Getting in Between People and Their Measurement

Psychometric Evaluation of (Intensive) Longitudinal Data

Tilburg University

Understanding Society

Leonie Vogelsmeier
&
Joran Jongerling

Department of Methodology & Statistics
Tilburg University

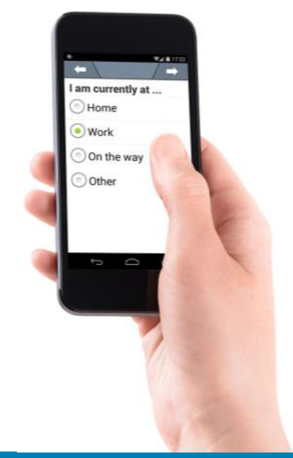
☛ Introduction to ILD

Experience Sampling

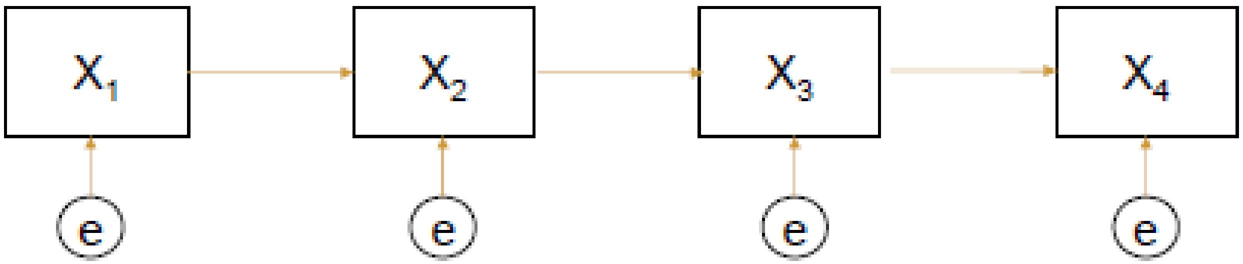
- Experience Sampling Methodology (ESM)
  - (or Ambulatory Assessment, Ecological Momentary Assessment)
- Many subjects
- Repeatedly fill in questionnaires (> 50 x)

- Investigate dynamics of psychological constructs/factors
  in daily life

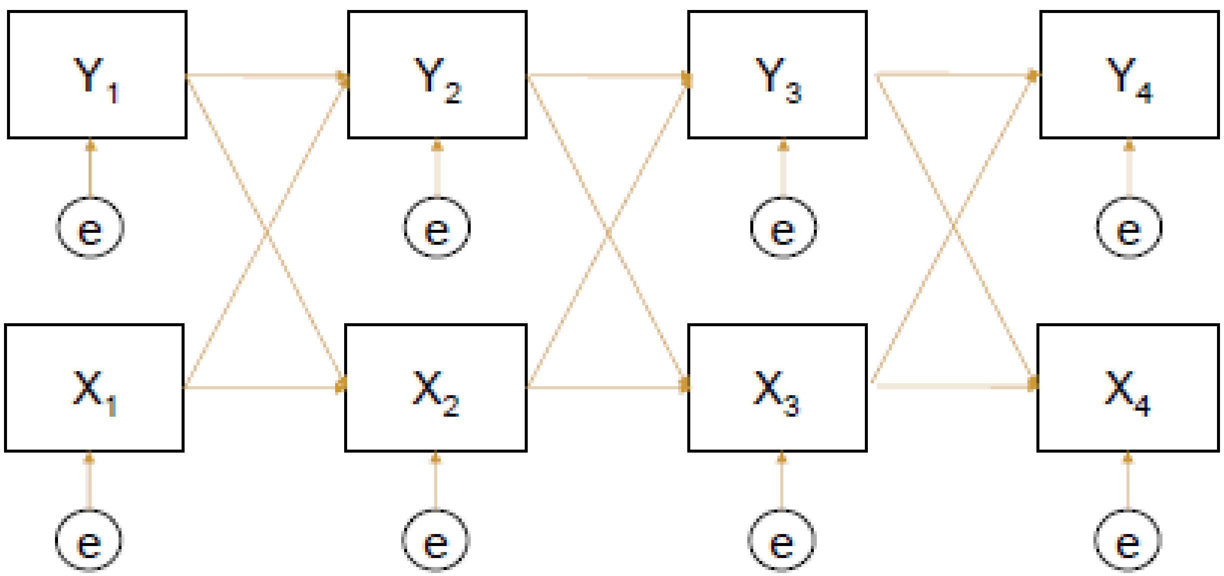
→ May be used, e.g., for personalized therapy



Experience Sampling



AR

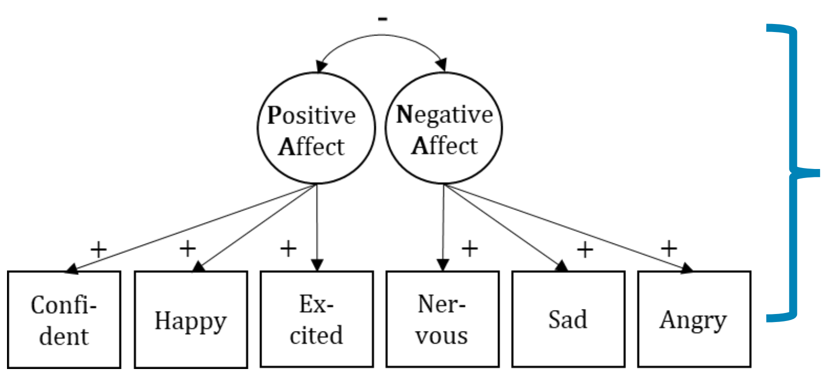


VAR

→ The use of ESM designs tremendously increased in the last years…

Experience Sampling

- Psychological constructs are usually latent (unobserved)
- Assessment via self-report measures in the form of questionnaires



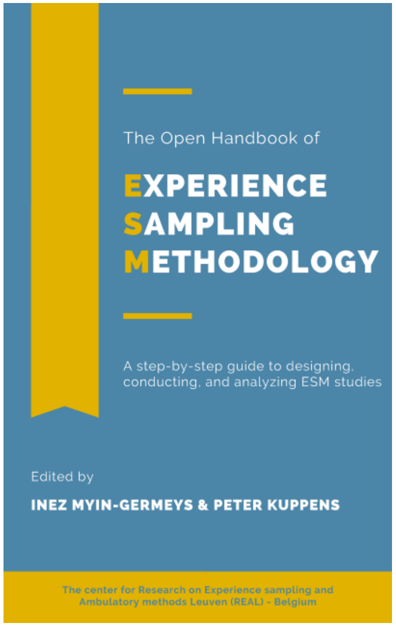
Measurement model

- We have to worry about the same things we always have to worry about with questionnaires, and it is more challenging.
- Generally, measurement models can be evaluated with (different types of) factor analysis

  $$\boldsymbol{y}_i = \tau + \boldsymbol{\Lambda}\,\boldsymbol{\eta}_i + \epsilon_i$$

- The measurement model(s) for an ESM study at hand can differ compared to cross-sectional studies and other ESM studies!

Experience Sampling



- Chapter 4 from the "Free ESM Handbook" gives guidelines on how to construct questionnaires for ESM studies (think of questionnaire length, wording, response scale, order of questions…).
- In this lecture, we'll focus on how to evaluate your measurements after the data has been collected.
- Per topic, we'll start by recapping some of the basics of proper measurement in cross-sectional and panel data.
- We'll then extend the insights from "traditional" data types to ILD and will highlight what additional steps and checks you need to do with this new type of data.

Outline

Part 1 ~14:05-14:50

- Introduction to Different Data Types
- Factor Models
- Invariance

Break ~14:50-15:00

Part 2 ~15:00-15:50

- Discussion
- Reliability

Break~15:50-16:00

Part 3 ~16:00-17:00

- Lab
- If there is time: Exploring Non-Invariance for ILD + Discussion

Introduction to Different Data Types

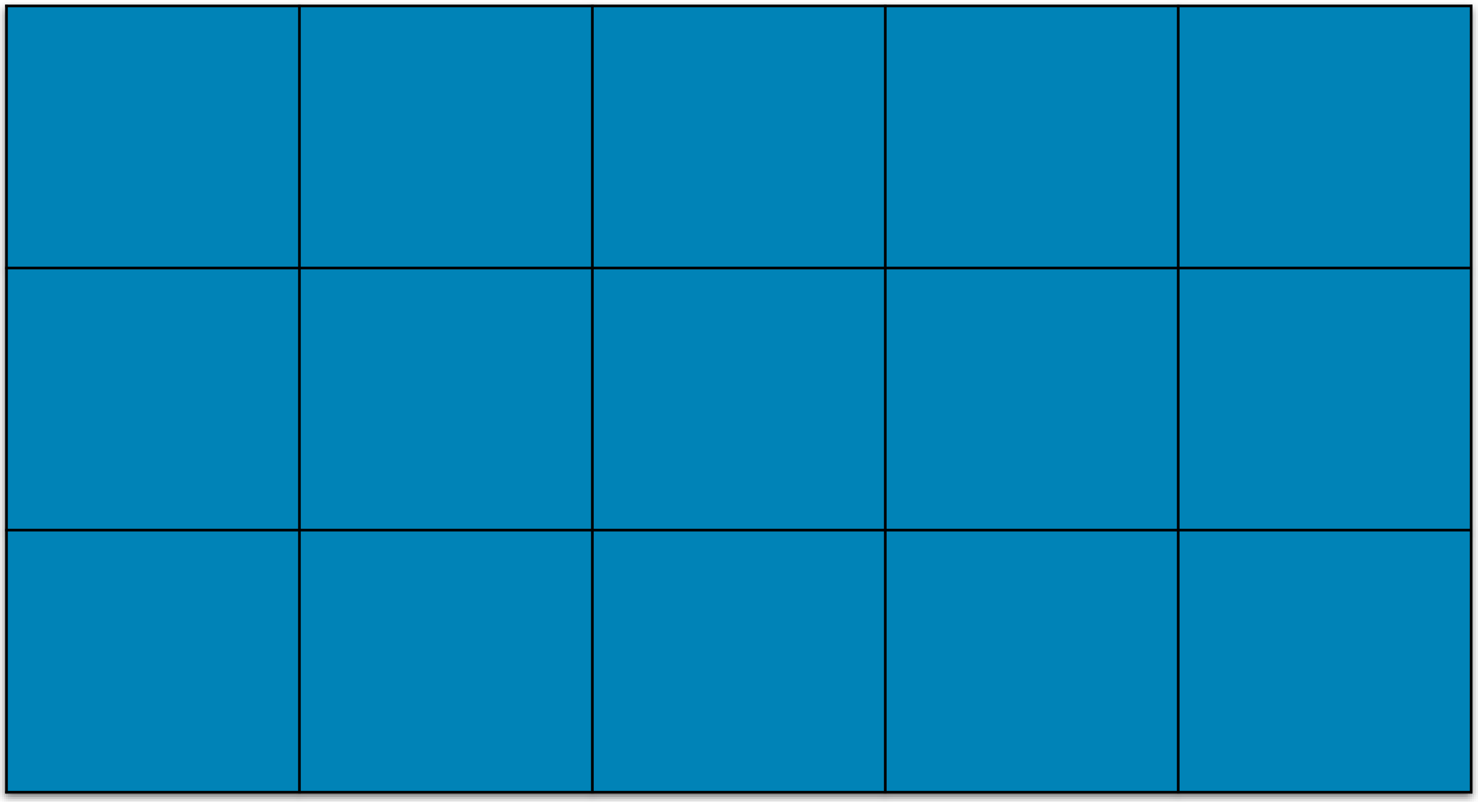
Introduction to Different Data Types

- Questionnaires applied to three types of data (mainly)
  - Cross-sectional
  - Panel
  - Intensive Longitudinal

