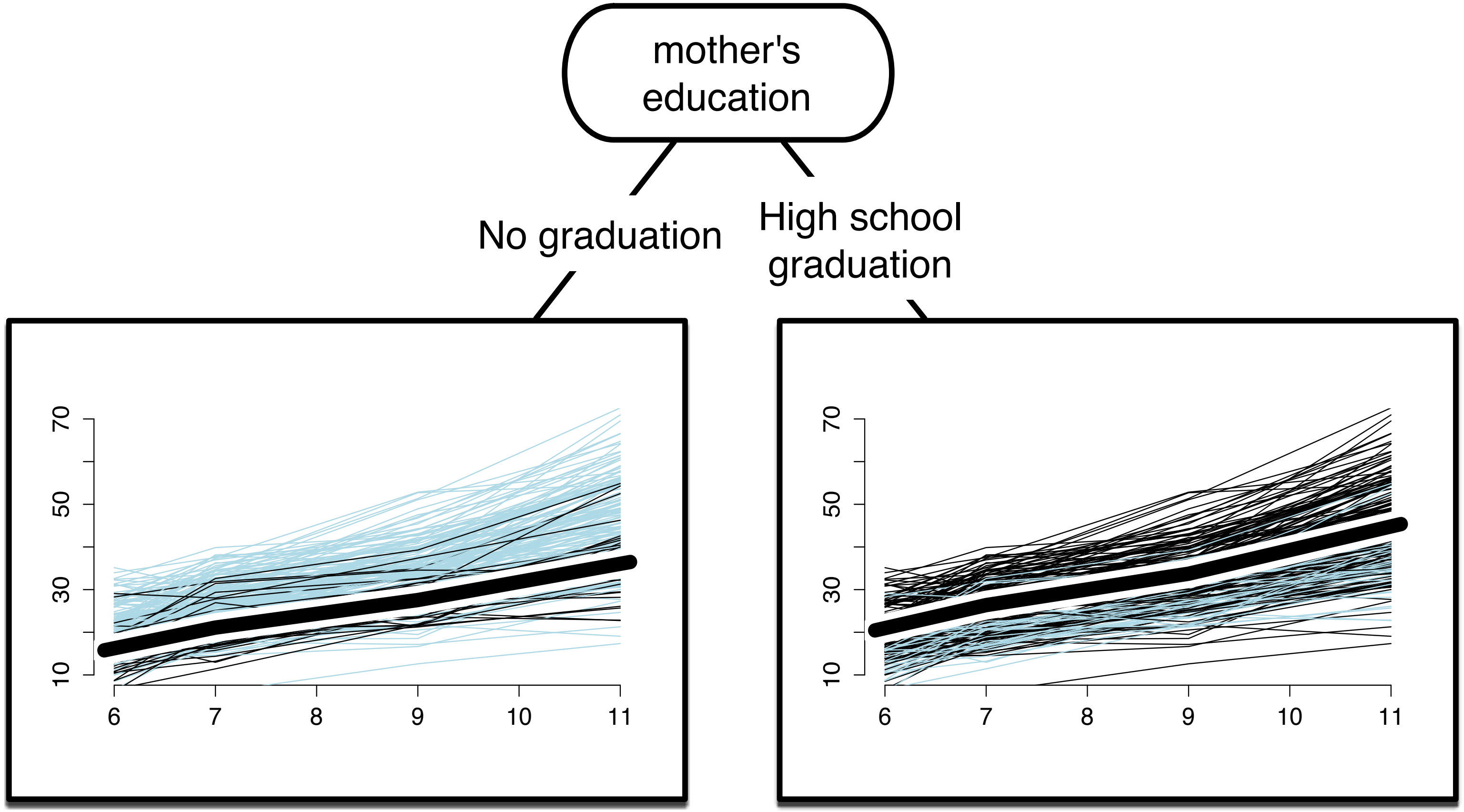
Brandmaier et al., Psychol Methods, 2013

