

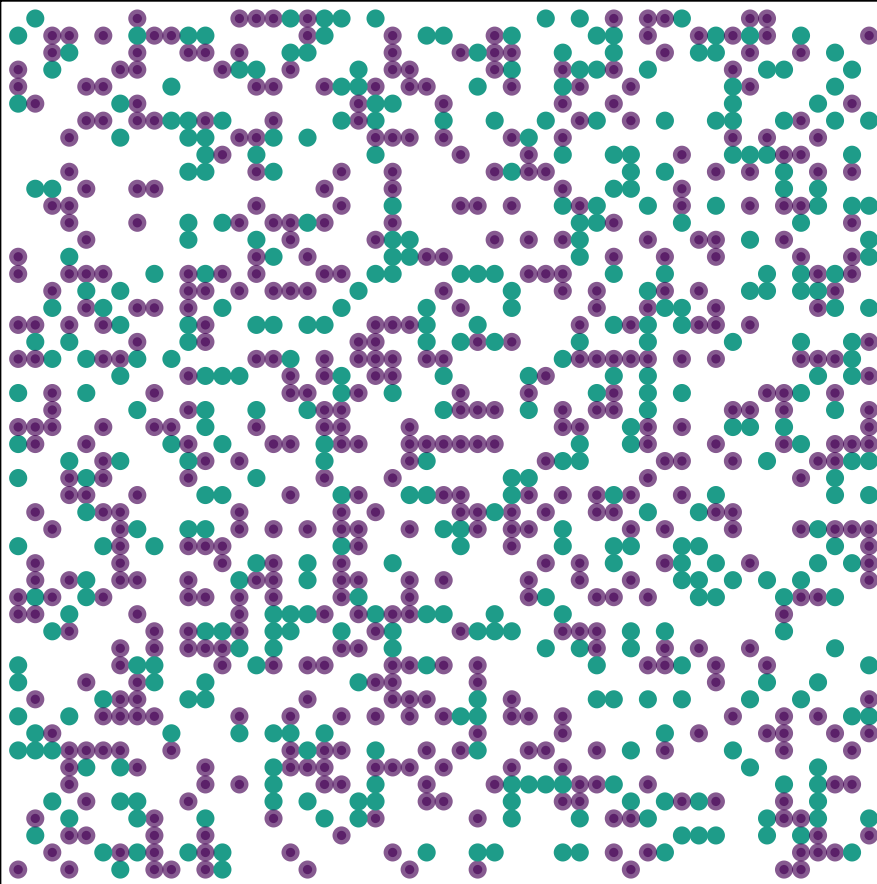
THE PSYCHOMETRICS OF INTERVAL RESPONSES

Matthias Kloft

DAGSTAT 2025, Berlin

MOTIVATING EXAMPLE

What is the percentage of purple dots?