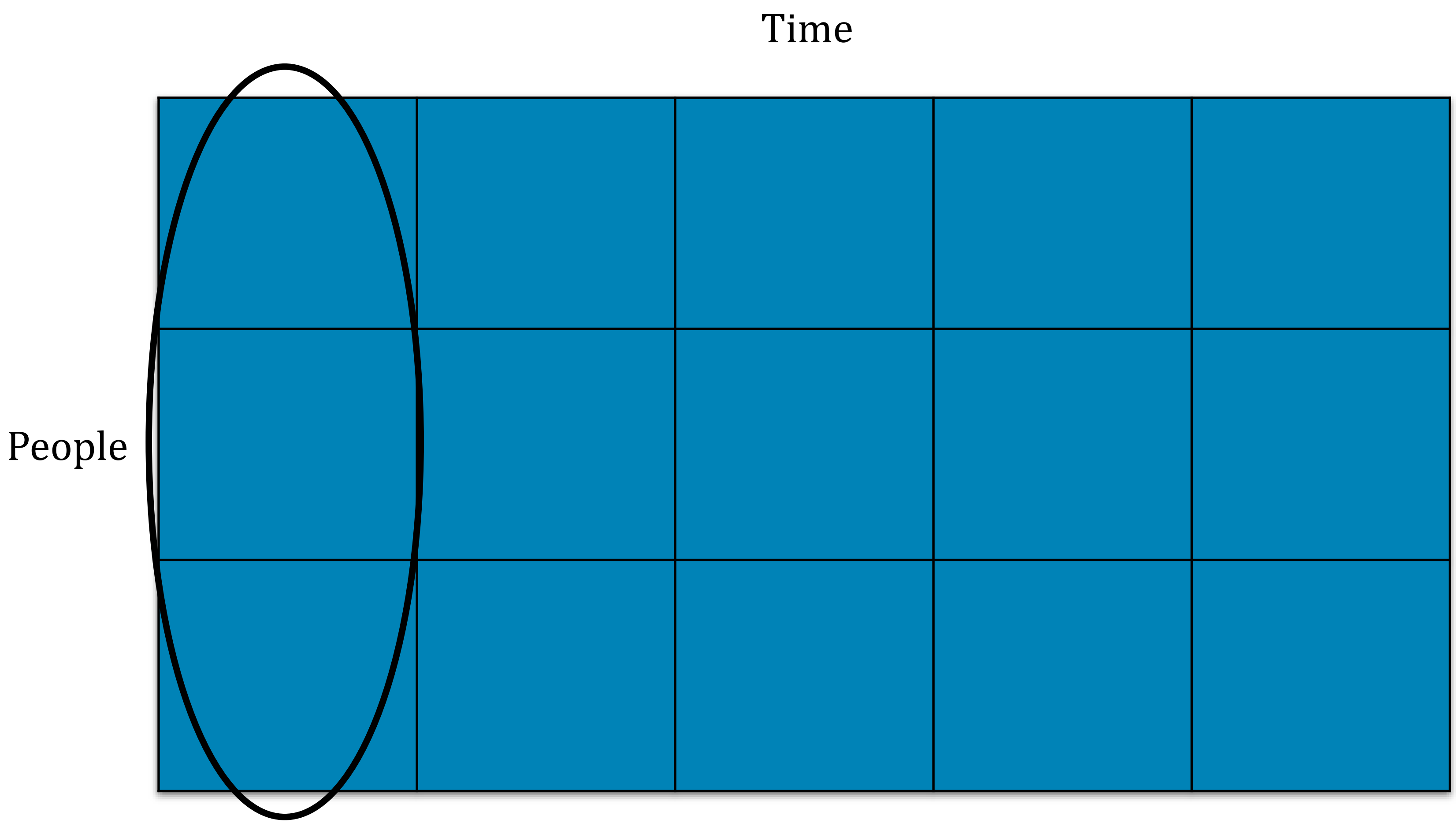
Three Types of Data

Time

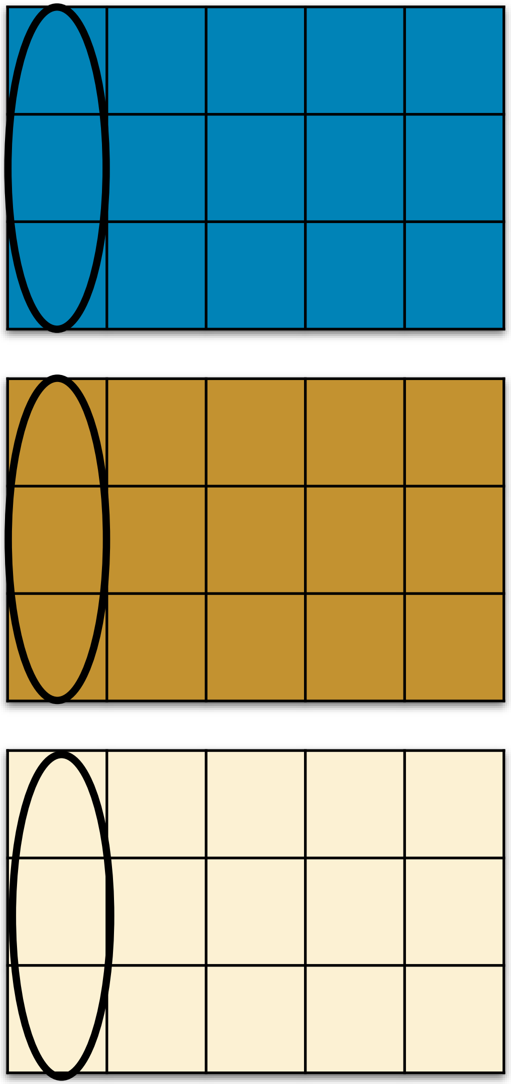
People

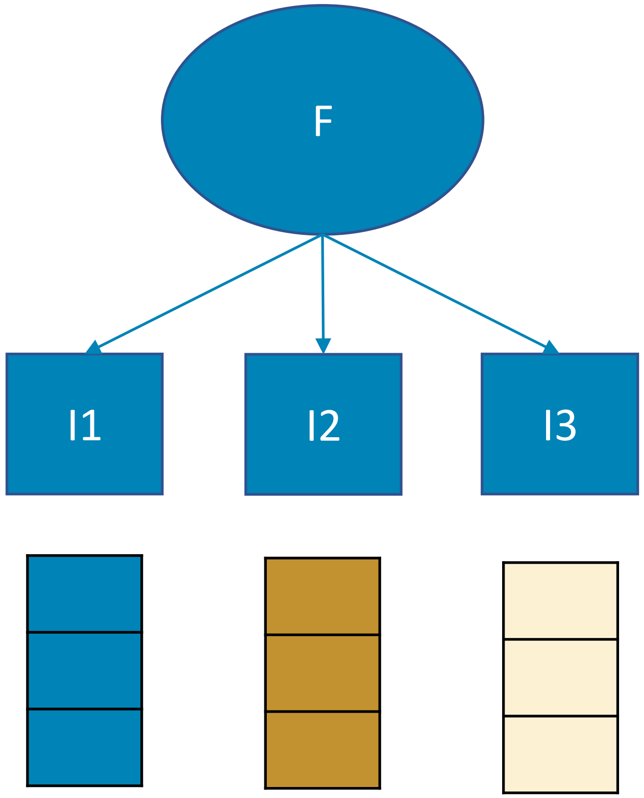


Three Types of Data

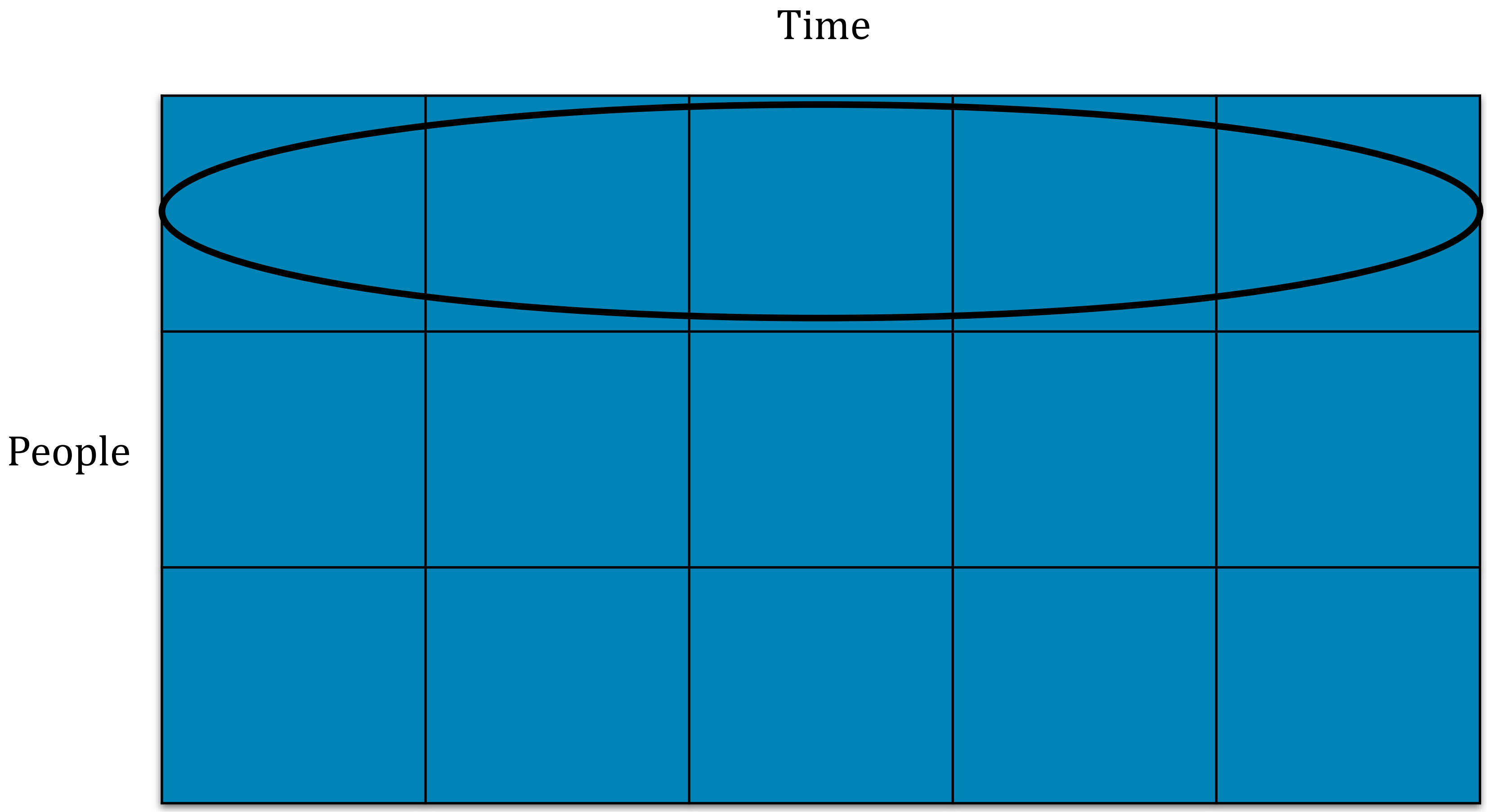


Three Types of Data

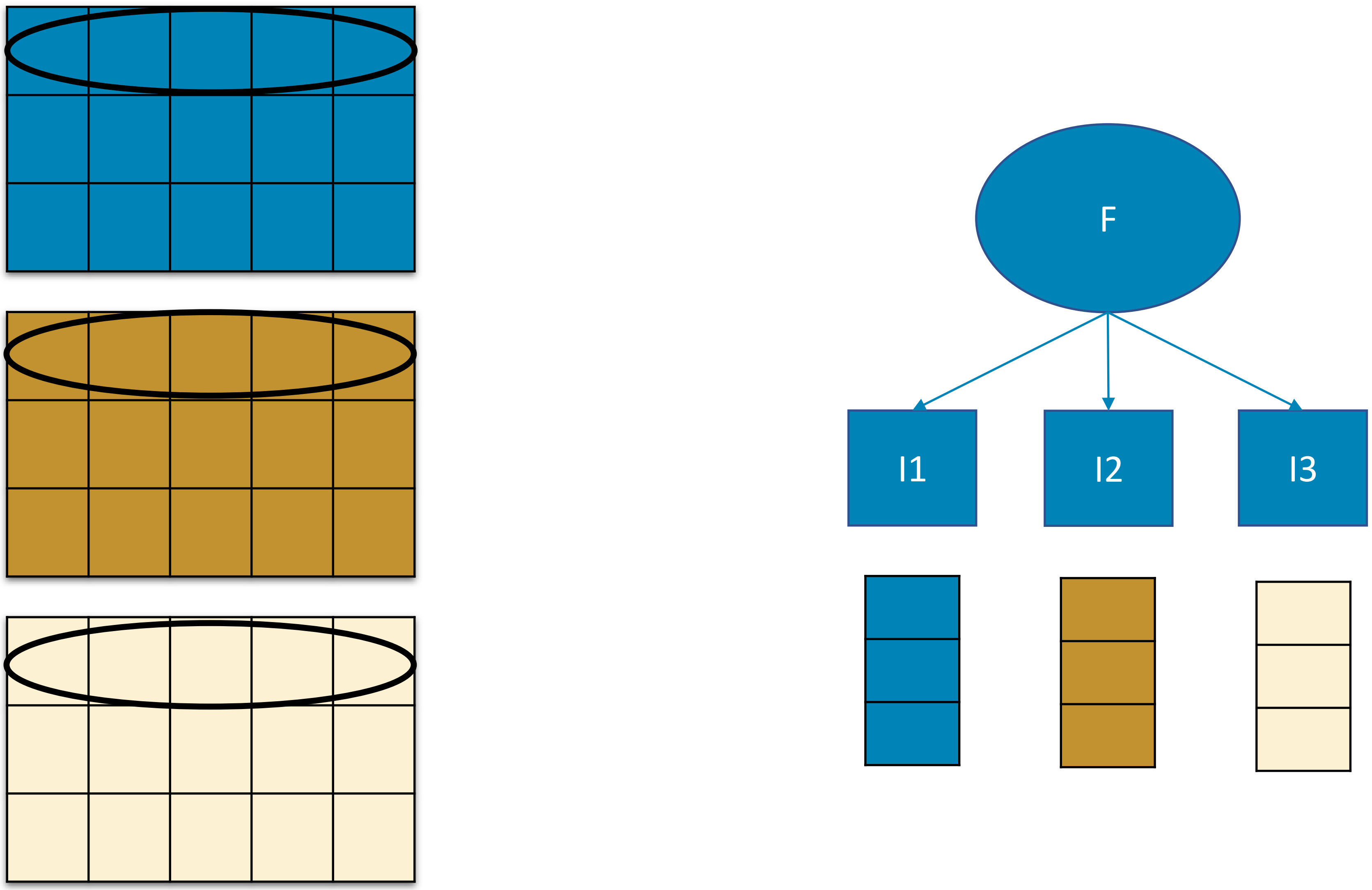




Three Types of Data



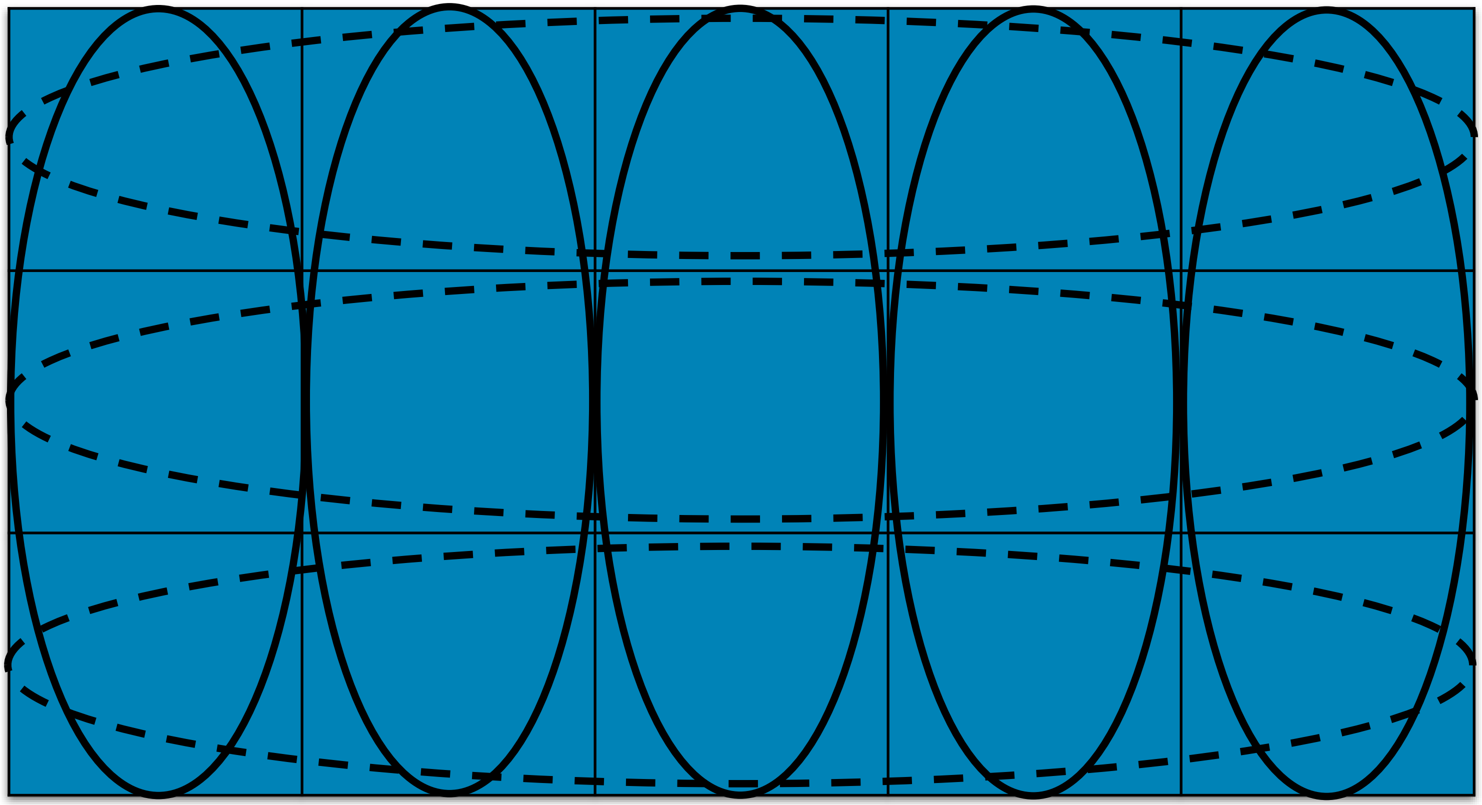
Three Types of Data



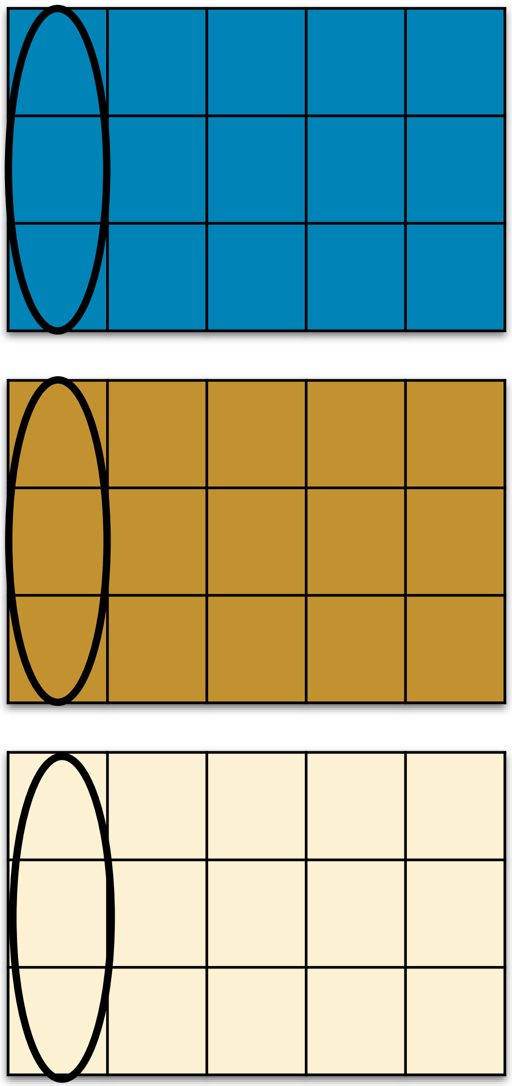
Three Types of Data

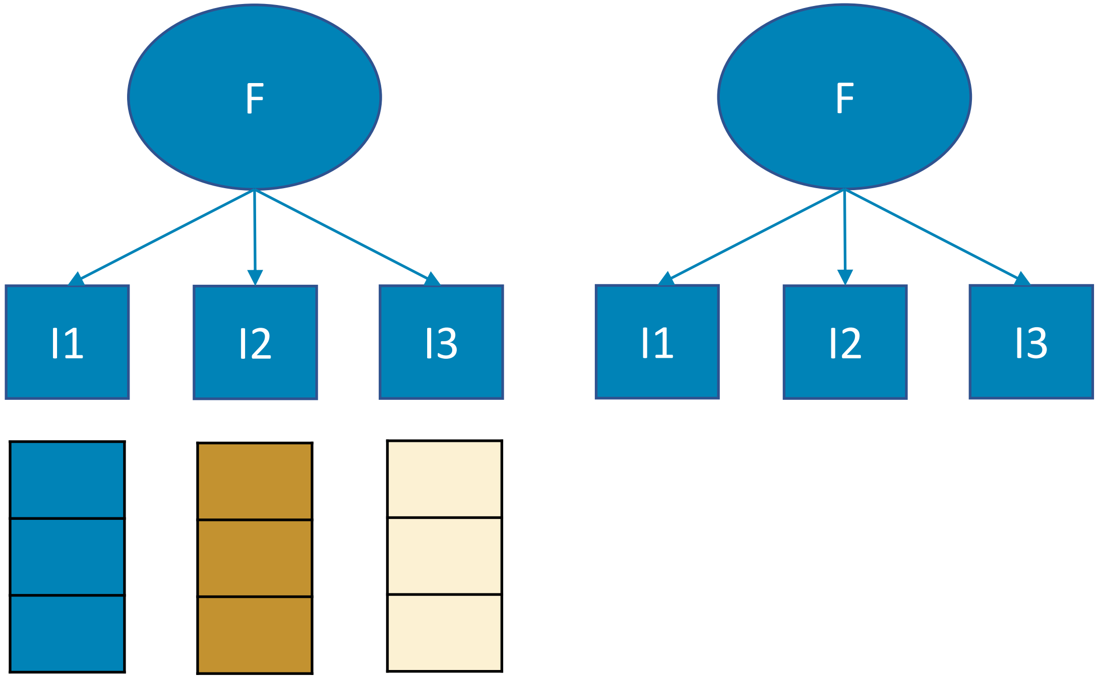
Time

People

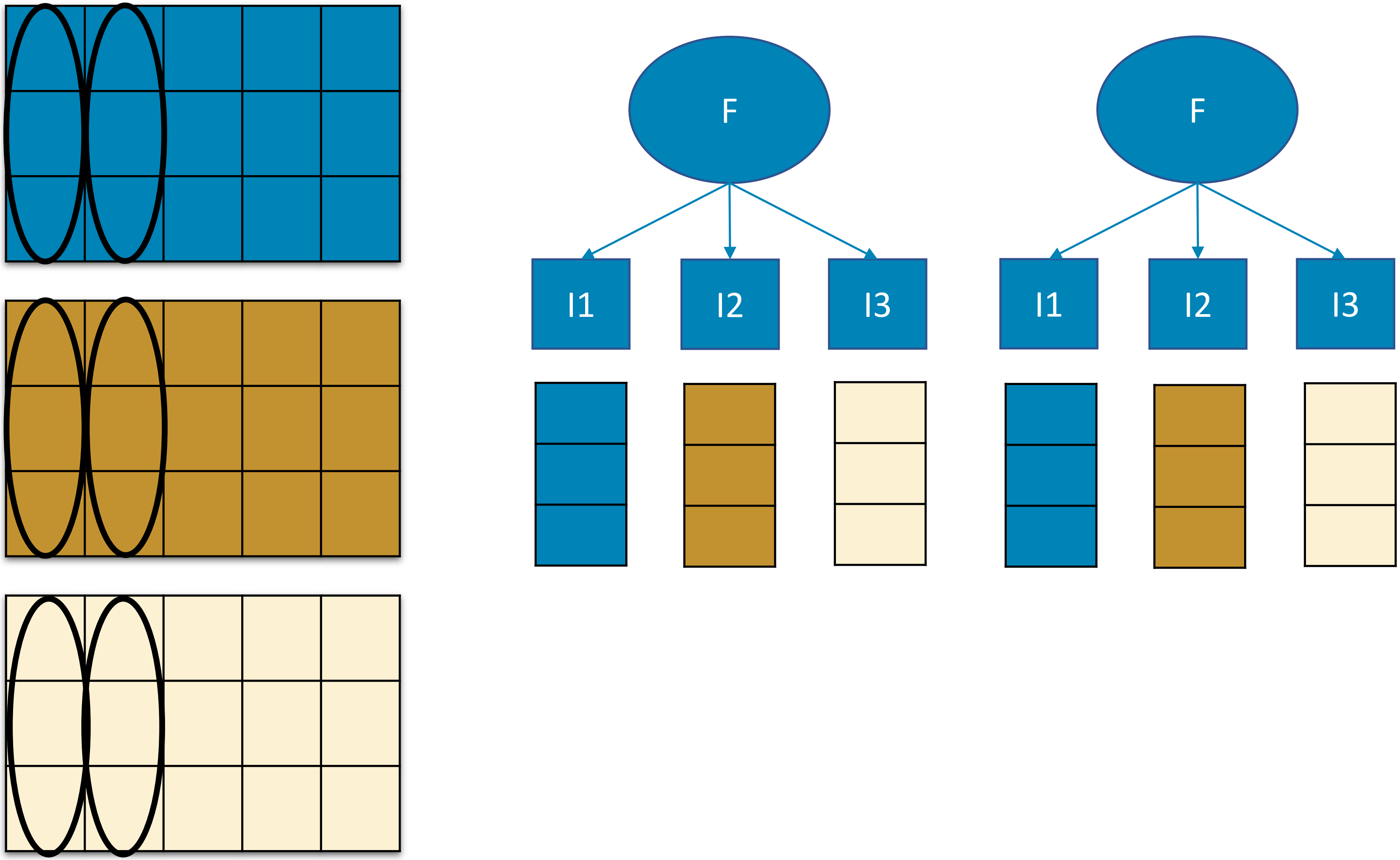


Three Types of Data

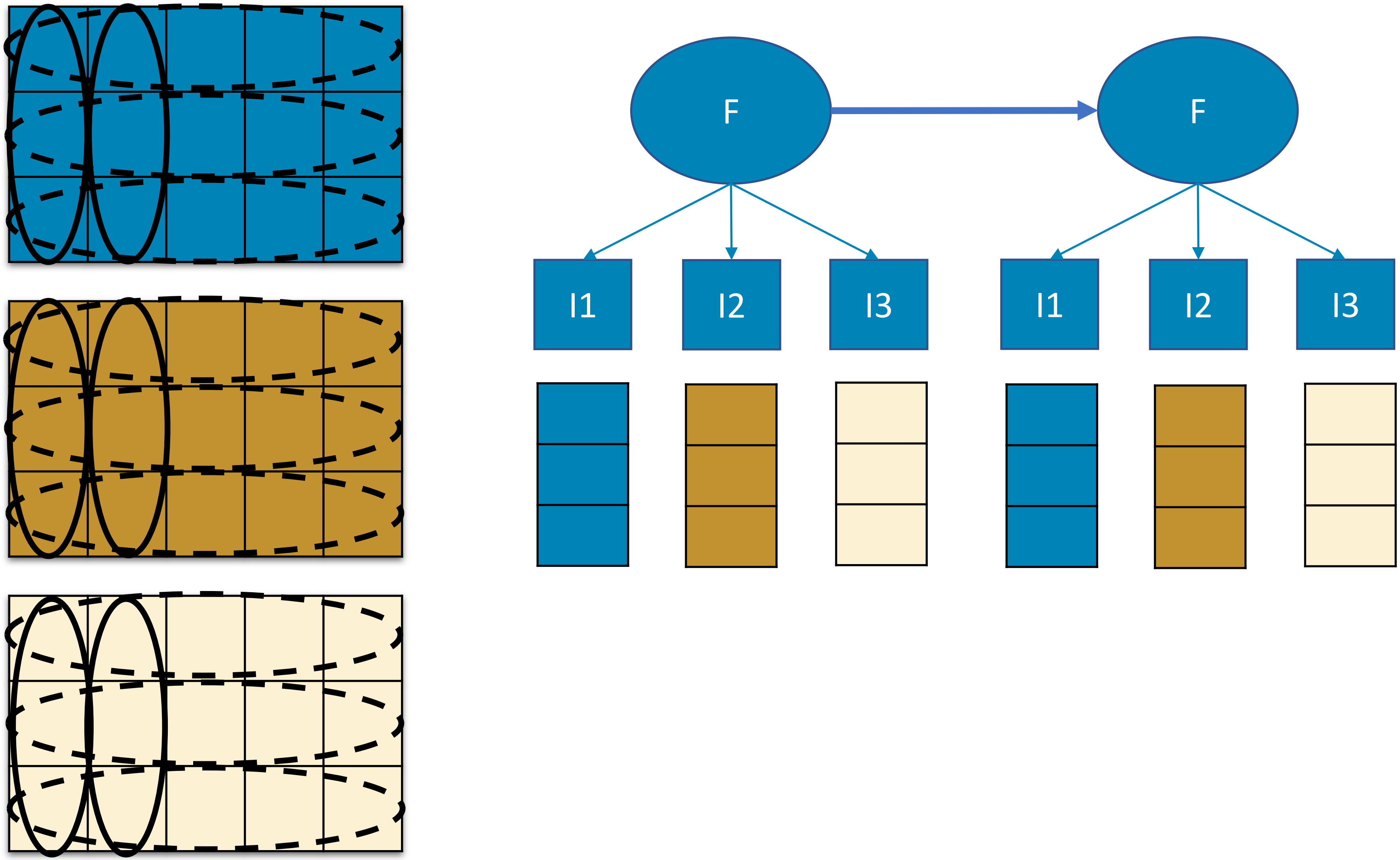




Three Types of Data



Three Types of Data



Introduce Cross-Sectional & Panel Data

To what extent do the following statements apply to you?	1-7
I often check the location of security cameras.	
I count the number of security guards.	
I sense whether someone is an easy target.	
I enjoy getting others in trouble.	
I enjoy other people's pain.	
I dream about stealing money.	
I do not sympathize with others.	
I root for the bad guy in movies.	
I have been told that I am not a good roommate.	
I am more important than others.	
I am bored by other people.	
I often walk around naked outside.	
I often bully/bullied my siblings.	
I don't have many friends.	
Money is more important than relationships.	
I would hurt others to get what I want.	