Split Candidate: Sex



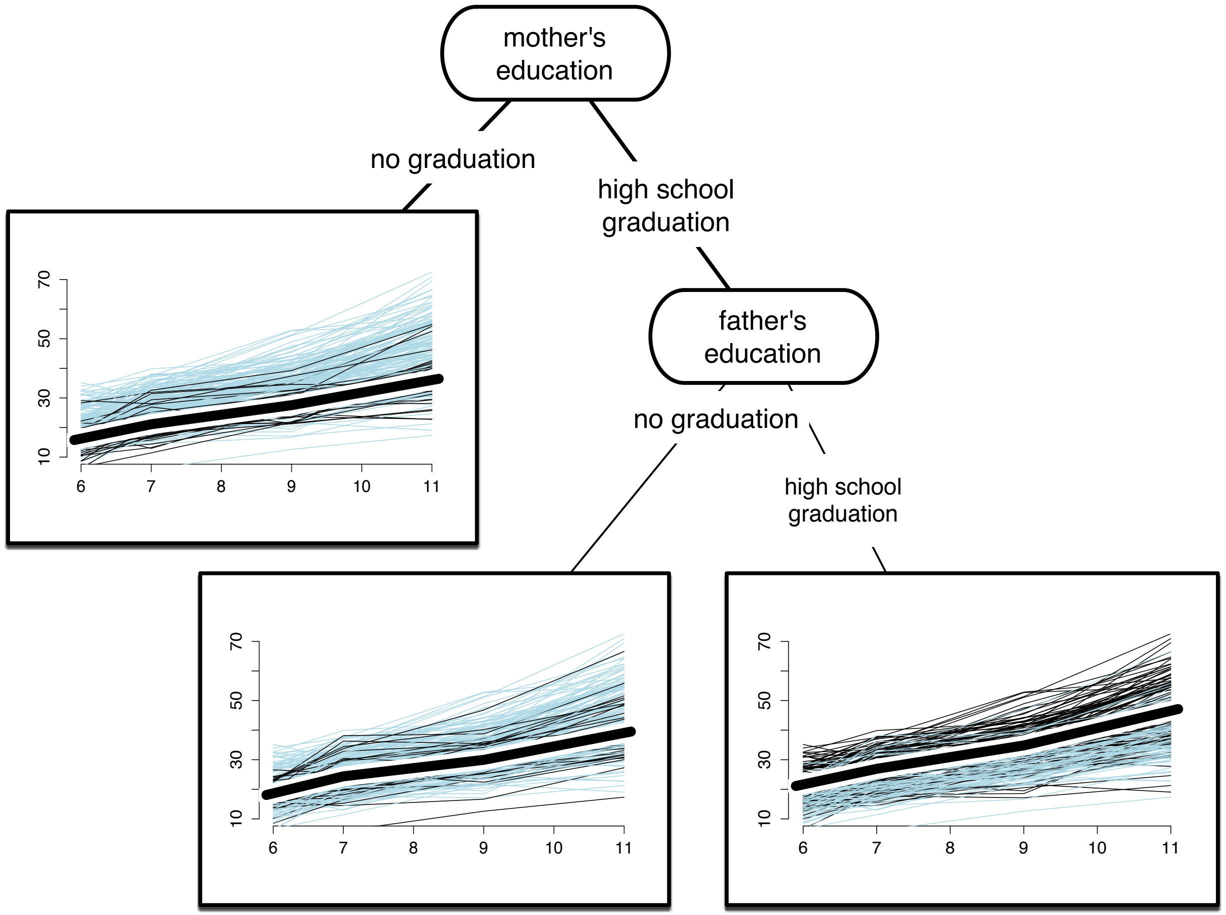
Brandmaier et al., Psychol Methods, 2013