Introduce ILD

Question	0-100
How much does the following apply to you: I can do anything	
How much does the following apply to you: I don’t fail	
How much does the following apply to you: If something bad happens, it's someone else’s fault	
How much does the following apply to you: I'm smarter than other people	
How much does the following apply to you: I avoid making the mistakes other people make	
How much does the following apply to you: I would definitely hire myself	
How much does the following apply to you: I'm better looking than others	

Psychometrics

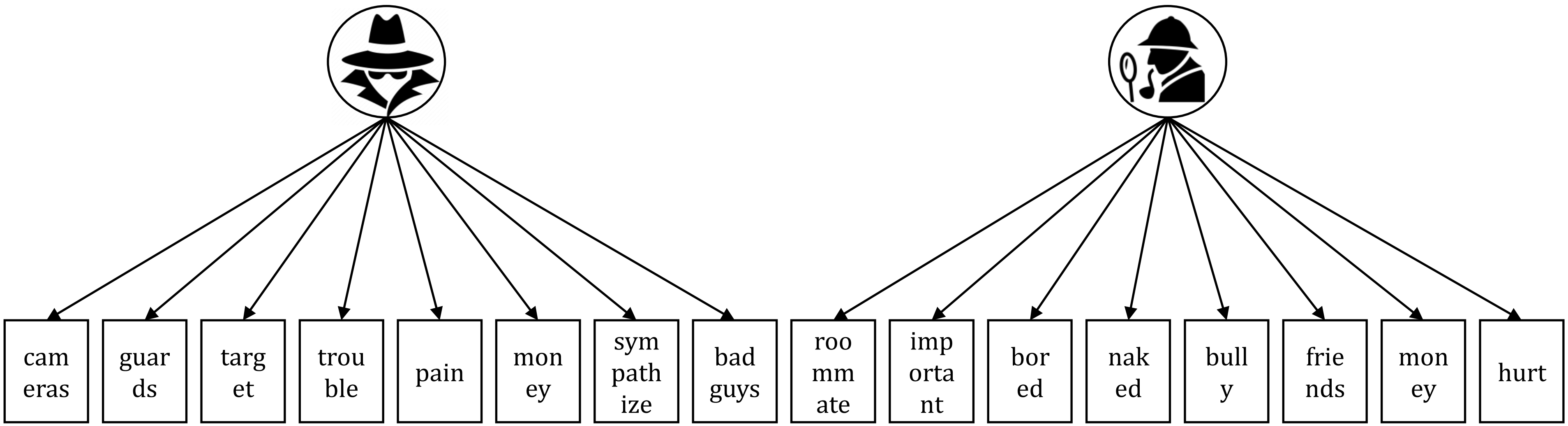
- Factor model
    - Do items measure what they are expected to measure?
    - It allows distinguishing the amount of variance in each item that is due to the common factor and the remaining error variance.

- Invariance
    - Is all systematic variability in item scores attributable to the psychological construct of interest?
    - Or is there variability due to group membership?
        - Does the factor model differ across groups?

- Reliability
    - Determining the reliability of assessing between-person differences.
    - When we have multiple items that measure the same underlying construct, internal consistency can be used as a measure of reliability.

Factor Model

Factor Model



- Allows to investigate whether items that are expected to measure a factor indeed measure the factor.

- Allows distinction between item variance that is due to the common factor and unique item variance.

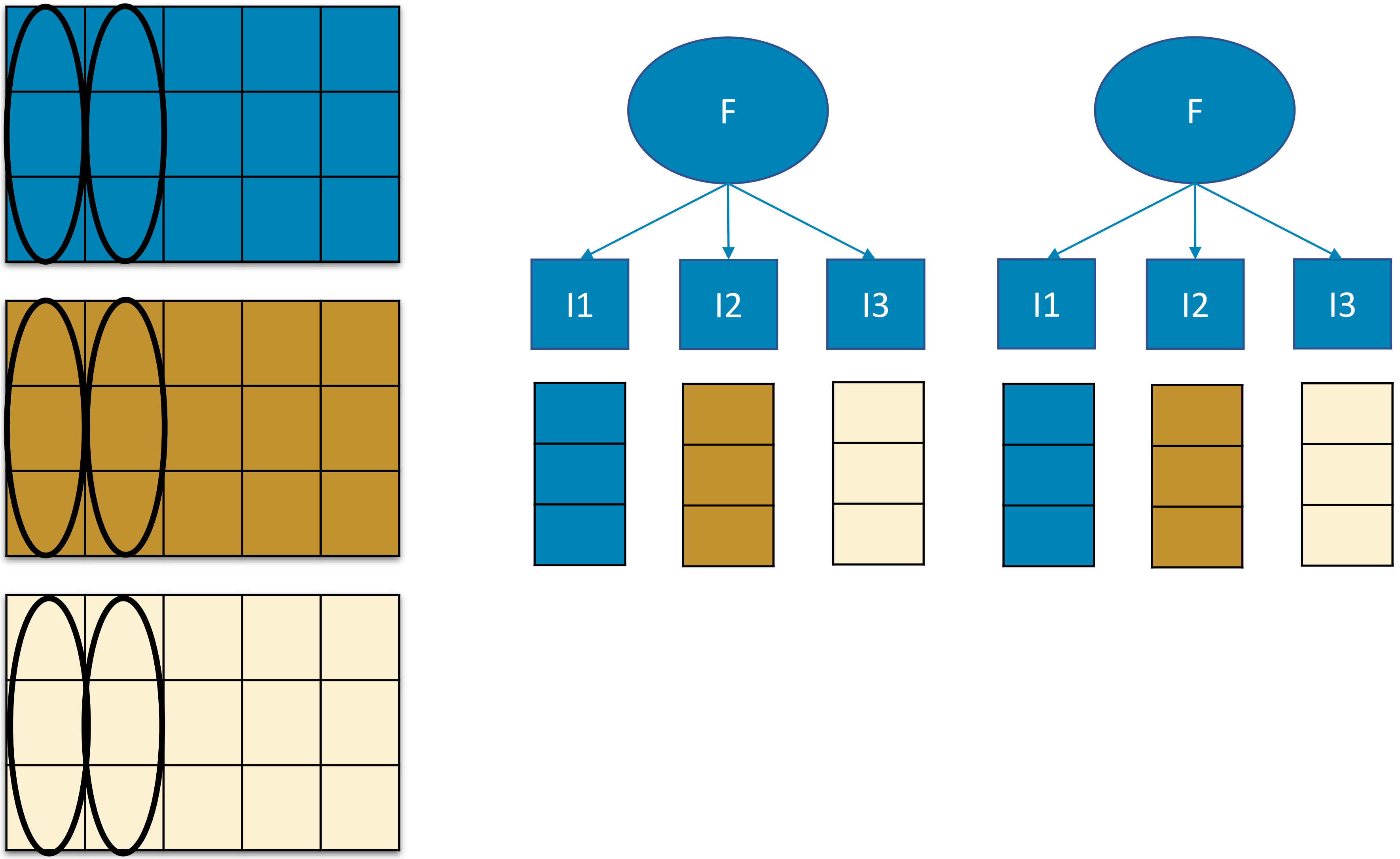
- You can think of factor analysis as regressing item scores on the latent factors. Regression weights are called loadings.

Factor Model Cross-Sectional Data

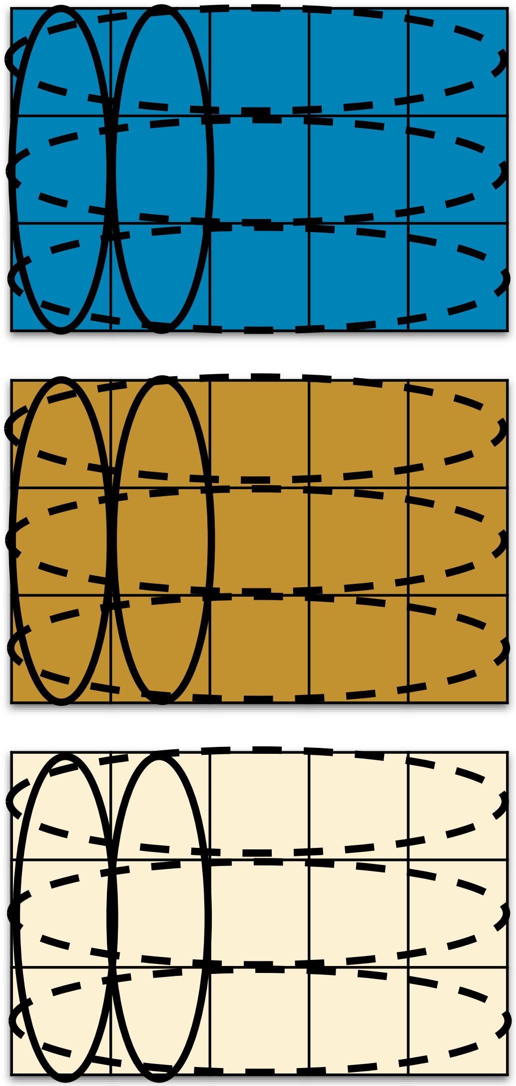
Regression notation:

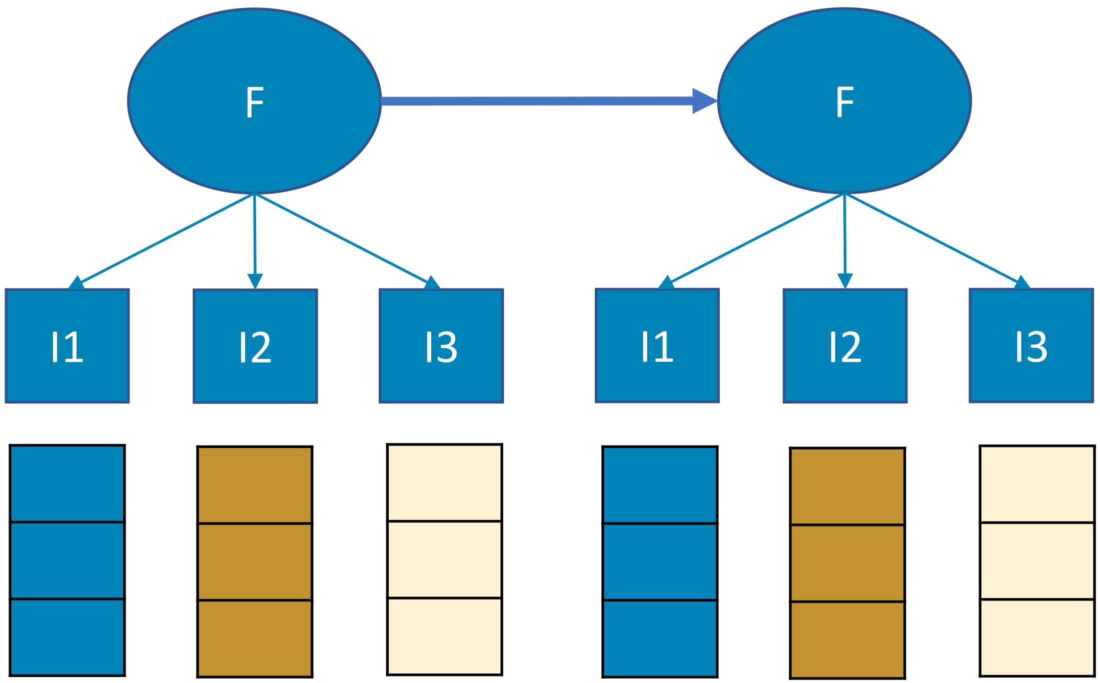
$$\boldsymbol{y}_i = \tau + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \epsilon_i$$

Factor Model for Panel Data



Factor Model for Panel Data



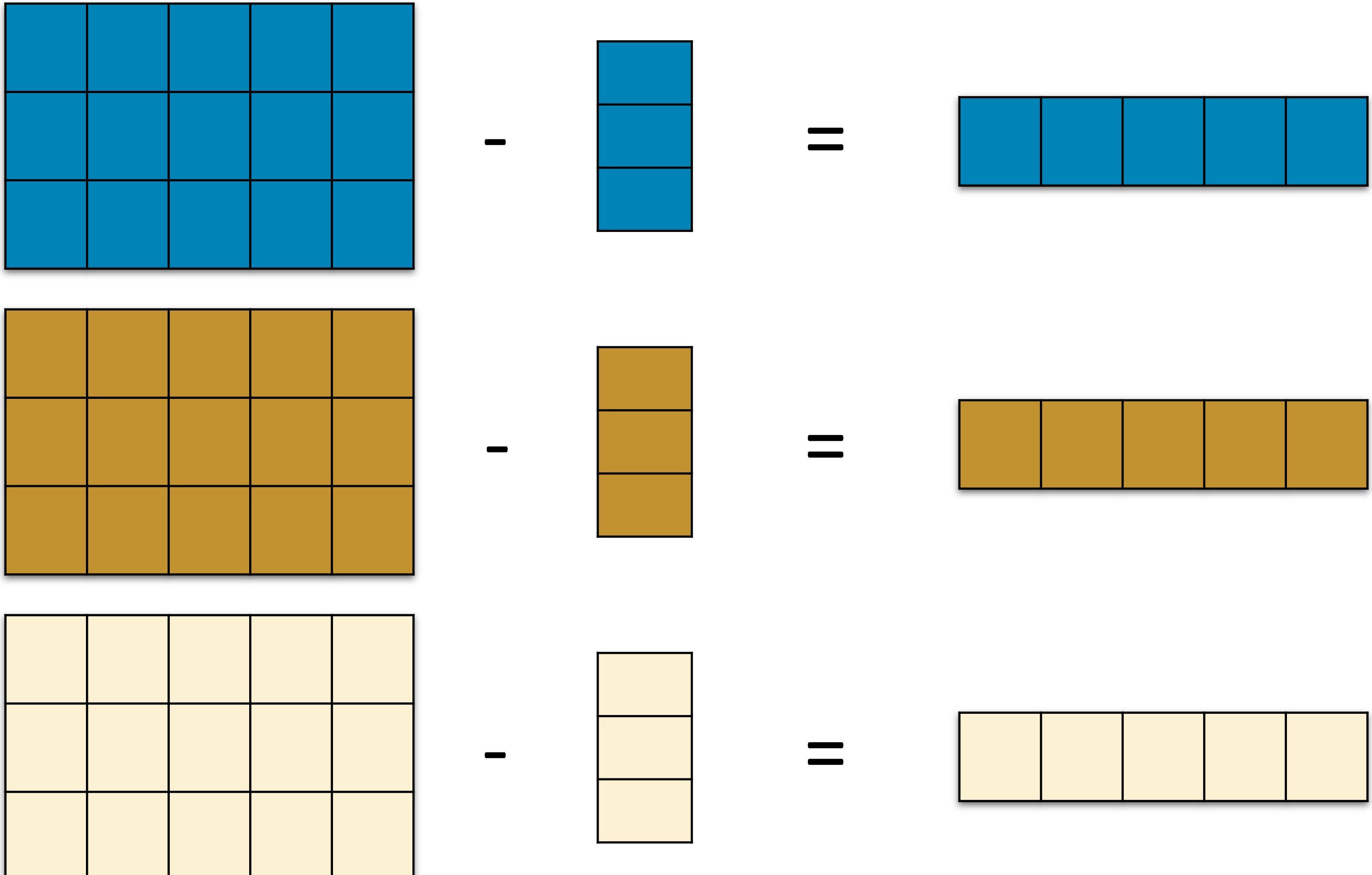


Factor Model Panel Data

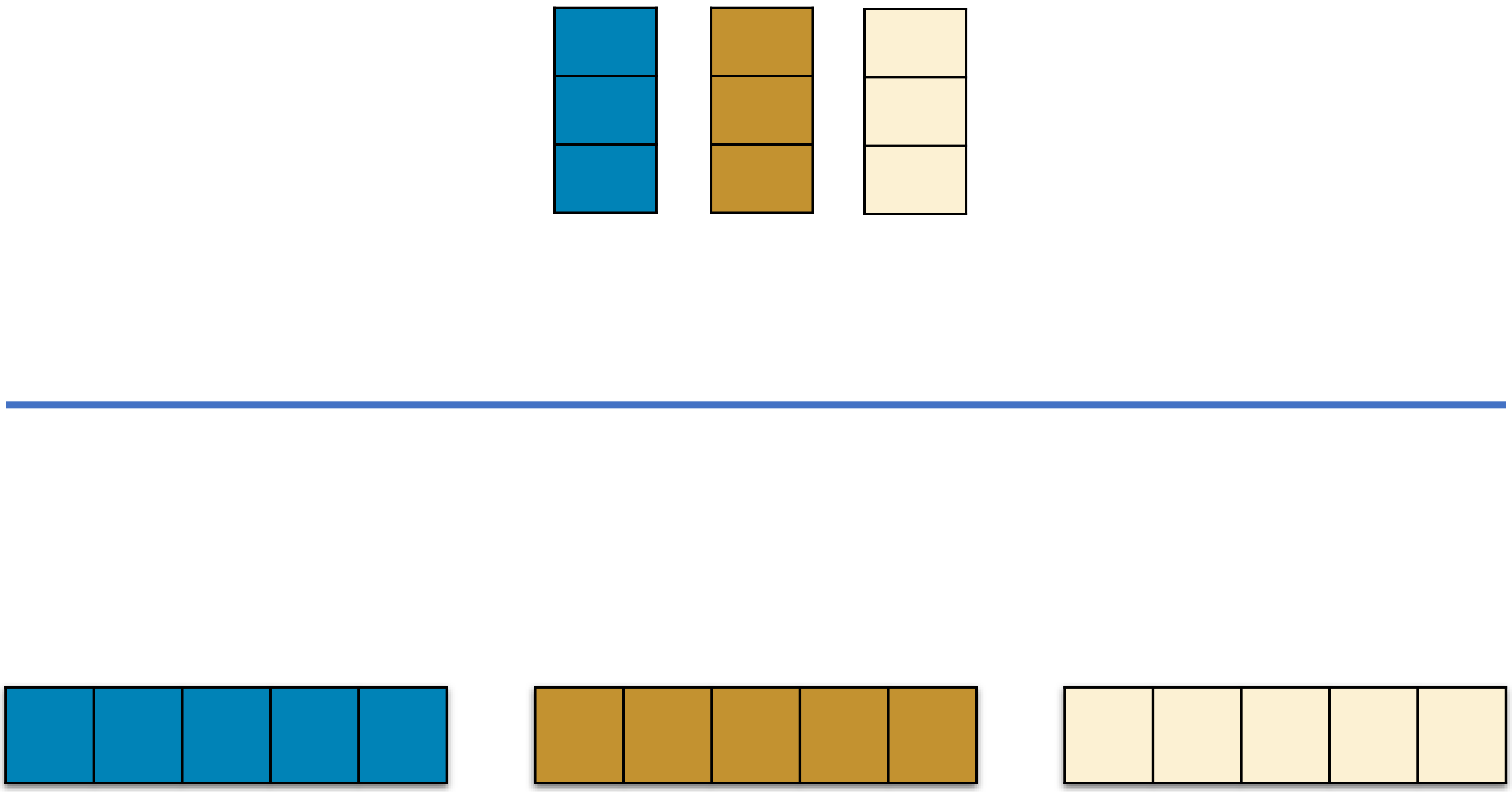
Regression notation:

$$\boldsymbol{y}_{i,t} = \tau + \boldsymbol{\Lambda}\boldsymbol{\eta}_{i,t} + \epsilon_{i,t}$$