Split Candidate: Mother's Education



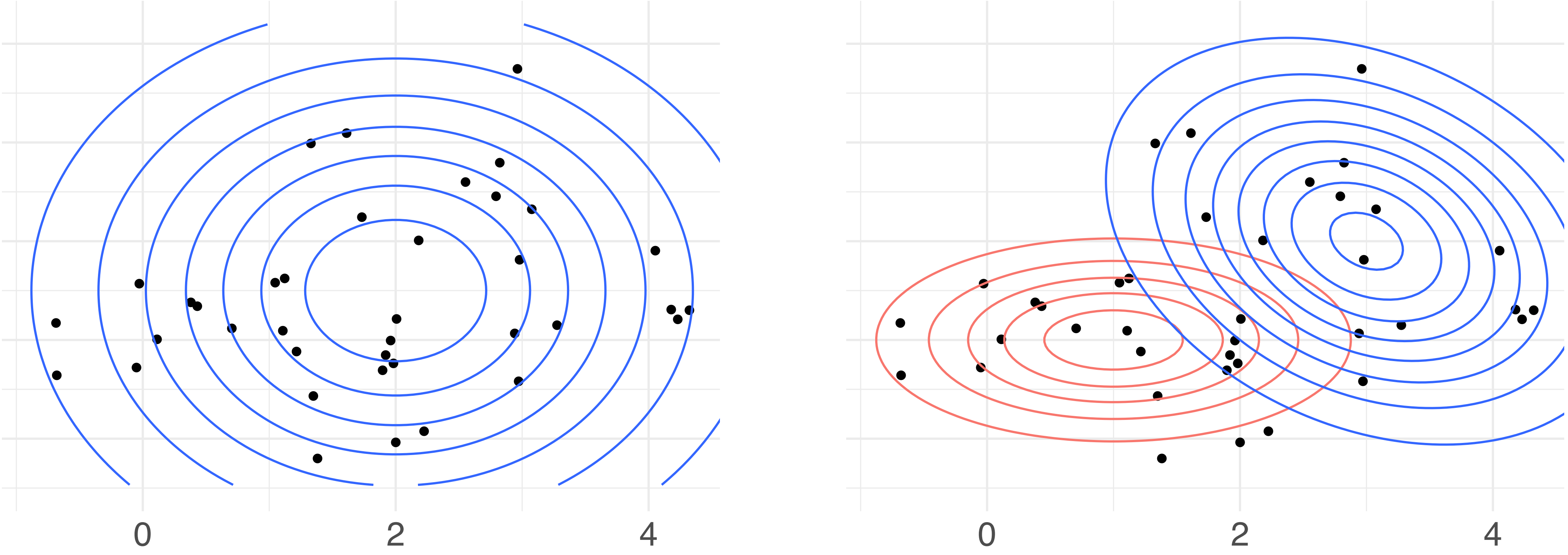
Brandmaier et al., Psychol Methods, 2013

Two-Level Tree



Brandmaier et al., Psychol Methods, 2013

Likelihood Ratio Splitting = Surprise Minimization

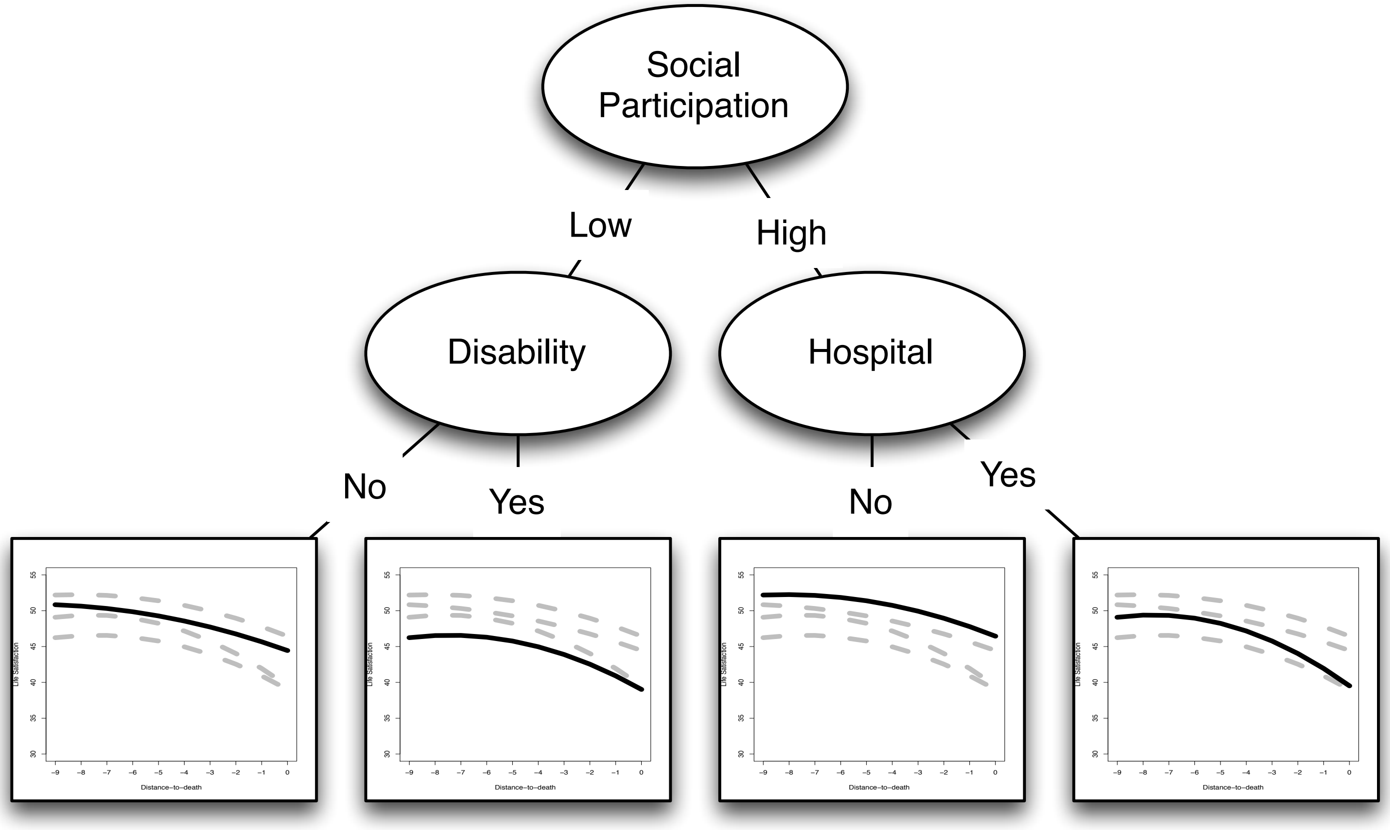


H0: “Split is uninformative == Information gain is zero == No reduction in surprise”