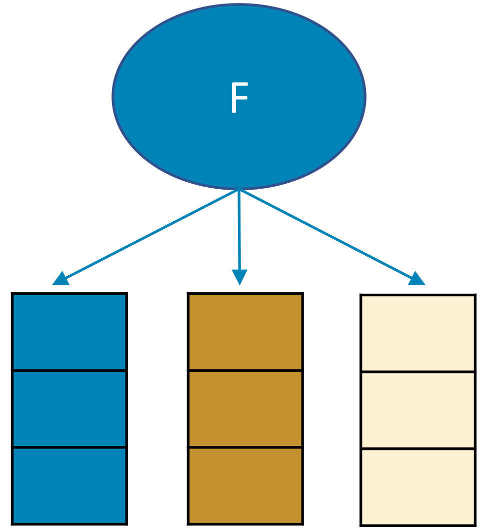
Multilevel

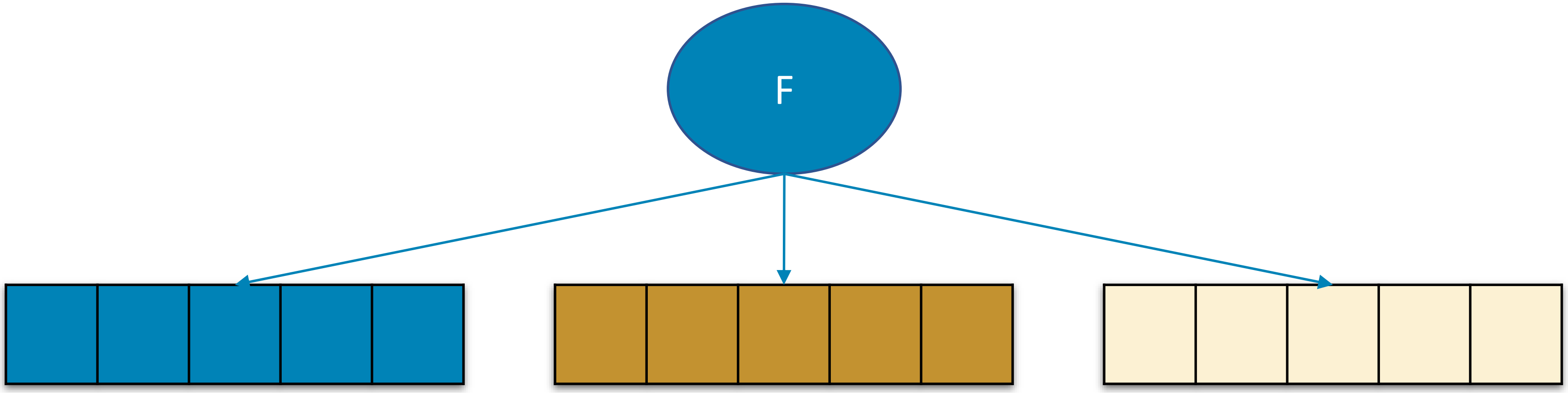


Multilevel

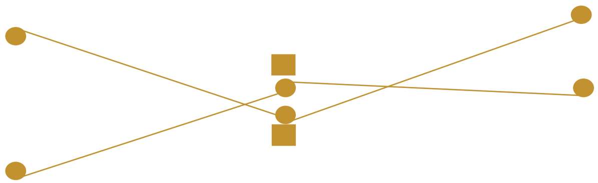


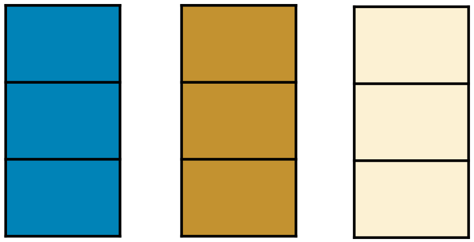
Multilevel

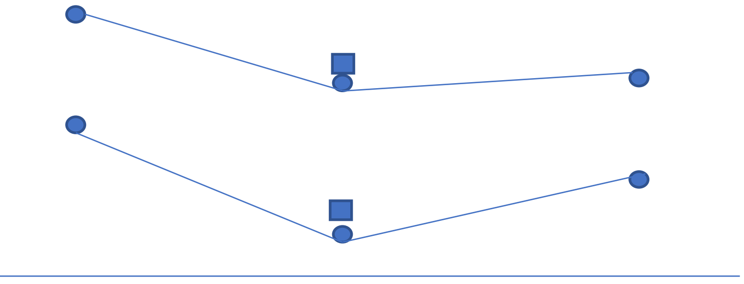


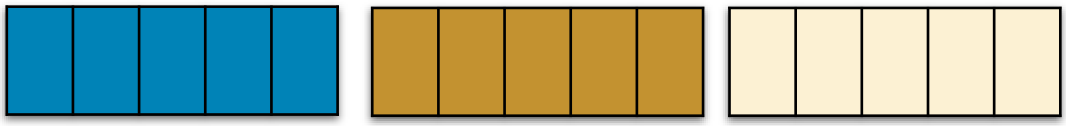


Multilevel



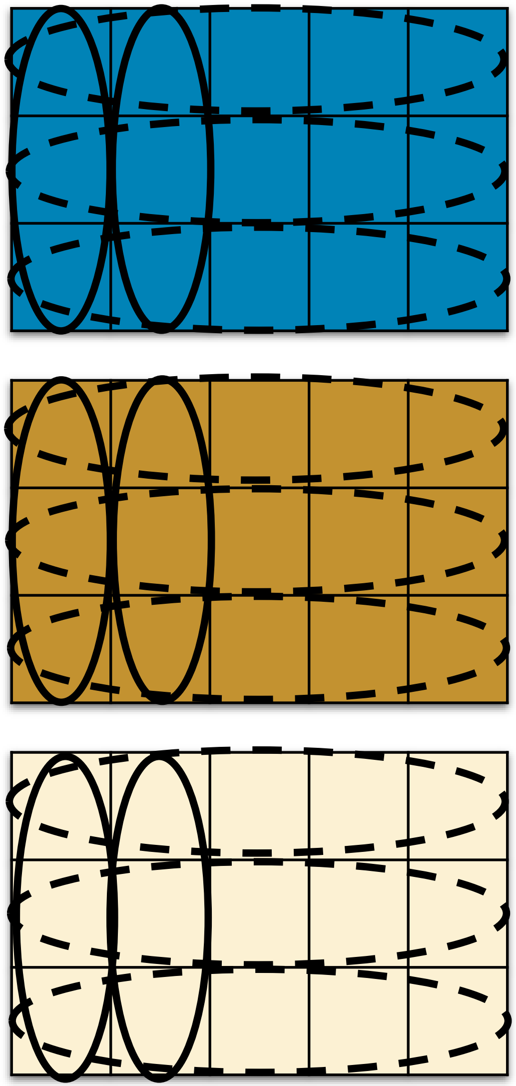


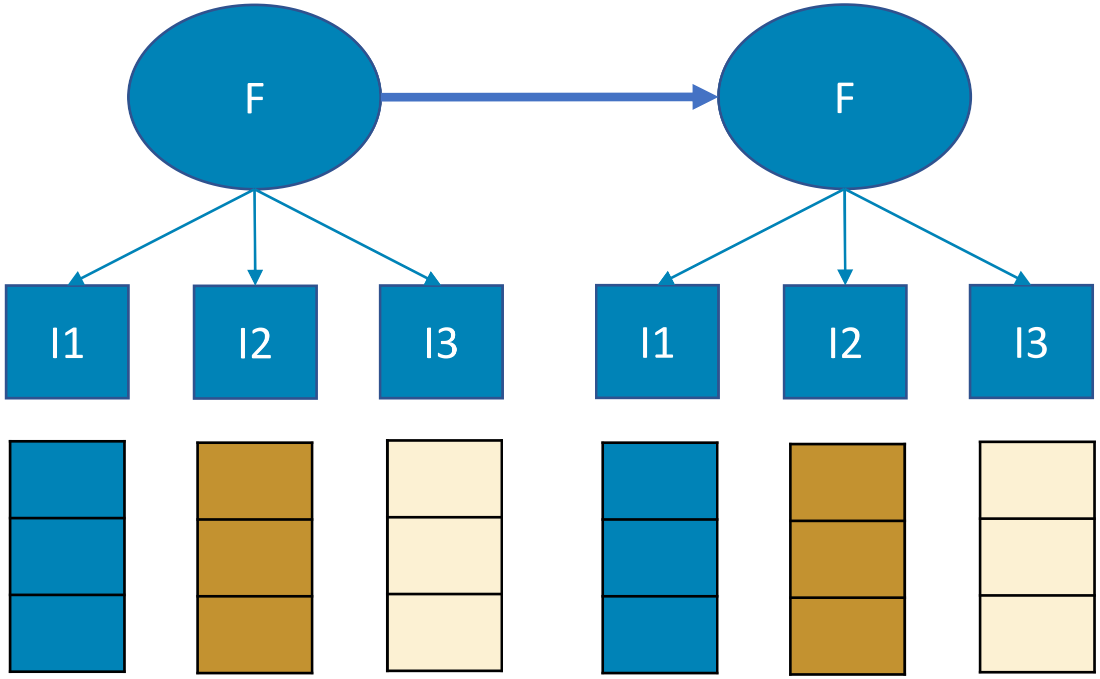




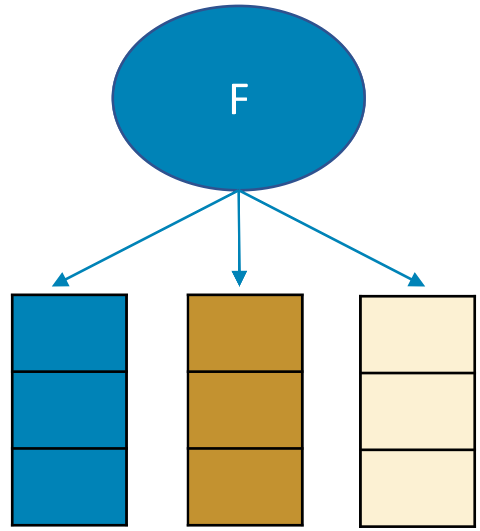
Panel vs Multilevel

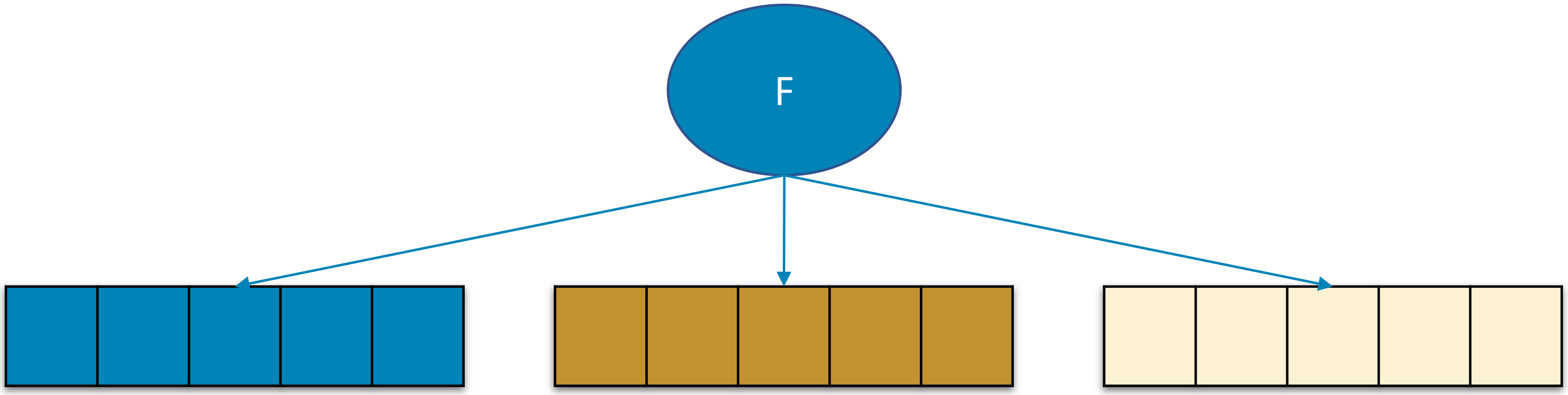
Panel vs Multilevel





Panel vs Multilevel

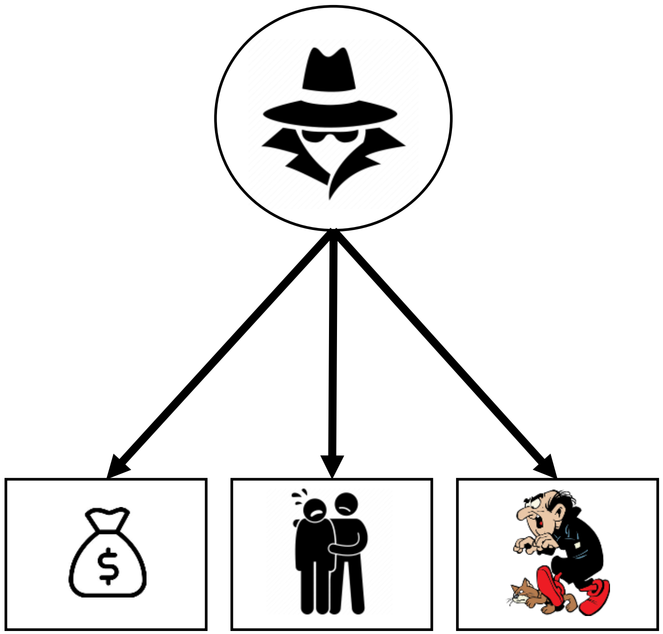


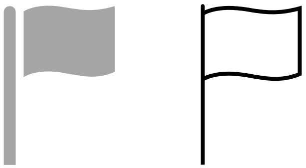


Measurement Invariance

Measurement Invariance

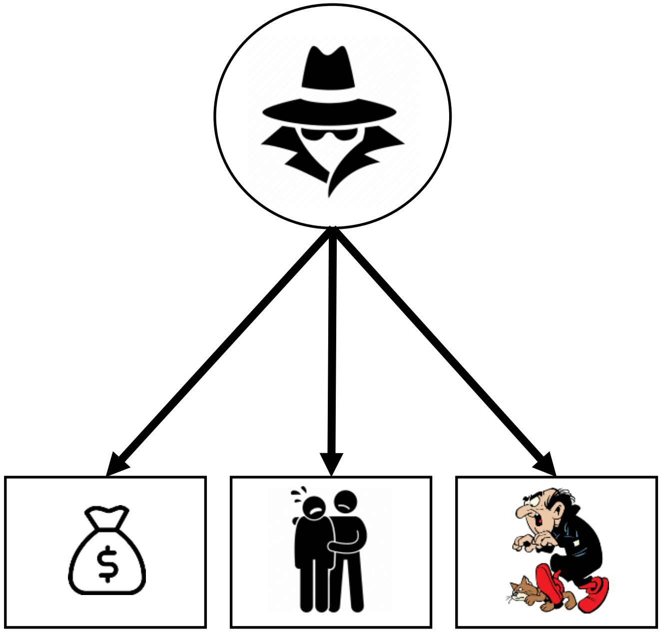
Different levels of invariance (that build up on each other) need to hold for valid between-group comparisons of the latent factor.

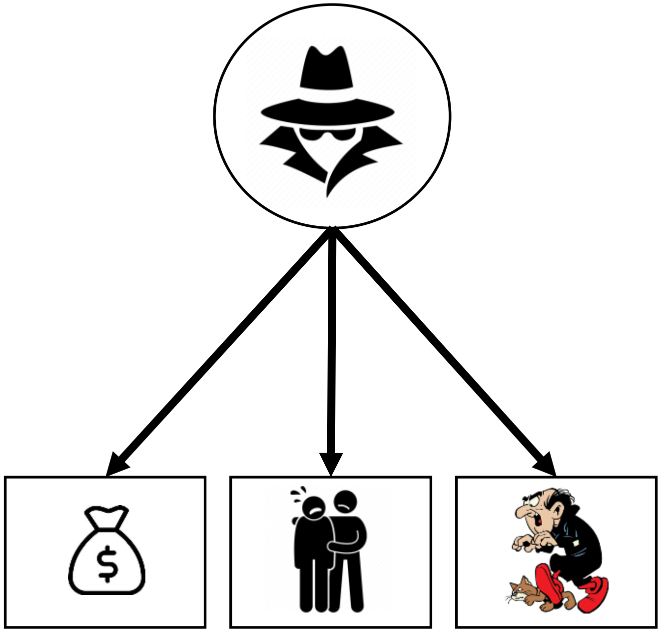




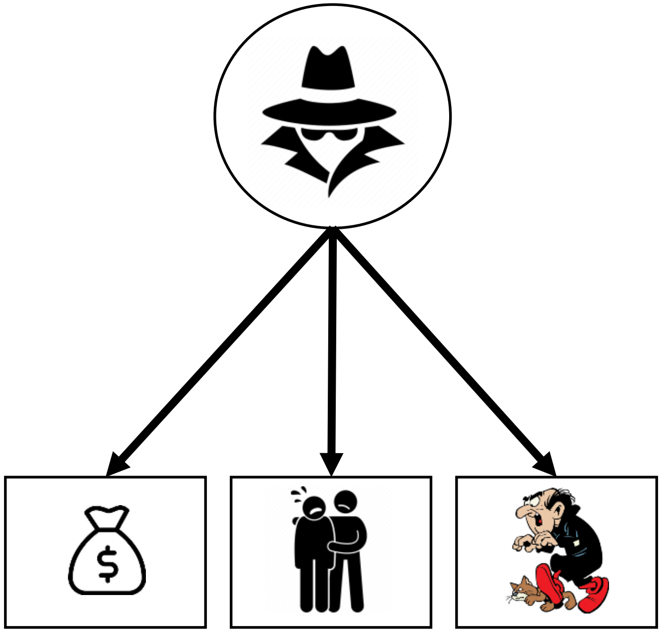
I dream about stealing money.
I do not sympathize with others.
I root for the bad guy in movies.