Example: Terminal Decline in Well-being using SOEP

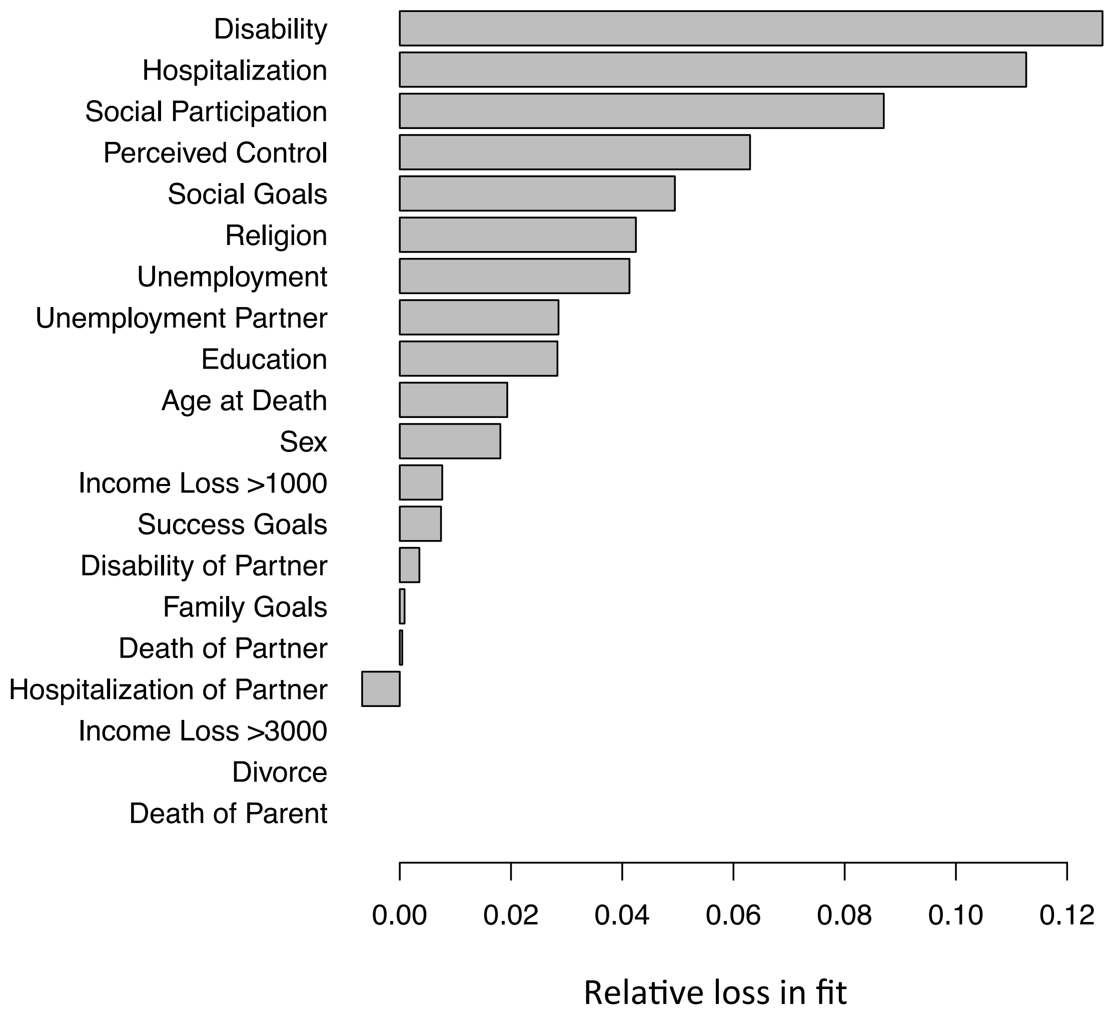
- 4,404 now-deceased participants of the nationwide German SOEP (age at death: M = 73.2 years; 17-102 years; SD = 14.3 years; 52% women)
- Terminal decline, all available observations obtained in the last 10 years of life realigned along a time-to-death time metric
- Outcome: “How satisfied are you currently with your life, all things considered?”, 11-point scale
- Predictors: socio-demographic (e.g., age at death, education, religion), health and burden (e.g., disability, unemployment, divorce), psychosocial (e.g., social participation, perceived control, life goals).

First Two Levels of the Well-Being Tree



Brandmaier et al., Dev. Psychol., 2017

Variable Importance in Well-Being



Summary

SEM Trees and Forests

- combine **model-based and data-driven** modelling
- are tools to **recursively** identify **sub groups** and their predictors in the data given a **model**
- explain **heterogeneity** in a sample
- by **reducing surprise** (maximizing information gain)
- potentially discover differences both on the **construct level** and on the **measurement level**

Caveats

Prediction ≠ Explanation

- No short-cut from data to theory or knowledge
- The model with **best predictions** may **not** be the **true** model
- Shmueli et al. (2010): **parsimonious** but less „true“ model can **have a higher predictive validi**ty than a „truer“ but more complex model, particularly when
  - Data are noisy
  - When the true effects of the left-out variables are small
  - Sample size is small

Outlook

SEM Trees and Forests as a hybrid of two modeling cultures allows us:

- **Challenge established models** when comparing predictive accuracy (**hold out set or cross-validation!**).
- Tree/forest may lead to a **revision of the substantial theory** and the formulation of a new parametric model and/or experiment
- Conclusion that **postulated model applies only to a limited range of subjects**