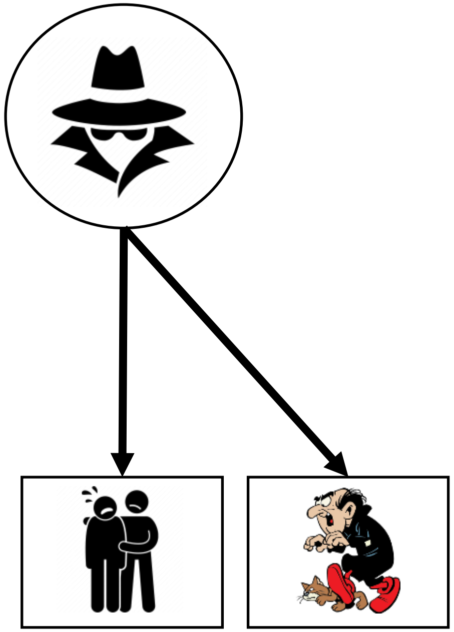
Configural/Pattern Invariance



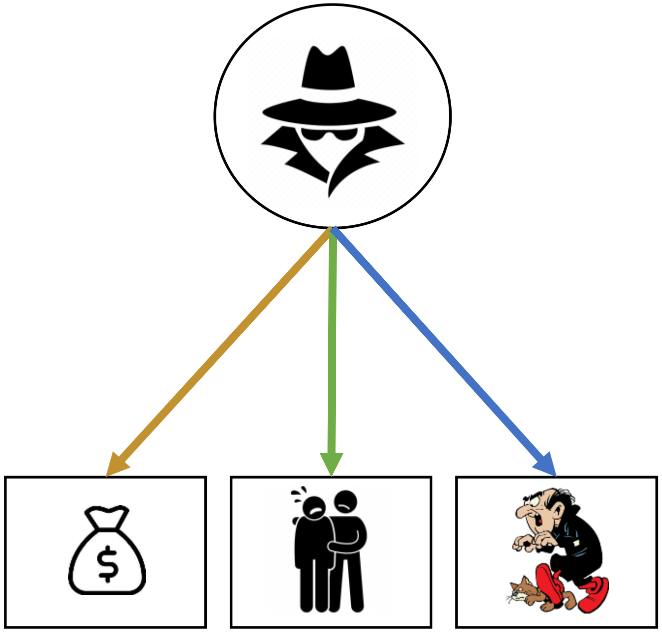


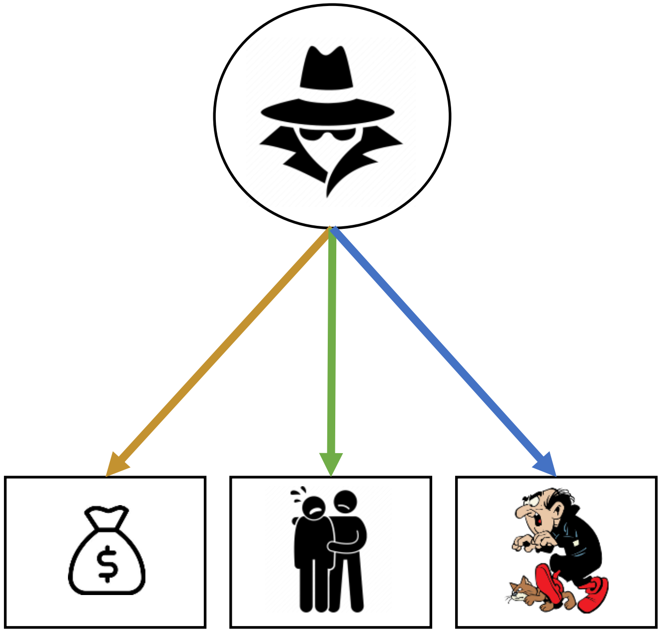
Configural/Pattern Invariance





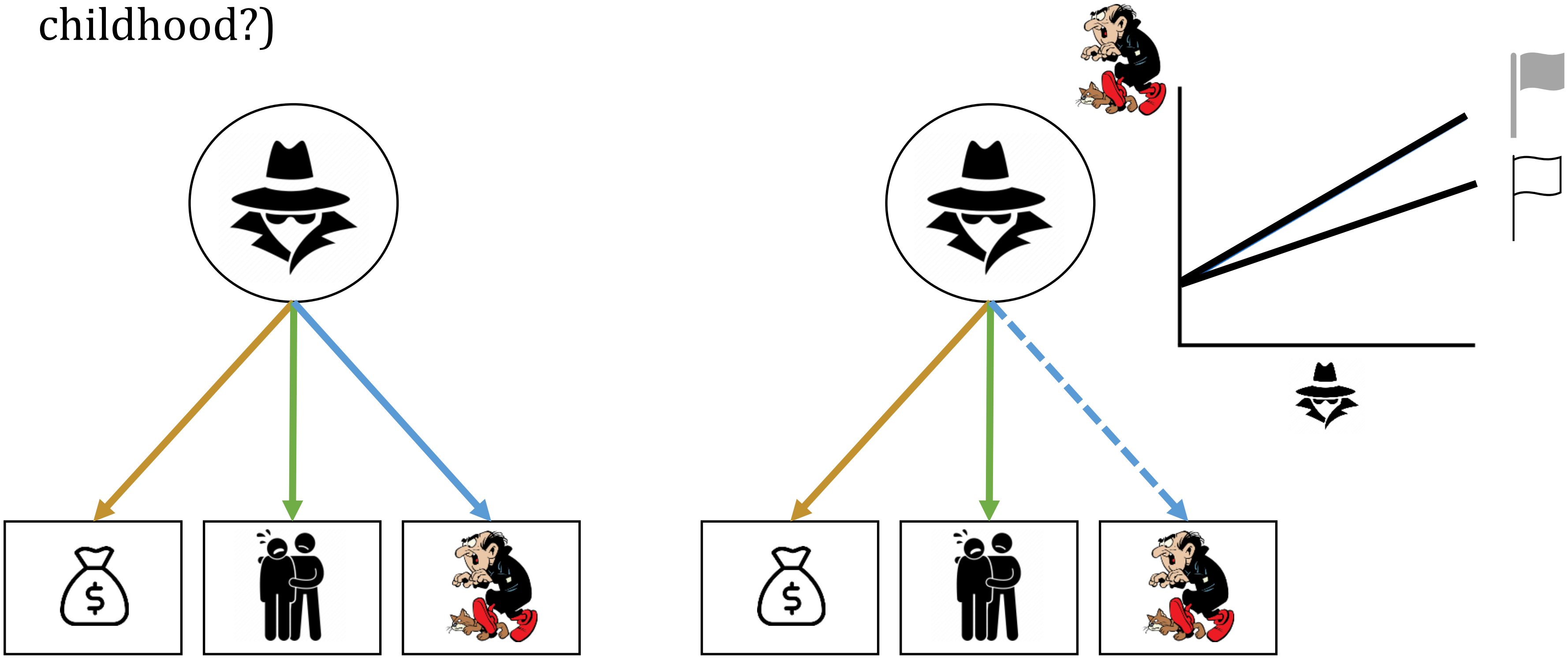
Loading Invariance





Metric invariance

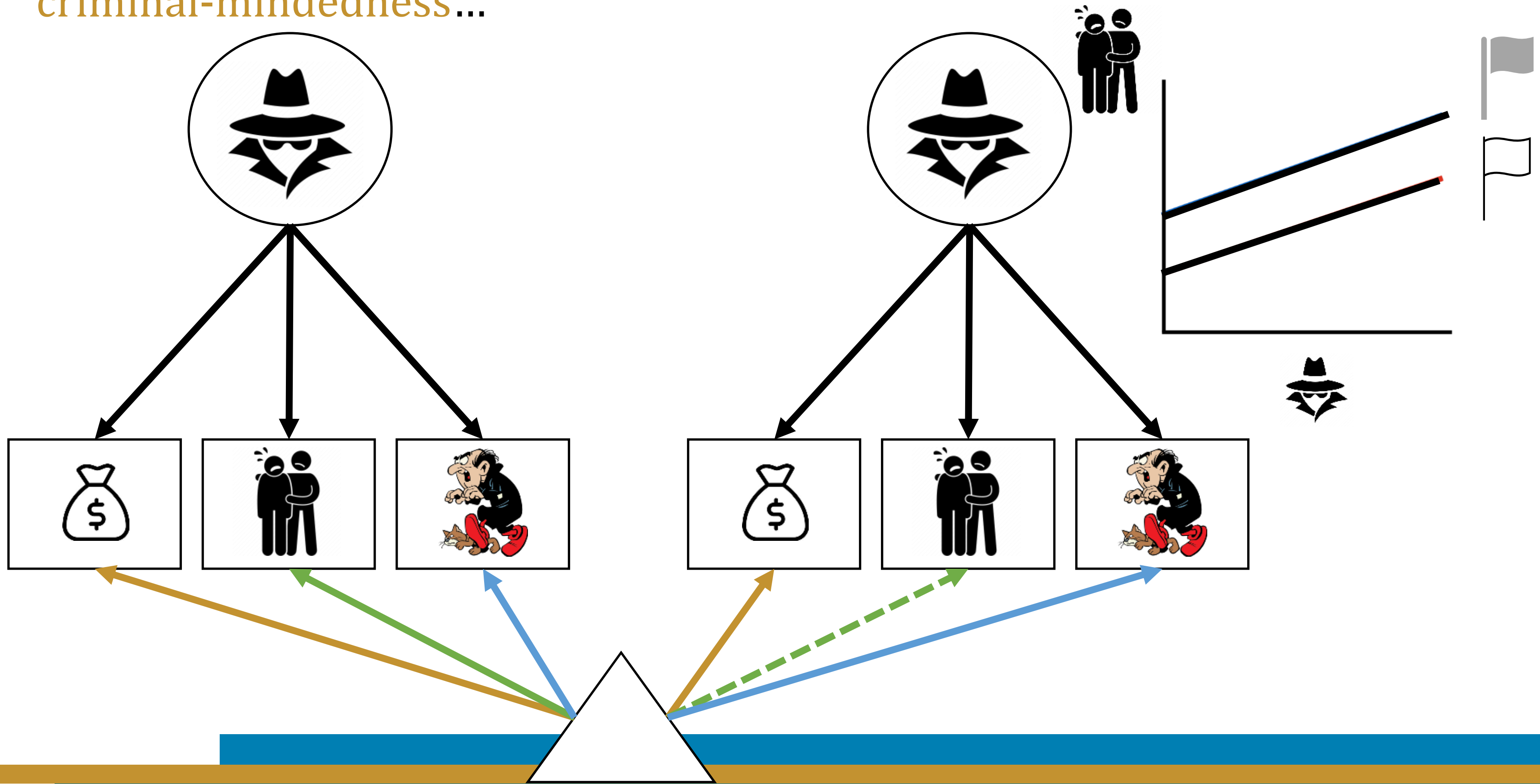
- *Identifying with the bad guy* is a worse indicator of criminal-mindedness in the white country; more other reasons for that (empathy for bad childhood?)



- If we ignore this, then we might misinterpret differences in item scores as differences in criminal-mindedness, although they are at least partly induced by differences in the regression weights.

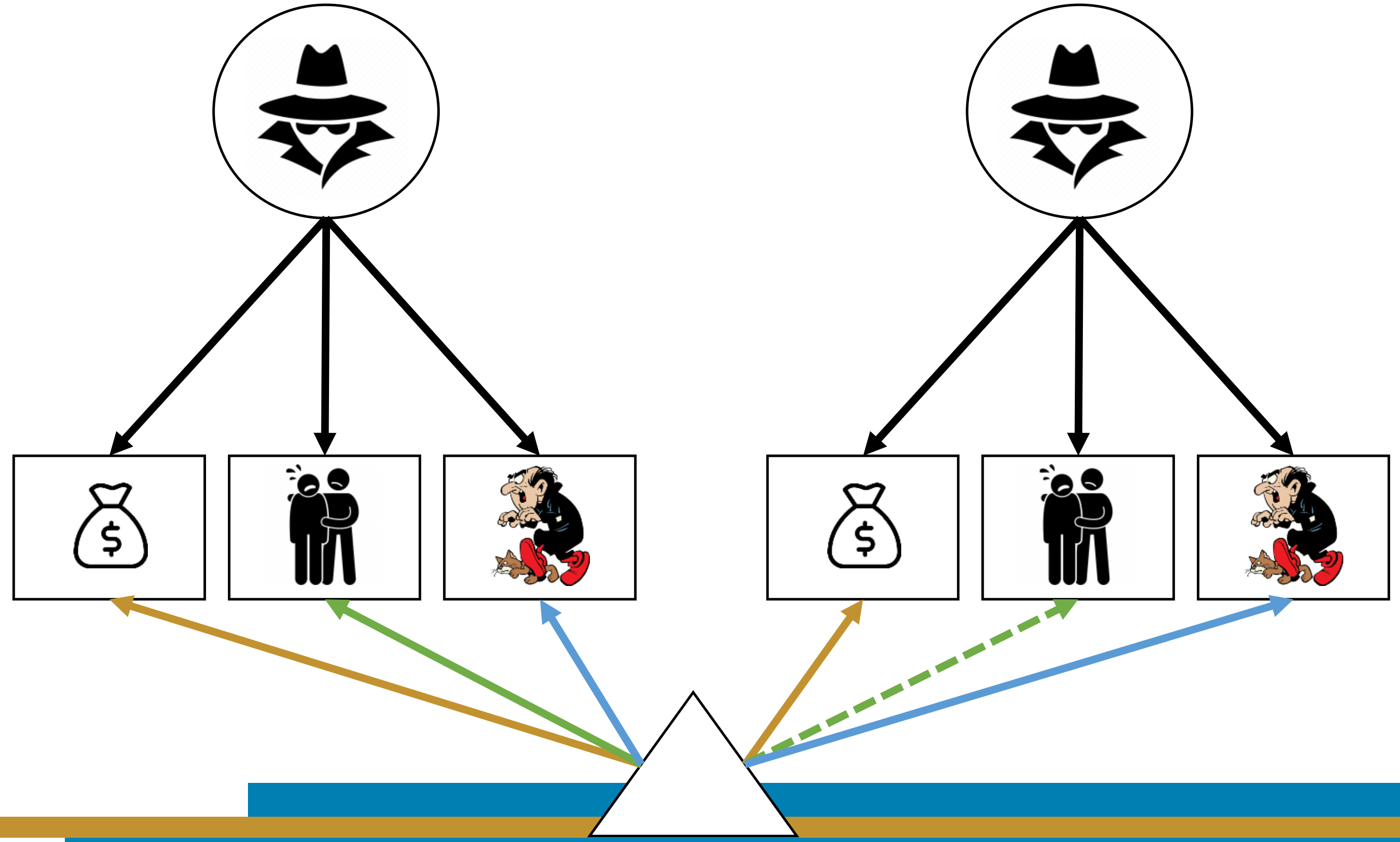
Scalar invariance

- One (perhaps very individualistic) country scores generally lower on *sympathizing with others*, regardless of the criminal-mindedness.
- Again: differences in item scores would be misinterpreted as differences in criminal-mindedness…



Partial invariance

- You can allow for small differences in loadings and intercepts across groups!
- You will see that in the lab session.



Regression Notation

- For cross-sectional data

$$\boldsymbol{y}_{i,g} = \boldsymbol{\tau}_g + \boldsymbol{\Lambda}_g \boldsymbol{\eta}_{i,g} + \boldsymbol{\epsilon}_{i,g}$$

- For (intensive) longitudinal data

$$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i,t} + \boldsymbol{\Lambda}_{i,t} \boldsymbol{\eta}_{i,t} + \boldsymbol{\epsilon}_{i,t}$$

Invariance for longitudinal data

Person

Time		No restrictions	Invariance constraints on the measurement model
	No restrictions	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i,t} + \boldsymbol{\Lambda}_{i,t}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over time and subjects	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{t} + \boldsymbol{\Lambda}_{t}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over time Measurement invariance over subjects
	Invariance constraints on the measurement model	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i} + \boldsymbol{\Lambda}_{i}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over subjects Measurement invariance over time	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ Measurement invariance over time and subjects

Adolf et al. 2014

Invariance for longitudinal data

Person

Time	No restrictions	Invariance constraints on the measurement model
No restrictions	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i,t} + \boldsymbol{\Lambda}_{i,t}\,\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over time and subjects	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{t} + \boldsymbol{\Lambda}_{t}\,\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over time Measurement invariance over subjects
Invariance constraints on the measurement model	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i} + \boldsymbol{\Lambda}_{i}\,\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over subjects Measurement invariance over time	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\,\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ Measurement invariance over time and subjects

Adolf et al. 2014

Invariance for longitudinal data

Person

Time	No restrictions	Invariance constraints on the measurement model
No restrictions	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i,t} + \boldsymbol{\Lambda}_{i,t}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over time and subjects	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{t} + \boldsymbol{\Lambda}_{t}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over time Measurement invariance over subjects
Invariance constraints on the measurement model	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i} + \boldsymbol{\Lambda}_{i}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over subjects Measurement invariance over time	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ Measurement invariance over time and subjects

Adolf et al. 2014

Invariance for longitudinal data

Person

Time	No restrictions	Invariance constraints on the measurement model
No restrictions	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i,t} + \boldsymbol{\Lambda}_{i,t}\,\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over time and subjects	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{t} + \boldsymbol{\Lambda}_{t}\,\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over time Measurement invariance over subjects
Invariance constraints on the measurement model	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i} + \boldsymbol{\Lambda}_{i}\,\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over subjects Measurement invariance over time	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\,\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ Measurement invariance over time and subjects

Adolf et al. 2014

Invariance for longitudinal data

This model is not identified.
We can't use this as a baseline.

Time \ Person	No restrictions	Invariance constraints on the measurement model
No restrictions	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i,t} + \boldsymbol{\Lambda}_{i,t}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over time and subjects	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{t} + \boldsymbol{\Lambda}_{t}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over time Measurement invariance over subjects
Invariance constraints on the measurement model	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i} + \boldsymbol{\Lambda}_{i}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over subjects Measurement invariance over time	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ Measurement invariance over time and subjects

Adolf et al. 2014

Invariance for longitudinal data

We could start by assuming invariance over time and compare a model with and without restrictions across persons.

Time \ Person	No restrictions	Invariance constraints on the measurement model
No restrictions	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i,t} + \boldsymbol{\Lambda}_{i,t}\,\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over time and subjects	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{t} + \boldsymbol{\Lambda}_{t}\,\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over time Measurement invariance over subjects
Invariance constraints on the measurement model	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i} + \boldsymbol{\Lambda}_{i}\,\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over subjects Measurement invariance over time	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\,\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ Measurement invariance over time and subjects

Adolf et al. 2014

Invariance for longitudinal data

Person

Time	No restrictions	Invariance constraints on the measurement model
No restrictions	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i,t} + \boldsymbol{\Lambda}_{i,t}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over time and subjects	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{t} + \boldsymbol{\Lambda}_{t}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over time Measurement invariance over subjects
Invariance constraints on the measurement model	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i} + \boldsymbol{\Lambda}_{i}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over subjects Measurement invariance over time	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ Measurement invariance over time and subjects

We could start by assuming invariance over persons and compare a model with and without restrictions across time.

Adolf et al. 2014

Invariance for longitudinal data

Person

Time

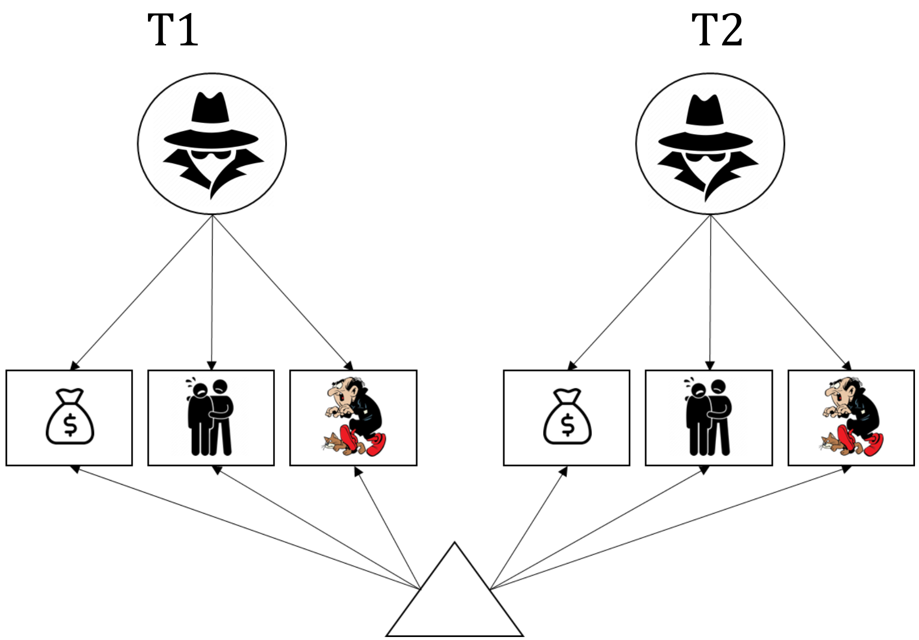
	No restrictions	Invariance constraints on the measurement model
No restrictions	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i,t} + \boldsymbol{\Lambda}_{i,t}\,\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over time and subjects	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{t} + \boldsymbol{\Lambda}_{t}\,\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over time Measurement invariance over subjects
Invariance constraints on the measurement model	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i} + \boldsymbol{\Lambda}_{i}\,\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ No invariance over subjects Measurement invariance over time	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\,\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ Measurement invariance over time and subjects

Often unrealistic.

Adolf et al. 2014

Invariance for longitudinal data

- For Panel Data: You can still simply run pairwise comparisons.
- For ILD: This becomes too much!!



Invariance for ILD

- Alternative to testing: Consider intensive longitudinal data as cross-classified (McNeish et al. 2020).
  - He builds up on multilevel analysis but now lets observations be nested in both individuals and time, resulting in twee variance terms.

	ID = 1	ID = 2	ID = 3
T = 1	Y_{11}	Y_{12}	Y_{13}
T = 2	Y_{21}	Y_{22}	Y_{23}
T = 3	Y_{31}	Y_{32}	Y_{33}

	ID = 1	ID = 2	ID = 3
T = 1	Y_{11}	Y_{12}	Y_{13}
T = 2	Y_{21}	Y_{22}	Y_{23}
T = 3	Y_{31}	Y_{32}	Y_{33}

- What is high variance?

Invariance for ILD

- There is variance/non-invariance across time/subjects. And then?
  - You may try to explain the variance by adding covariates
  - In DSEM (Mplus),
    - it is possible to have
      - random intercepts and slopes to capture between-person differences
      - random intercepts for differences across time
    - it is NOT possible to have
      - loadings vary across time
  - But do we even want that? Are constructs at all comparable?
  - Should we not rather embrace measurement model changes and differences?
  - What if even the number and nature of factors differ?

Break

Discussion

Reliability

Reliability for Panel Data & ILD

- Reliability (assuming invariance)
- In addition to the cross-sectional/between-person reliability, we can also check:
  - Longitudinal reliability:
    - Factor correlation (like a test-retest reliability)
    - Interrater reliability: whether rank ordering of factor scores is stable across time
    - Reliability per time-point
  - The reliability per person

- Whether you go “classic” or multilevel, it’s basically doing cross-sectional reliability several times (plus some extras).

Reliability

- All reliability measures based on:

$$\frac{\text{True Variance}}{\text{Total Variance}}$$

- Just differences in how the constant (true) part is defined/quantified.
  - And sometimes even what the total variance we should consider is.
- This distinction harder to make with single-item measures, so will focus on multi-item measures
  - But see ME(V)AR (Schuurman & Hamaker, 2019) and immediate test-retest (Dejonckheere et al, 2021)

A practical guide to studying reliability of intensive longitudinal data

ROTTERDAM ESM GROUP

Leertouwer, IJ.[1], van Roekel, G.H.[2], Keijsers, L.G.M.T.[1], Schuurman, N.K.[3]

1 Erasmus Universiteit Rotterdam 2 Tilburg University 3 Universiteit Utrecht

The Challenge

$$rel(y) = \frac{\sigma_\theta^2}{\sigma_y^2} = 1 - \frac{\sigma_\epsilon^2}{\sigma_y^2}$$

In CTT: $y_t = \underbrace{\theta}_{\text{true score}} + \underbrace{\epsilon_t}_{\text{measurement error}}$

In ILD: $y_t = \theta_t + \epsilon_t$

How to separate true score variance from error variance?

The Conditions	The Method	The Trick	The Math
How many items for the construct? 1 → Lagged relation(s) plausible? Yes	(ML) ME(V)AR(1) (Schuurman et al., 2015; Schuurman & Hamaker, 2019) & Other models that add random errors	Estimate variance of random error(s) at each measurement occasion	Example within; between possible $y_t = \mu + \tilde{y}_t + \epsilon_t$ $\tilde{y}_t = \phi\tilde{y}_{t-1} + \zeta_t$ $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ $\zeta_t \sim N(0, \sigma_\zeta^2)$ $rel(y) = \frac{\sigma_\zeta^2/(1-\phi^2)}{\sigma_\zeta^2/(1-\phi^2) + \sigma_\epsilon^2}$
1 → Lagged relation(s) plausible? No (or Yes)	(Immediate) test-retest (Dejonckheere et al., 2021)	Repeat some measurements really soon; assume no change in true score	$\sigma_\epsilon^2 = \frac{E[\Delta_{TRT}]^2}{2}$ $\quad rel(y) = 1 - \frac{\sigma_\epsilon^2}{\sigma_{TRT}^2}$
1 → Lagged relation(s) plausible? No (or Yes)	Uncertainty per occasion (in prep.)	Ask for upper and lower bound at each measurement	$\sigma_\epsilon^2 = \frac{E[\Delta_{UL}]^2}{2}$ $\quad rel(y) = 1 - \frac{\sigma_\epsilon^2}{\sigma_{UL}^2}$
2 (or more) + Tau-equivalence + No lagged relations → Fixed measurements? (similar times per participant) Yes	Generalizability Theory (Cranford et al., 2006; Shrout & Lane, 2012)	Decompose variance: person, item, timepoint, interactions, error	$y_{itp} = \mu + I_i + T_t + P_p + (IT)_{it} + (TP)_{tp} + (IP)_{ip} + (ITP)_{itp} + \epsilon_{itp}$ $R_{KF} = \frac{\sigma_P^2 + (\sigma_{IP}^2/m)}{\sigma_P^2 + (\sigma_{IP}^2/m) + (\sigma_\epsilon^2/km)}$ $R_{1R} = \frac{\sigma_P^2 + (\sigma_{IP}^2/m)}{\sigma_P^2 + (\sigma_{IP}^2/m) + (\sigma_T^2) + (\sigma_{TP}^2) + (\sigma_\epsilon^2/m)}$ +Others
2 (or more) → Fixed measurements? No	Alpha multilevel (Nezlek, 2017; Shrout & Lane, 2012)	Decompose variance: items within occasions within persons and error	$y_{itp} = \mu + P_p + T(P)_{tp} + \epsilon_{itp}$ $R_{CN} = \frac{\sigma_{T(P)}^2}{\sigma_{T(P)}^2 + \sigma_\epsilon^2/m}$ $R_{KRN} = \frac{\sigma_P^2}{\sigma_P^2 + (\sigma_{T(P)}^2/k)\,(\sigma_\epsilon^2/km)}$
2 (or more) → Fixed measurements?	Alpha multilevel (Geldhof et al., 2014; Lai, 2021)	Compute alpha on within-person and between-person covariance matrices	$\alpha_W = \frac{K^2\bar{\sigma}_{ijW}}{\sigma_{yW}^2}$ $\quad \alpha_B = \frac{K^2\bar{\sigma}_{ijB}}{\sigma_{yB}^2 + (\sigma_{yW}^2/\tilde{m})}$
3 (or more): Measurement invariance / Tau-equivalence testable! → Lagged relation(s) plausible? No	Omega (multilevel) (Geldhof et al., 2014; Lai, 2021)	Compute omega on within-person and between-person loadings and errors	$\omega_W = \frac{\left(\sum_{i=1}^K \lambda_i\right)^2 \sigma_{FW}^2}{\left(\sum_{i=1}^K \lambda_i\right)^2 \sigma_{FW}^2 + \sum_{i=1}^K \epsilon_{iW}}$ $\omega_B = \frac{\left(\sum_{i=1}^K \lambda_{iB}\right)^2 \sigma_{FB}^2}{\left(\sum_{i=1}^K \lambda_{iB}\right)^2 (\sigma_{FB}^2 + \sigma_{FW}^2/\tilde{m}) + \sum_{i=1}^K \epsilon_{kB} + \sum_{i=1}^K \epsilon_{iW}/\tilde{m}}$
3 (or more) → Lagged relation(s) plausible? Yes	DFAR(1) (Molenaar, 1985; Song & Ferrer, 2012)	Adjust omega for lagged relation(s)	Example within; between possible $y_t = \mu + \lambda\tilde{y}_t + \epsilon_t$ $\tilde{y}_t = \phi\tilde{y}_{t-1} + \zeta_t$ $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ $\zeta_t \sim N(0, \sigma_\zeta^2)$ $rel(y) = \frac{\left(\sum_{i=1}^K \lambda_i\right)^2 (\sigma_\zeta^2/1-\phi^2)}{\left(\sum_{i=1}^K \lambda_i\right)^2 (\sigma_\zeta^2/1-\phi^2) + \left(\sum_{i=1}^K \sigma_{\epsilon i}^2\right)}$

y = observed datapoint(s), σ^2 = variance, θ = true score, ϵ = measurement error, t = timepoint, μ = mean(s), ϕ = AR parameter, $\tilde{y}$ = dynamic process, ζ = innovation, Δ = difference, TRT = test retest, UL = uppper lower bound, i = item, p = person, K = total number of items, $\bar{\sigma}_{ij}$ = average covariance, W = within person, B = between person, m = number of timepoints, $\tilde{m}$ = harmonic mean number of timepoints, α = coefficient alpha, ω = coefficient omega, λ = factor loadings

Erasmus University Rotterdam

Leertouwer@essb.eur.nl: R code for all methods is almost ready!

The Many Reliabilities of Affective Dynamics

Sebastian Castro-Alvarez[1], Laura F. Bringmann[2], Jason Back[3], and Siwei Liu[1]

[1]University of California, Davis

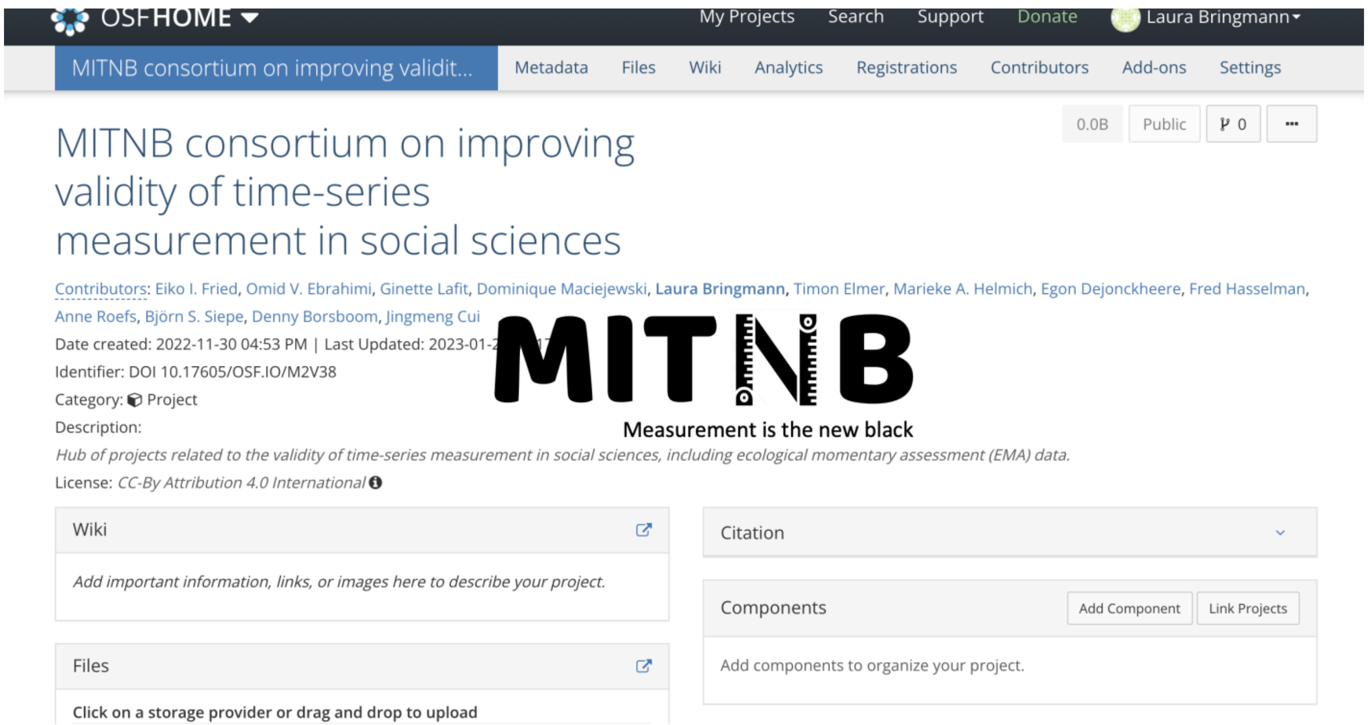
[2]University of Groningen

[3]California State University, Sacramento

Abstract

Reliability is a key concept in psychology that has been broadly studied since the introduction of Cronbach's alpha, which is a measure of the internal consistency of a test. Despite its importance, this is a topic that is relatively understudied when dealing with intensive longitudinal data. In particular, when studying the psychological dynamics of affective states, there is no warranty that intensive longitudinal measurements are reliable. Given this, empirical researchers need tools to study and report the reliability of the scales used in intensive longitudinal research. In recent years, different approaches to estimate the reliability of the scales and the items used when studying psychological dynamics have been proposed. However, the advantages and disadvantages of each of these methods are unclear, making it difficult to determine when a certain approach would be preferred over the others. Specifically, these diverse approaches estimate reliability indices based on statistical models such as linear multilevel analysis, vector autoregressive models and dynamic factor models. Furthermore, while some methods suggest estimating one reliability index for the scale that applies to the whole sample, others estimate specific reliability indices for each individual in the sample. This wide variety of approaches can provoke some confusion for empirical researchers. Therefore, we aim to highlight the advantages and disadvantages of each of the available methods used to estimate the reliability of intensive longitudinal data. We also showcase their use with empirical data.

Keywords: reliability, dynamic factor analysis, experience sampling methods

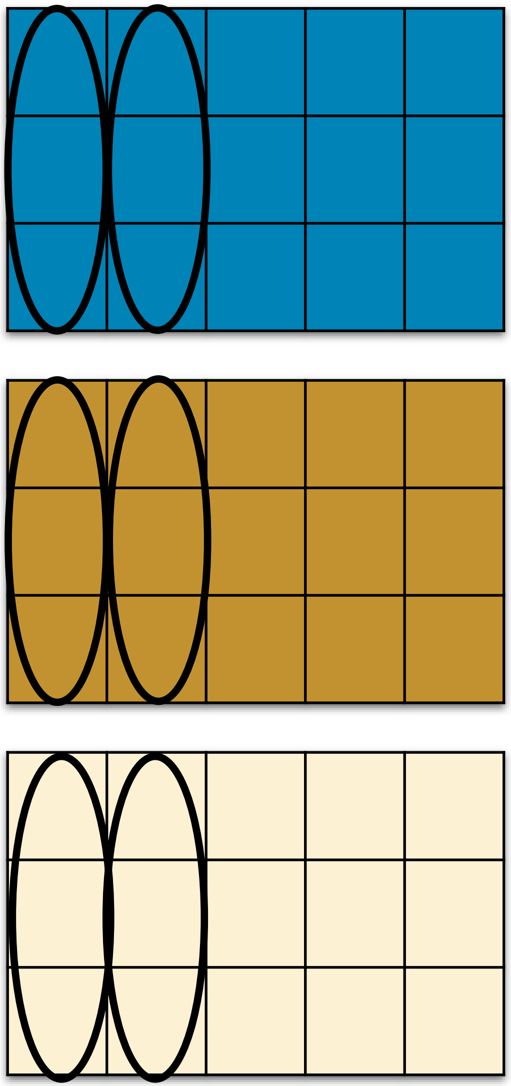


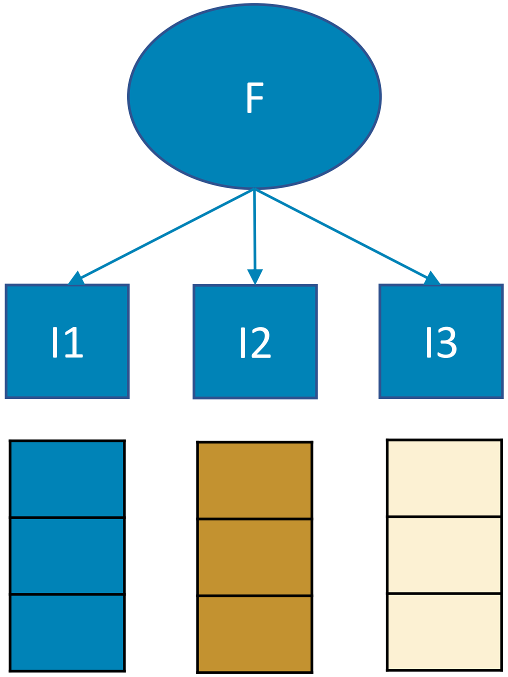
https://osf.io/m2v38/

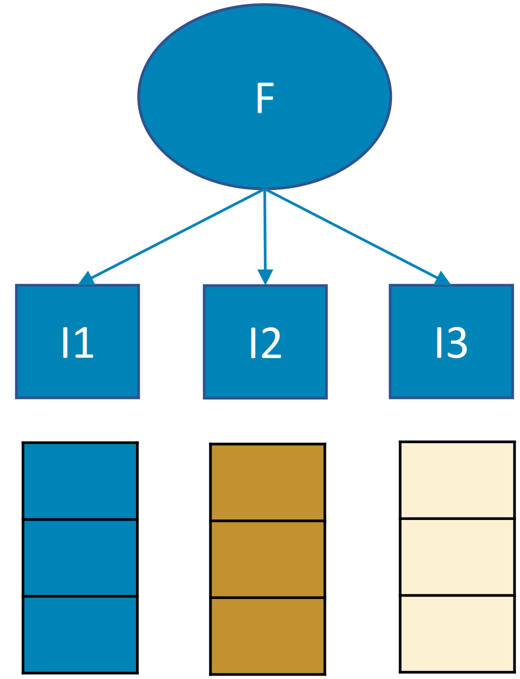
Reliability (assuming invariance)

- Several statistics developed in 1930s-1950s as short cut estimates of reliability, for instance, Cronbach's alpha.
  - Developed before technological advances to find factor structures of scales.
  - Now we can use model-based estimates for reliability that do consider factor structure of scales.
  - This is important in case not all items measure the factors equally well.
  - We will use omega:
  - $\omega = \frac{\left(\sum_{i=1}^{k} \lambda_i\right)^2 Var(\psi)}{\mathbf{1}'\hat{\Sigma}\mathbf{1}}$
    - with $\hat{\Sigma}$ as the observed covariance matrix,
    - k as the number of items
    - $Var(\psi)$ as the variance of the factor scores
    - $\mathbf{1}$ as a k-dimensional row-vector used to sum elements in the matrix.

Reliability for Panel Data



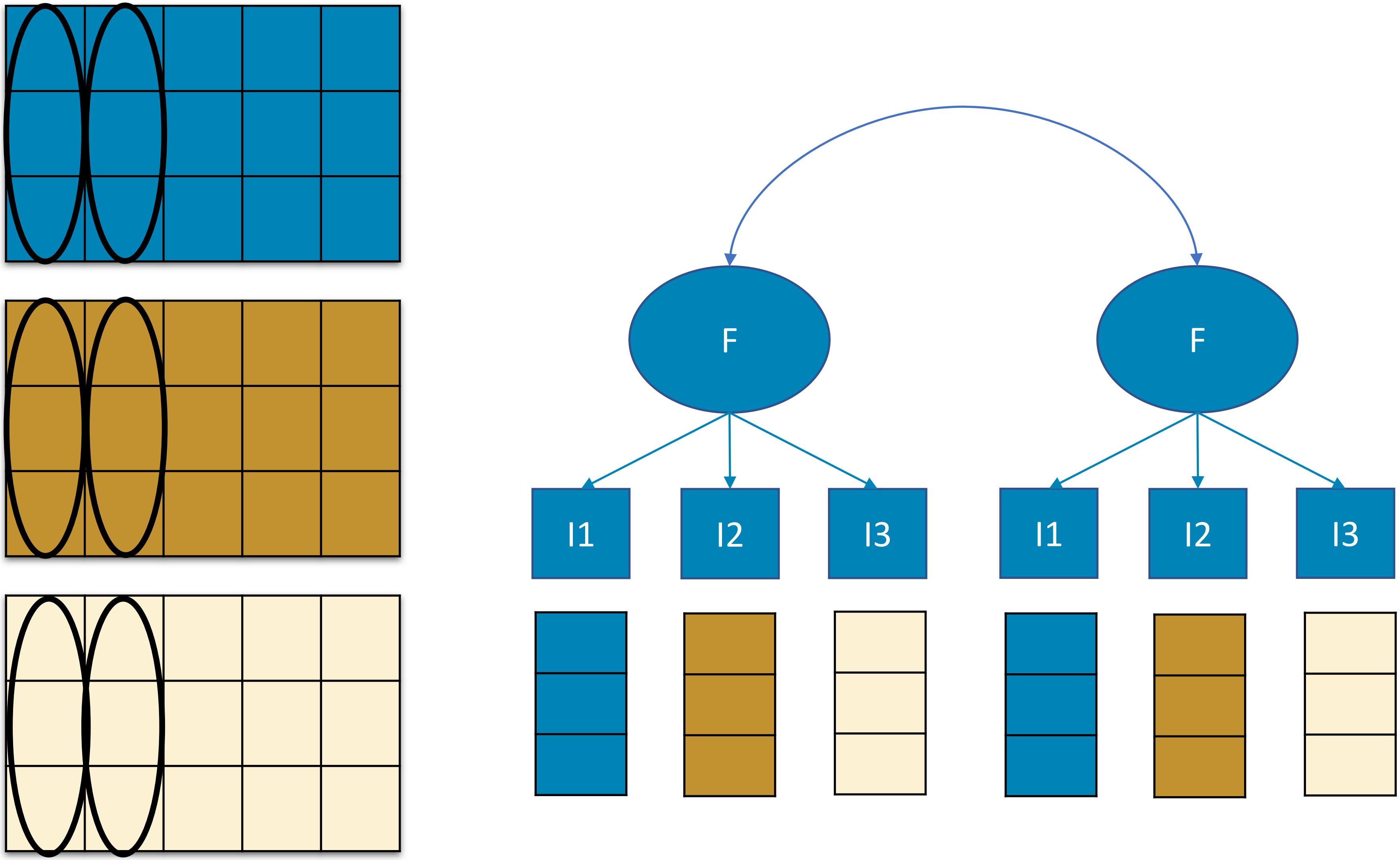




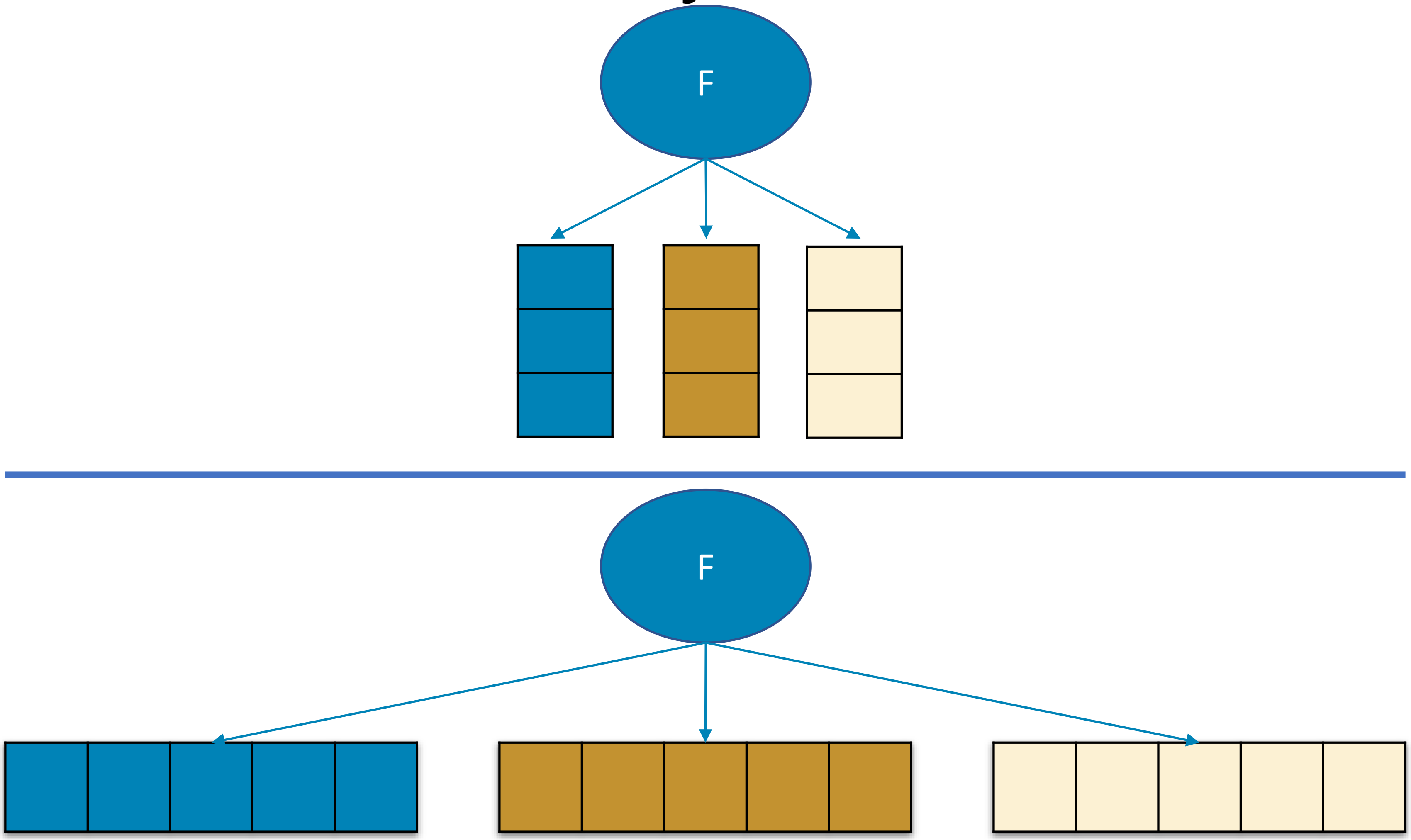
$$\omega = \frac{\left(\sum_{i=1}^{k} \lambda_i\right)^2 Var(\psi)}{\mathbf{1}'\hat{\Sigma}\mathbf{1}}$$

$$\omega = \frac{\left(\sum_{i=1}^{k} \lambda_i\right)^2 Var(\psi)}{\mathbf{1}'\hat{\Sigma}\mathbf{1}}$$

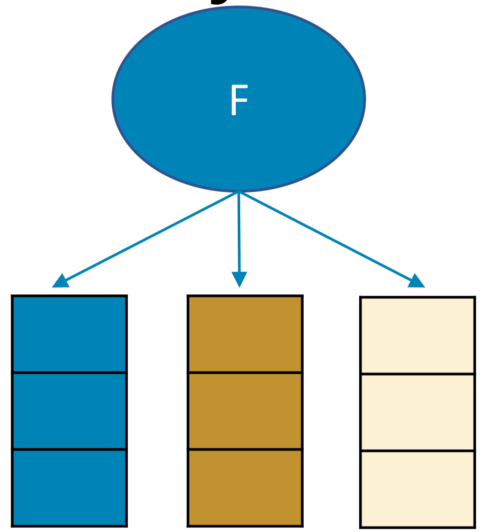
Reliability for Panel Data



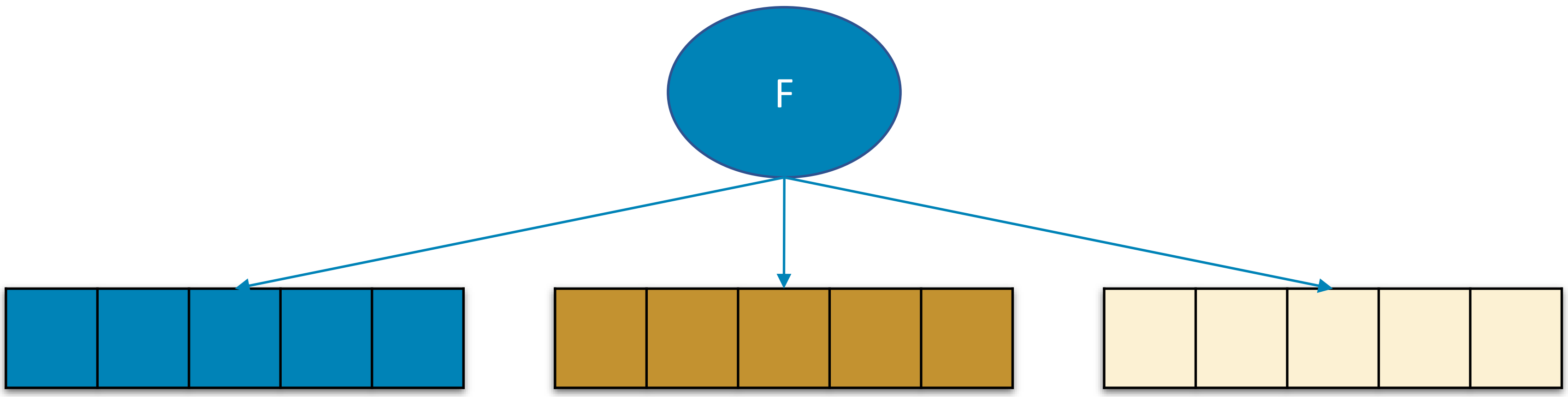
Multilevel Reliability



Multilevel Reliability



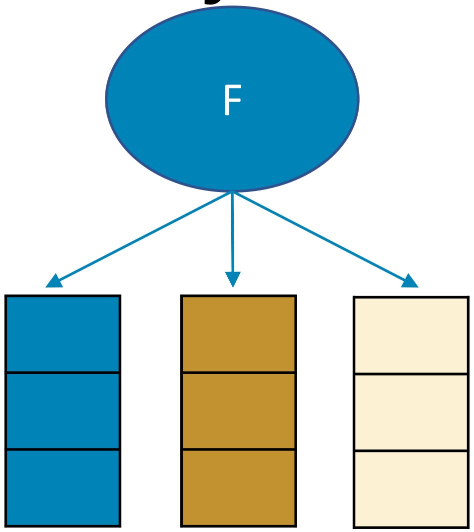
$$\omega = \frac{\left(\sum_{i=1}^{k} \lambda_i\right)^2 Var(\psi)}{\mathbf{1}'\hat{\Sigma}\mathbf{1}}$$



Multilevel Reliability

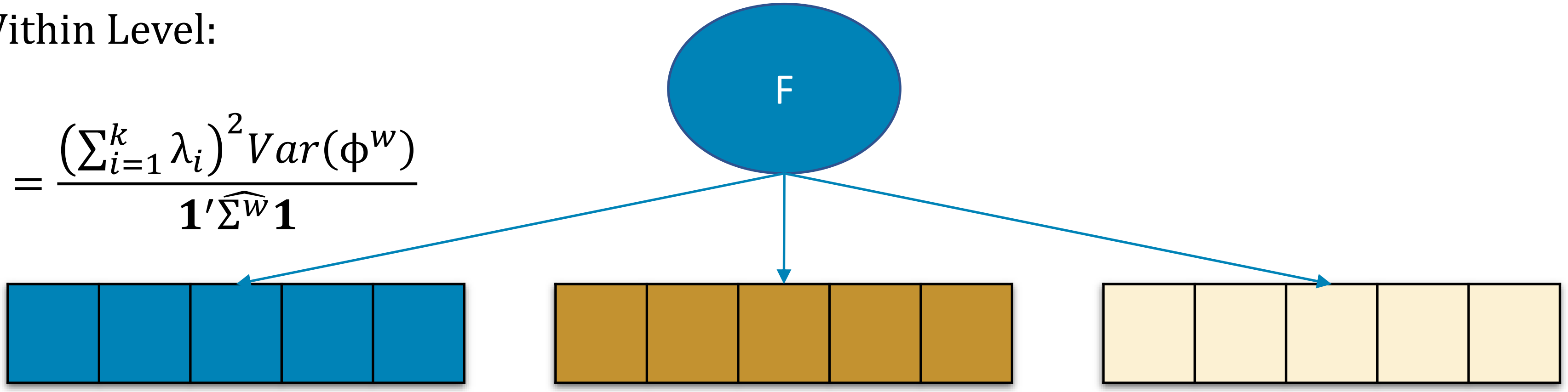
Between Level:

$$\omega = \frac{\left(\sum_{i=1}^{k} \lambda_i\right)^2 Var(\phi^b)}{\mathbf{1}'\widehat{\Sigma^b}\mathbf{1}}$$

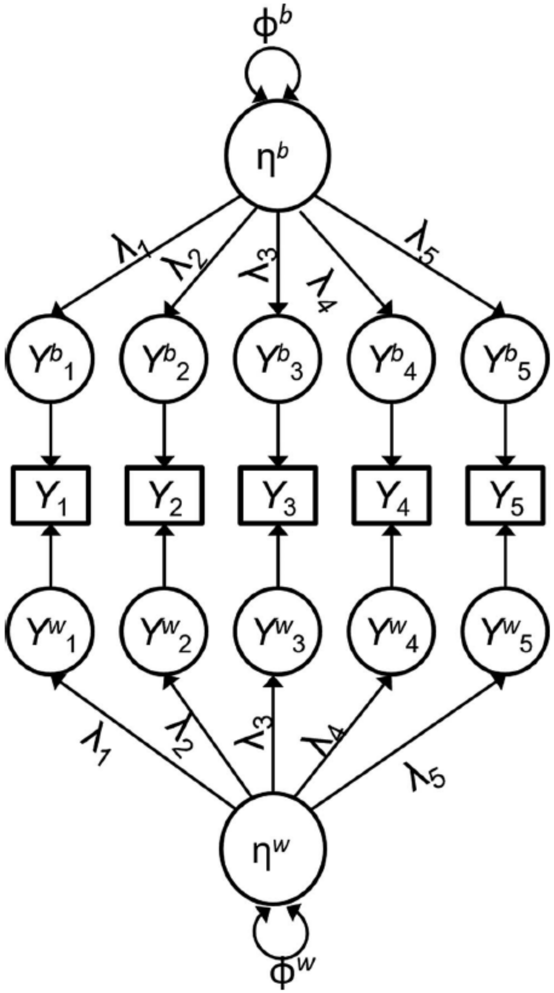


Within Level:

$$\omega = \frac{\left(\sum_{i=1}^{k} \lambda_i\right)^2 Var(\phi^w)}{\mathbf{1}'\widehat{\Sigma^w}\mathbf{1}}$$



Reliability for Panel Data & ILD



Assume invariance between the within- and between factor model and use

$$\omega = \frac{\left(\sum_{i=1}^{k} \lambda_i\right)^2 Var(\psi)}{\mathbf{1}'\hat{\Sigma}\mathbf{1}}$$

$$ICC = \frac{\boldsymbol{\eta}^b}{(\boldsymbol{\eta}^b + \boldsymbol{\eta}^w)}$$

Reliability for Panel Data & ILD

- Previous slide shows that there are different ways to get reliabilities with multilevel analysis for the same basic model.

- But there is even more flexibility, and so, even more ways to calculate reliability with the exact same data.

- Above we split variance in between-person variance and within-person variance, but is that the only important distinction?
  - Do different items behave differently?
  - Do we need to distinguish between specific time-points?

Reliability for Panel Data & ILD

- Different partitionings lead to different "true" variances (that we care about) and total variances.

$$R_{1F} = \frac{\sigma^2_{PERSON} + [\sigma^2_{PERSON*ITEM}/m]}{\sigma^2_{PERSON} + [\sigma^2_{PERSON*ITEM}/m] + [\sigma^2_{ERROR}/m]}.$$

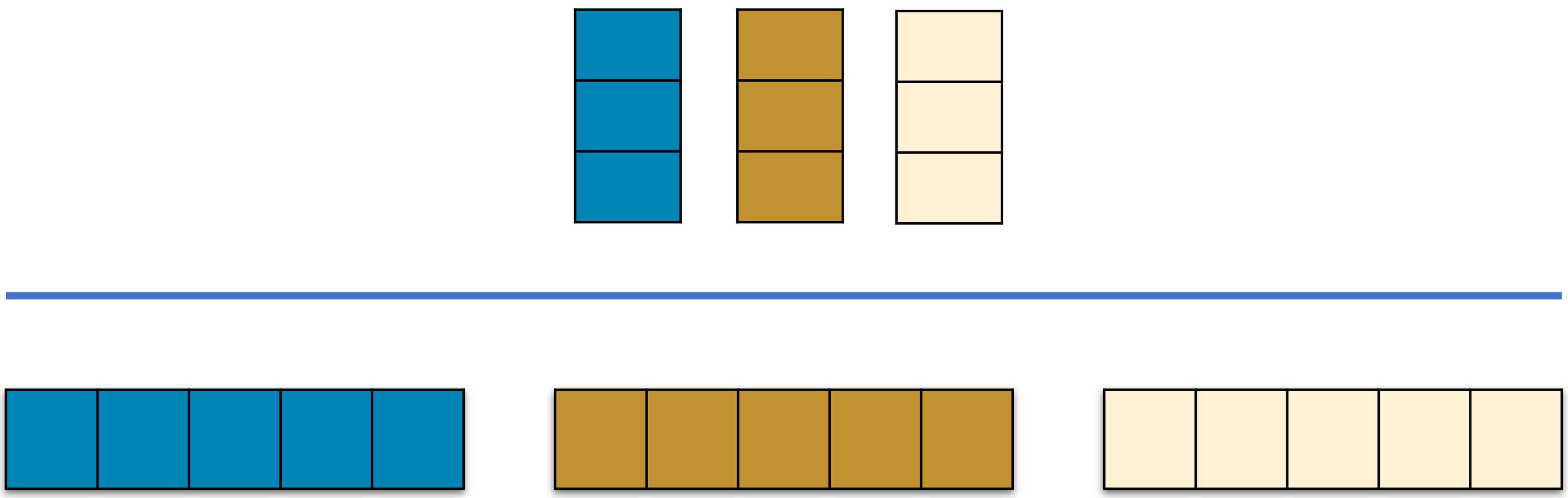
$$\omega = \frac{\left(\sum_{i=1}^{k} \lambda_i\right)^2 Var(\psi)}{\mathbf{1}'\hat{\Sigma}\mathbf{1}}$$

$$R_{1R} = \frac{\sigma^2_{PERSON} + [\sigma^2_{PERSON*ITEM}/m]}{\sigma^2_{PERSON} + [\sigma^2_{PERSON*ITEM}/m] + \sigma^2_{Day} + \sigma^2_{PERSON*DAY} + [\sigma^2_{ERROR}/m]}.$$

$$R_{KF} = \frac{\sigma^2_{PERSON} + [\sigma^2_{PERSON*ITEM}/m]}{\sigma^2_{PERSON} + [\sigma^2_{PERSON*ITEM}/m] + [\sigma^2_{ERROR}/Km]}$$

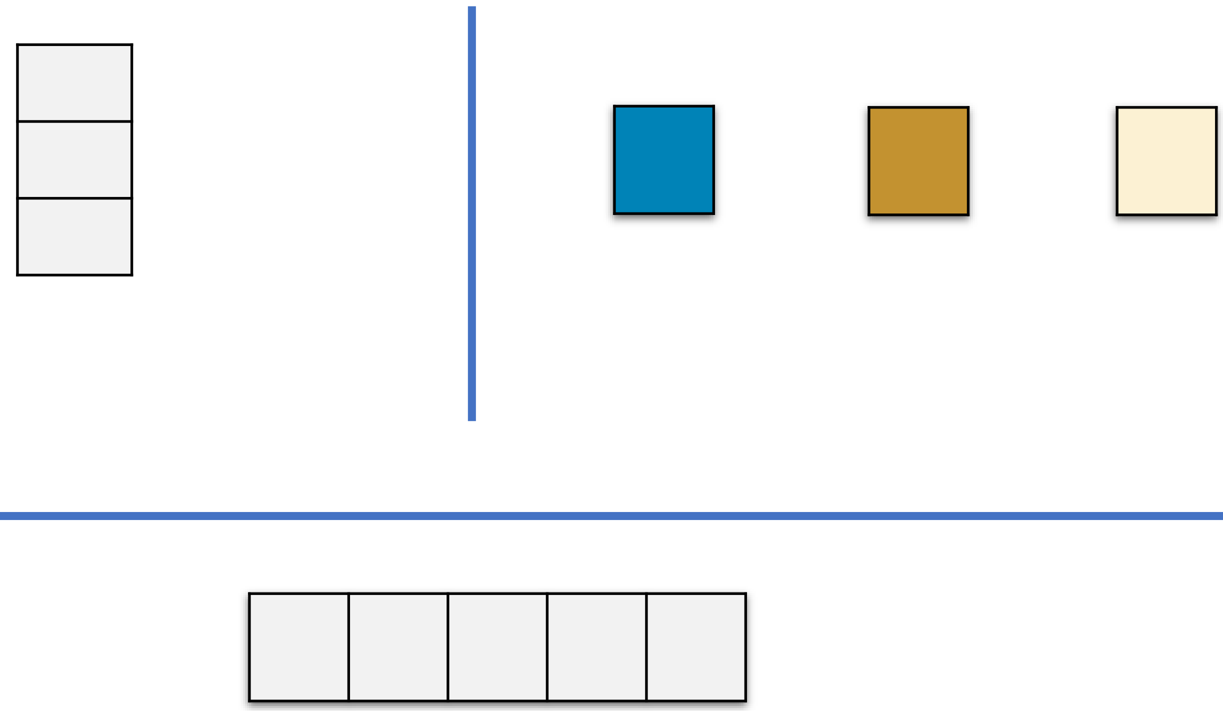
Reliability for Panel Data & ILD

$$Y_{tj} = \mu + Person_j + \ e_{tj}$$



Reliability for Panel Data & ILD

$$Y_{itj} = \mu + Item_i + Person_j + e_{itj}$$



Reliability for Panel Data & ILD

$$Y_{tj} = \mu + Person_j + e_{tj}$$

$$Y_{itj} = \mu + Item_i + Person_j + e_{itj}$$

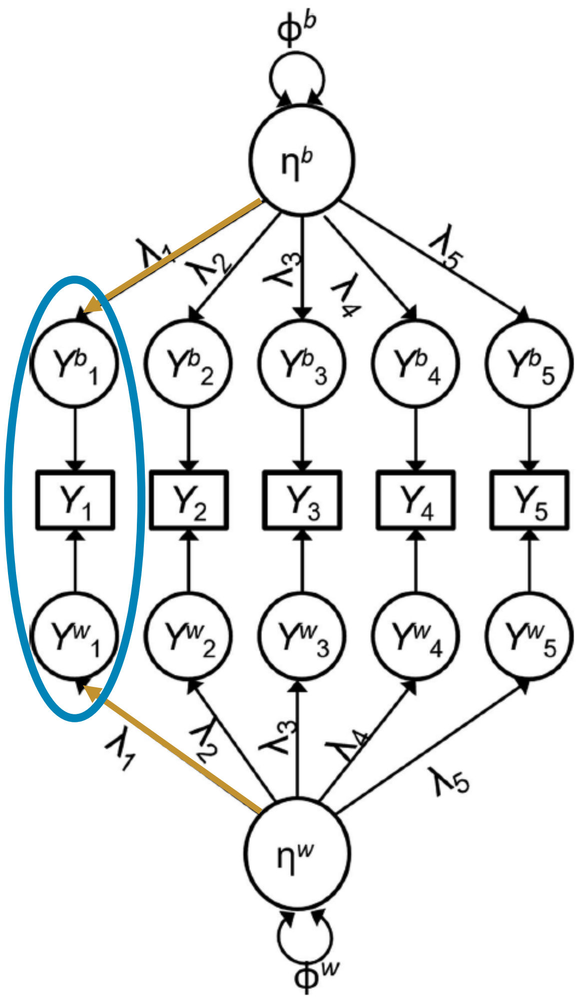
$$Y_{itj} = \mu + Item_i + Day_t + Person_j + e_{itj}$$

$$Y_{itj} = \mu + Item_i + Day_t + Person_j + (ID)_{it} + (IP)_{ij} + (DP)_{tj} + (IDP)_{itj} + e_{itj}$$

Reliability for Panel Data & ILD

- Lai (2021): "Reliability is a characteristic of an observed composite"
  - So doesn't like using latent variables, should use actual composites you can calculate from the data (e.g., sum-scores).

Reliability for Panel Data & ILD



True score variance in Y1

For raw-scores:

$$Y_{ip} = T_i + e_i$$

$$Var(\eta) = \phi^b + \phi^w$$

$$\omega^{2L} = \frac{\left(\sum_{i=1}^{k} \lambda_i\right)^2 (\phi^b + \phi^w)}{\left(\sum_{i=1}^{k} \lambda_i\right)^2 (\phi^b + \phi^w) + \mathbf{1}'\boldsymbol{\Theta}^B\mathbf{1} + \mathbf{1}'\boldsymbol{\Theta}^W\mathbf{1}}$$

$$\omega = \frac{\left(\sum_{i=1}^{k} \lambda_i\right)^2 Var(\psi)}{\mathbf{1}'\hat{\Sigma}\mathbf{1}}$$

Reliability for Panel Data & ILD

- Always compare true score variance to total variance.

- Different methods only differ in what they consider true- and total variance, and in terms of possible constraints on the multilevel factor-model.
    - In factor models, true-score variance is the variance of the factors
    - For composites, true-score variance can be determined based on the factor-model.

Break

Lab

Exploring Non-Invariance for Intensive Longitudinal Data

Exploring Longitudinal Invariance

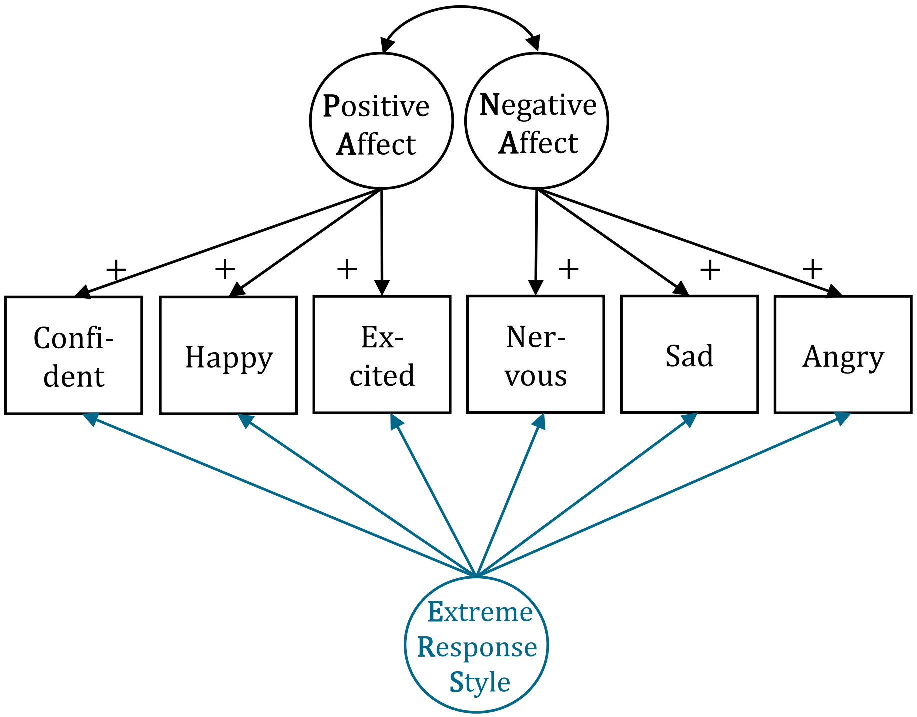
- Pairwise comparisons to detect all types of differences and changes are almost impossible.

- No method, then…

Response Styles

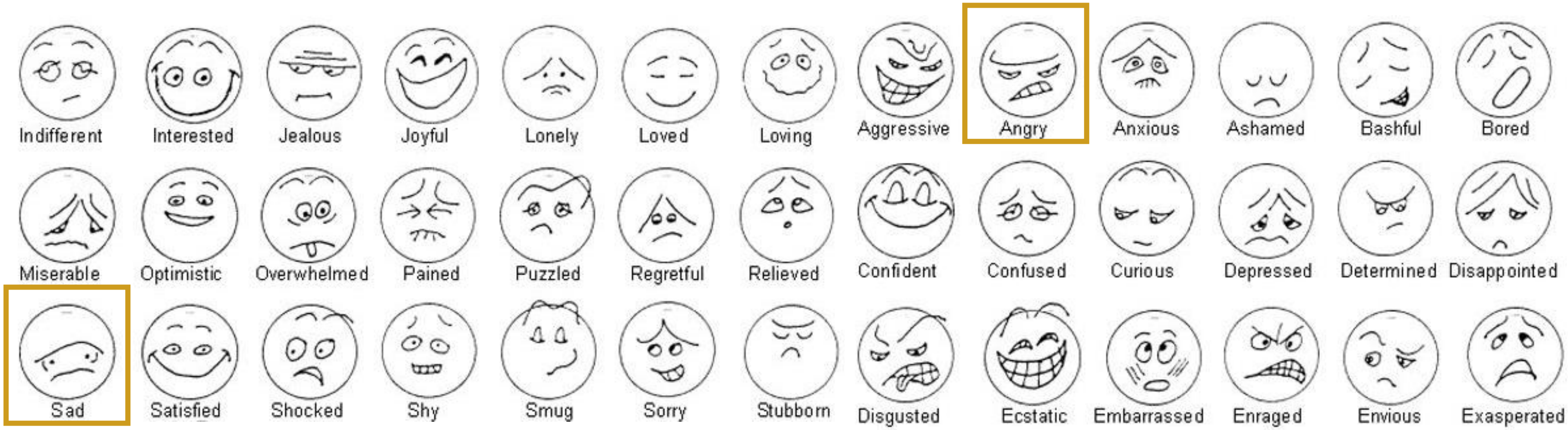
Changes in the factor models (the measurement models) are even more likely than in panel or cross-sectional data.

- Distracting situations
- Getting demotivated



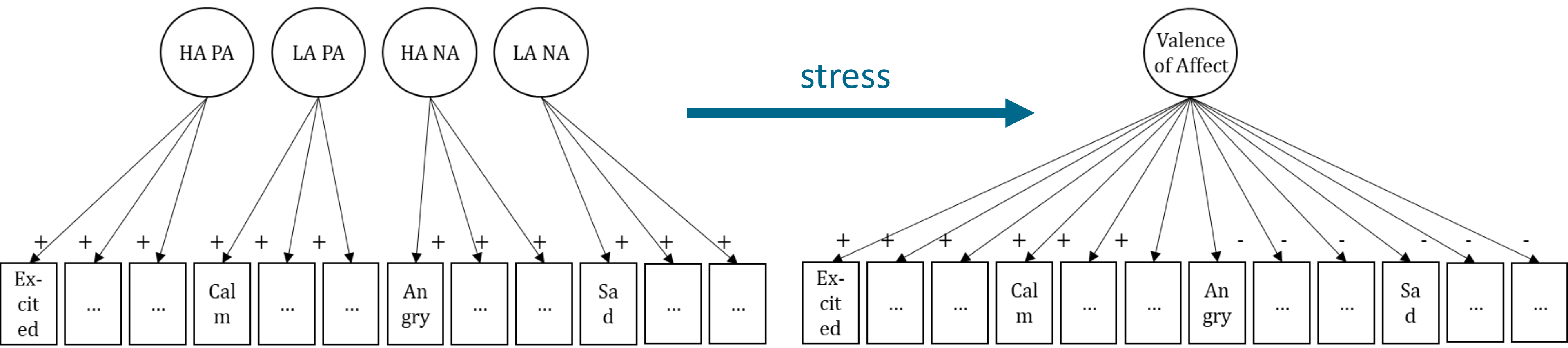
Substantive differences

- Ability to differentiate between emotions
  - High differentiators: Label in a differentiated and context dependent way
  - Low differentiators: Less specific emotional experiences



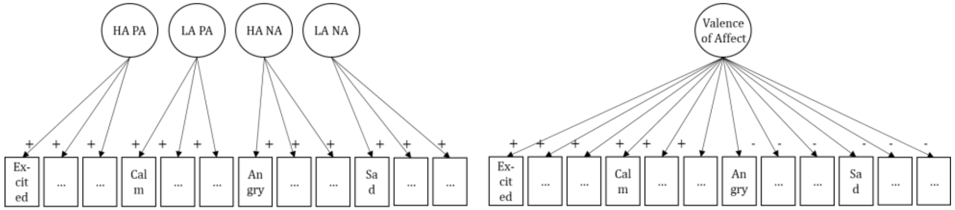
Substantive differences

- Ability to differentiate between emotions
  - High differentiators: Label in a differentiated and context dependent way
  - Low differentiators: Less specific emotional experiences



Existing Approaches Were Limited

- Only tested if invariance across subjects OR if invariance across time was violated
- Assumed that the number and nature of factors are the same



- No pinpointing for which subjects and time-points the MM differs
- No insights into what the MM differences look like

Latent Markov Factor Analysis

- Vogelsmeier et al. 2019
- Latent Markov model:
  - Latent class model that allows for transitions
  - Initial state probabilities
  - Transition probabilities
  - Probabilities may depend on individual- or time-point-specific covariates
- Exploratory factor analysis per state
  - State-specific intercepts, loadings, and unique variances
  - States may differ regarding all levels of invariance, thus, also regarding number of factors !!
  - For observations within the same state, invariance holds

$$\mathbf{y}_{itk} = \tau_k + \mathbf{\Lambda}_k \boldsymbol{\eta}_{it} + \epsilon_{itk},$$

Estimation

- Can be estimated with a FIML or three-step approach

LATENT GOLD

```
install.packages("devtools")
library("devtools")
install_github("LeonieVm/lmfa@0.1.3")
library("lmfa")
```

Step 1: Investigating measurement models

Step 1a: Selecting the number of states and factors

- run `step1()` with the modelselection option and store the results (e.g., into the object "`modelselection`")
- inspect the models using `summary(modelselection)` and rerun non-converged models using `step1()` without the model selection option
- choose best model(s) based on the BIC and CHull using `plot(modelselection)` and `chull_lmfa(modelselection)`
- store the best model(s) (e.g., into the object "`measurementmodel`")

Step 1b: Interpreting the measurement models

- use `summary(measurementmodel)` to obtain the measurement model parameters from the chosen model(s)
- investigate state-specific (obliquely rotated and standardized) loadings
- investigate state-specific intercepts
- investigate state-specific unique variances

Step 1c: Attach factor scores to the dataset

- store the outcome of `factorscores_lmfa(ESM, measurementmodel)` into a new dataset

Step 2: Obtaining state assignments & classification errors

- run `step2()` to calculate classification errors, posterior state-membership probabilities, and modal state assignments and store the results (e.g., into the object "`classification`")
- obtain the results with `summary(classification)`
- if desired, attach posterior state-membership probabilities and the modal state assignments to the dataset by storing `classification$data` into a new dataset

Step 3: Investigating transition model

Step 3a: Selecting the covariates for the transition model

- run `step3()` using the posterior state-membership probabilities of the `classification` object from `step2()` and all covariates of interest and store the results (e.g., into the object "`transitionmodel`")
- inspect Wald test results with `summary(transitionmodel)` and look at the p-values to decide which covariates should be included in the final transition model

Step 3b: Interpreting the transition model

- use `summary(transitionmodel)` to obtain the transition model parameters and probabilities for covariates being equal to their sample means
- use `probabilities(transitionmodel)` to obtain initial state and transition probabilities for any covariate value (and interval length) of interest

Step 3c: Updating state assignments & investigating state memberships

- if desired, attach posterior state-membership probabilities and the modal state assignments to the dataset by storing `transitionmodel$data` into a new dataset
- use `invariance(transitionmodel)` to investigate for which subjects within- and between-person invariance holds
- use `plot(transitionmodel)` to investigate subjects' individual transitions

Data (simulated for the *lmfa* package)

- Long format
- Column indicating time between previous and current observation
- Columns with indicator items
- Explanatory variables

	id	deltaT	negativeEvent	intervention	Interested	Joyful	Determined	Calm
1	1	0.00	53	0	45	16	8	75
2	1	0.56	37	0	52	42	35	50
3	1	1.04	55	0	71	80	70	78
4	1	1.81	59	0	62	77	75	94
5	1	0.80	73	0	27	40	46	17
6	1	2.45	49	0	55	53	18	45

Step 1: Investigating Measurement Models

- The state-specific MMs are estimated …
- …while disregarding the transitions and the covariate effects on these transitions.
- What is the best model in terms of the number of factors and states?
  - Model selection

Step 1: Investigating Measurement Models

```
modelselection <- step1(data = ESM,
                        indicators = c(
                          "Interested","Joyful","Determined","Calm",
                          "Lively","Enthusiastic","Relaxed","Cheerful",
                          "Content","Energetic","Upset","Gloomy",
                          "Sluggish","Anxious","Bored","Irritated",
                          "Nervous","Listless"),
                       modelselection = TRUE,
                       n_state_range = 1:4,
                       n_fact_range = 2:3)
```

Step 1: Investigating Measurement Models

```
summary(modelselection)

##                 LL      BIC convergence n par
## [323]  -353166.8 708485.3           1    254
## [333]  -353149.0 708602.3           1    272
## [3232] -353085.0 708940.1           1    327
## [3233] -353067.8 709058.2           0    345
## [3333] -353047.8 709170.6           0    363
## [3222] -353316.0 709249.7           1    309
## [322]  -353855.3 709709.8           1    236
## [33]   -354421.0 710375.3           1    181
## [2222] -353976.8 710418.8           1    291
## [32]   -355010.3 711401.4           1    163
## [222]  -355095.1 712037.0           1    218
## [22]   -356377.4 713983.1           1    145
## [3]    -361759.6 724281.6           1     90
## [2]    -363744.0 728098.0           1     72
```

Step 1: Investigating Measurement Models

Configural invariance clearly violated!

```
## Obliquely rotated standardized loadings:
##
##              S1F1  S1F2  S1F3   S2F1  S2F2   S3F1  S3F2  S3F3
## Interested   0.66  0.04  0.00   0.68  0.01   0.57 -0.01  0.02
## Joyful       0.60  0.02  0.02   0.65 -0.01   0.88  0.01  0.06
## Determined   0.37  0.03 -0.55   0.61  0.00   0.84  0.02 -0.01
## Calm         0.37 -0.58 -0.01   0.59  0.00   0.18 -0.15  0.82
## Lively       0.63  0.03  0.03   0.65  0.00   0.88 -0.01  0.01
## Enthusiastic 0.65 -0.01  0.02   0.64  0.00   0.89  0.02  0.00
## Relaxed      0.64  0.02  0.00   0.64  0.01   0.16 -0.14  0.85
## Cheerful     0.63  0.07  0.01   0.63 -0.01   0.91  0.01  0.02
## Content      0.61  0.00  0.03   0.67  0.02   0.93  0.02  0.01
## Energetic    0.64 -0.01  0.00   0.63 -0.01   0.90  0.05 -0.01
## Upset        0.09  0.62 -0.01   0.00  0.53   0.03  0.83 -0.03
## Gloomy      -0.24  0.39  0.44  -0.01  0.53   0.02  0.82 -0.01
## Sluggish     0.07 -0.01  0.73  -0.01  0.50  -0.29  0.34  0.77
## Anxious      0.09  0.70 -0.02   0.00  0.52   0.05  0.79 -0.01
## Bored        0.07 -0.01  0.74  -0.01  0.52   0.04  0.47 -0.04
## Irritated    0.06  0.51 -0.05   0.01  0.58   0.04  0.85 -0.02
## Nervous      0.08  0.73 -0.04   0.00  0.51   0.03  0.74  0.01
## Listless     0.06 -0.05  0.73   0.01  0.54   0.02  0.46 -0.03
```

Step 1: Investigating Measurement Models

```
## Intercepts:
##
##                 S1    S2    S3
## Interested     49.24 61.46 51.98
## Joyful         48.92 61.12 49.95
## Determined     46.60 61.20 50.35
## Calm           46.25 61.14 54.76
## Lively         49.29 60.85 50.57
## Enthusiastic   48.99 61.16 50.24
## Relaxed        49.00 61.12 54.90
## Cheerful       49.03 61.02 50.42
## Content        49.39 60.84 49.98
## Energetic      49.35 60.90 50.41
## Upset          44.12 26.54 36.42
## Gloomy         45.88 27.09 35.93
## Sluggish       44.95 26.54 33.26
## Anxious        45.81 26.48 35.83
## Bored          44.98 26.75 29.94
## Irritated      43.48 26.69 35.66
## Nervous        46.39 26.50 35.94
## Listless       45.35 26.84 29.67
```

```
## Unique variances:
##
##                 S1     S2     S3
## Interested     273.26 53.37  96.43
## Joyful         273.82 48.67  92.81
## Determined     261.88 49.93  92.70
## Calm           265.99 51.75  99.17
## Lively         286.09 48.20 104.04
## Enthusiastic   257.06 50.09 107.07
## Relaxed        270.54 49.55  99.75
## Cheerful       284.69 50.47  83.05
## Content        271.52 41.15  92.38
## Energetic      271.12 53.24  95.55
## Upset          278.71 46.13  92.89
## Gloomy         256.03 46.46  73.85
## Sluggish       245.57 51.70  82.24
## Anxious        276.61 45.65  87.14
## Bored          253.52 47.69 103.11
## Irritated      267.30 44.56  84.99
## Nervous        261.57 49.30  86.21
## Listless       269.07 47.29  92.10
```

Step 2: Obtaining State Assignments

- Each observation is assigned to the state with the highest state-membership probability.
- The inherent classification uncertainty is calculated.
- Relevant for obtaining unbiased estimates for the transition model

```
classification <- step2(data = ESM_fs, model = measurementmodel323)
```

Step 3: Investigating Transition Model

- The MMs (i.e., the factor parameters) are kept fixed
- The transitions between the states are estimated (while correcting for step 2's assignment uncertainty)

Step 3: Investigating Transition Model

```
transitionmodel <- step3(data = ESM,
                         identifier = "id",
                         n_state = 3,
                         postprobs =
                           classification$classification_posteriors[,-1],
                         timeintervals  = "deltaT",
                         initialCovariates = NULL,
                         transitionCovariates =
                           c("intervention", "negativeEvent"))
```

```
Wald tests:

                  Wald df p-value
intervention  213.3821  6       0
negativeEvent  55.7629  6       0
```

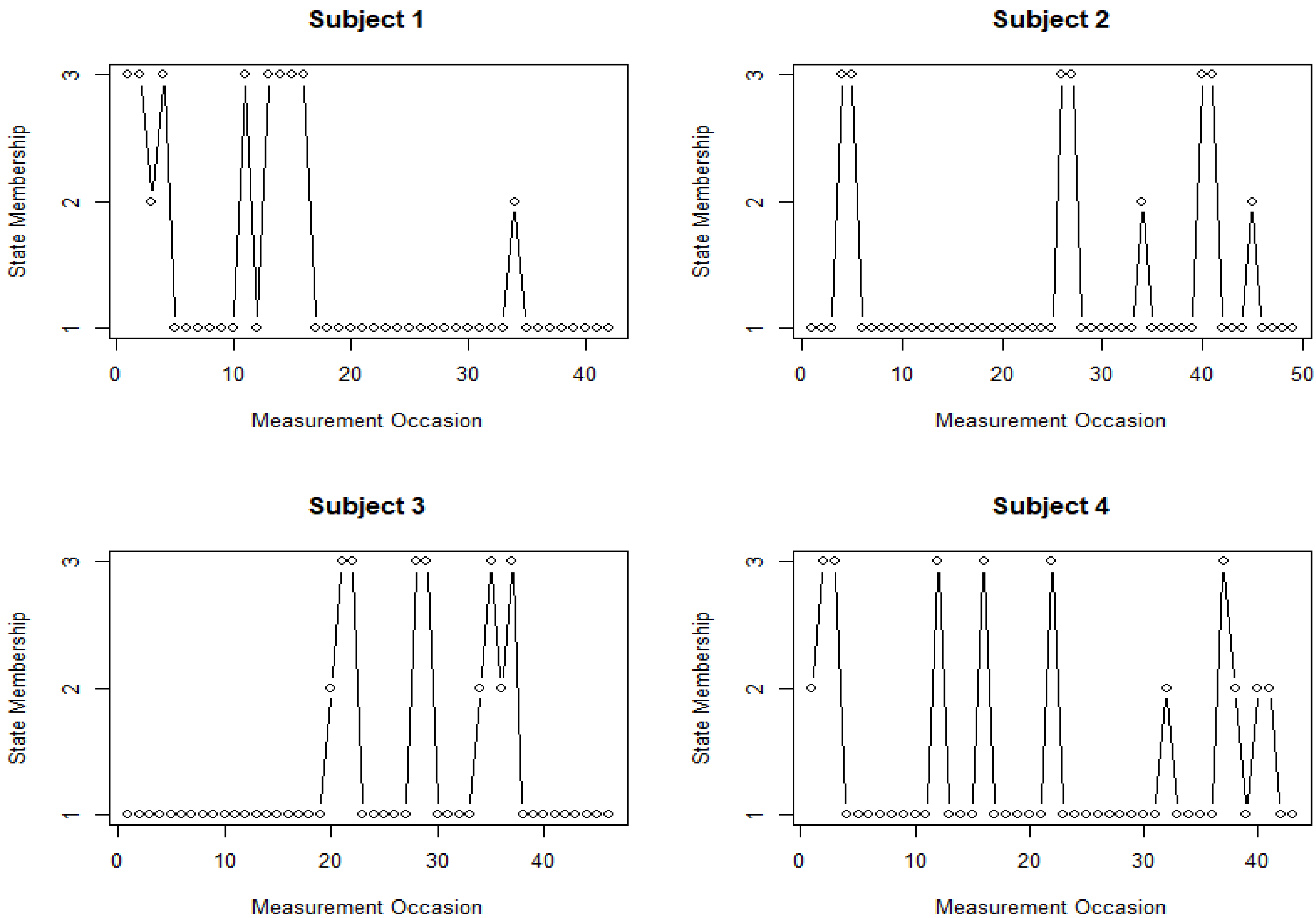
Step 3: Investigating Transition Model

```
probabilities(model = transitionmodel,
              deltaT = 1,
              initialCovariateScores = NULL,
              transitionCovariateScores = c(0, 49.65))
```

```
## 1. Initial state probabilities:
##
## (no covariates defined)
##
##   S1   S2   S3
## 0.42 0.34 0.24
##
## 2. Transition probabilities:
##
## interval length: 1
## intervention score: 0
## negativeEvent score: 49.65
##
##      S1   S2   S3
## S1 0.84 0.07 0.08
## S2 0.37 0.44 0.19
## S3 0.54 0.08 0.39
```

```
## 1. Initial state probabilities:
##
## (no covariates defined)
##
##   S1   S2   S3
## 0.42 0.34 0.24
##
## 2. Transition probabilities:
##
## interval length: 1
## intervention score: 1
## negativeEvent score: 49.65
##
##      S1   S2   S3
## S1 0.71 0.16 0.13
## S2 0.14 0.66 0.20
## S3 0.23 0.15 0.61
```

Step 3: Investigating Transition Model



Accounting for Non-Invariance?

- Obtain dataset with state memberships and state-specific factor scores

```
ESM_fs <- factorscores_lmfa(data = ESM, model = measurementmodel323)
ESM_fs_cl <- classification$data
```

- Continue with one state only
- Accept that invariance does not hold and focus on substantively interesting results:
    - E.g., learning about situations in which emotion differentiation is reduced

Discussion

References

- Adolf, J., Schuurman, N. K., Borkenau, P., Borsboom, D., & Dolan, C. V. (2014). Measurement invariance within and between individuals: a distinct problem in testing the equivalence of intra- and inter-individual model structures. *Frontiers in Psychology, 5*, 1–14. doi:10.3389/fpsyg.2014.00883
- Lai, M. H. C. (2021). Composite reliability of multilevel data: It's about observed scores and construct meanings. Psychological Methods, 26, 90–102. doi:https://doi.org/10.1037/met0000287
- McNeish, D., Mackinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2021). Measurement in Intensive Longitudinal Data. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–16. doi:10.1080/10705511.2021.1915788
- Vogelsmeier, L. V. D. E., Vermunt, J. K., van Roekel, E., & De Roover, K. (2019). Latent Markov factor analysis for exploring measurement model changes in time-intensive longitudinal studies. Structural Equation Modeling: A Multidisciplinary Journal, 26, 557–575. doi:10.1080/10705511.2018.1554445