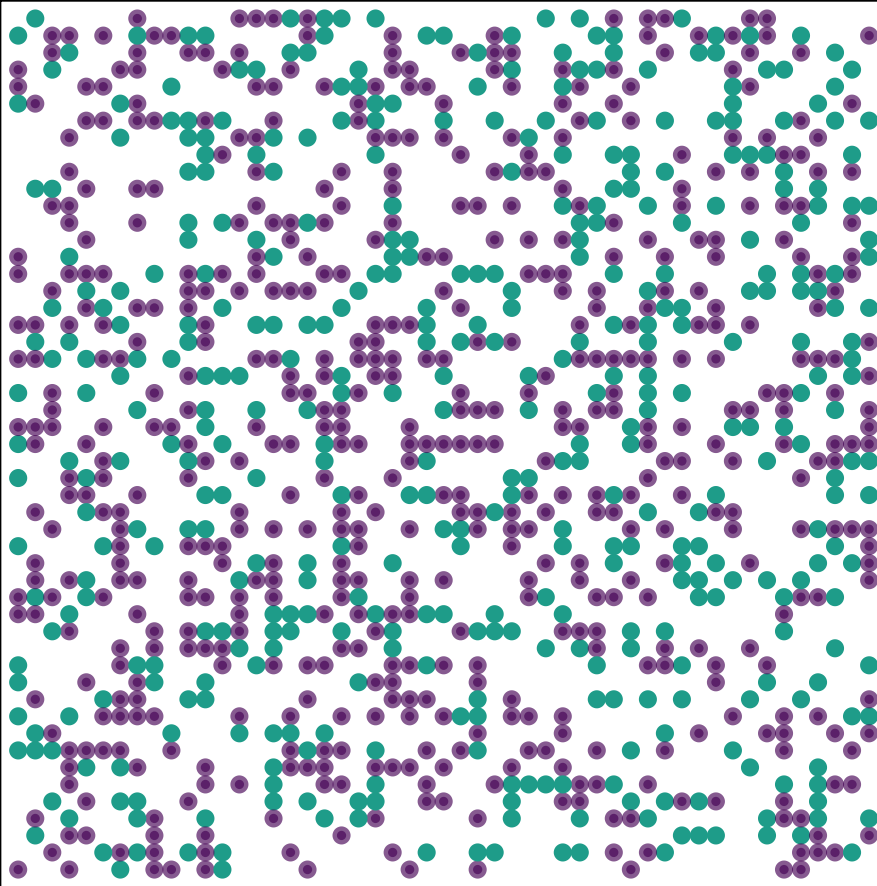


Think of a:

- Best guess
- Lower bound
- Upper bound

MOTIVATING EXAMPLE

What is the percentage of purple dots?



True percentage = 60%

INTERVAL RESPONSES

Single-range slider / visual analog scale (**VAS**)



Dual-range slider (**DRS**)

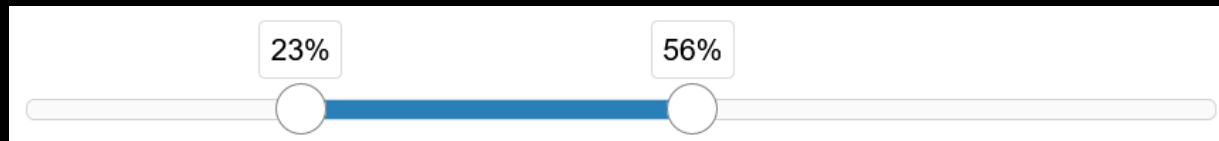


INTERVAL RESPONSES

Single-range slider / visual analog scale (**VAS**)



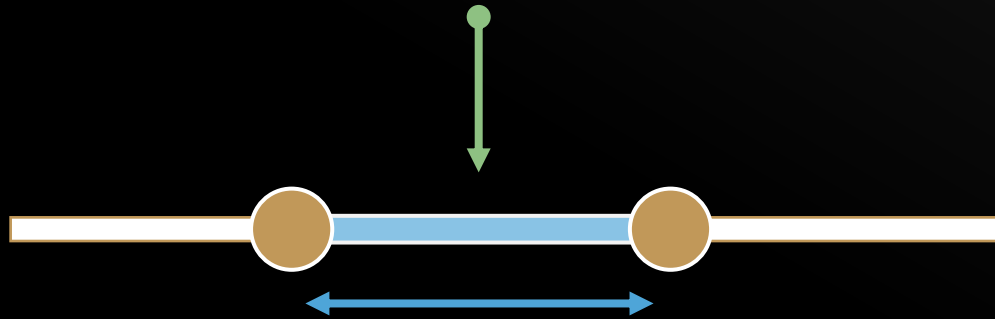
Dual-range slider (**DRS**)



noUISlider JavaScript range slider (Gersen, 2024)

INTERVAL RESPONSES: DESCRIPTIVE VALUES

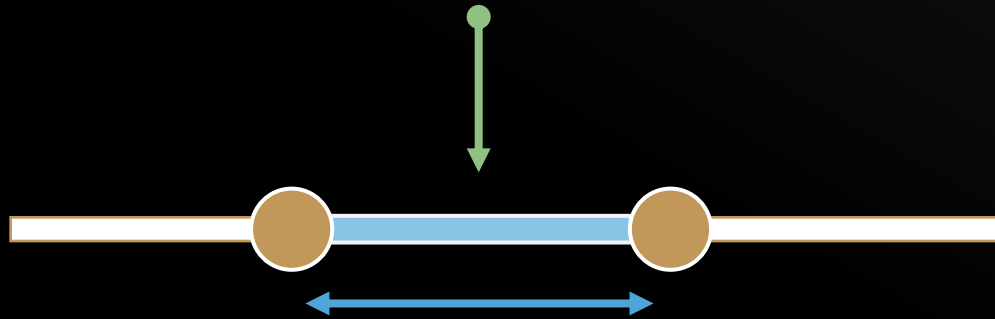
DRS Location: $\frac{y^{(L)} + y^{(U)}}{2}$



DRS Width: $y^{(U)} - y^{(L)}$

INTERVAL RESPONSES: DESCRIPTIVE VALUES

DRS Location: $\frac{y^{(L)} + y^{(U)}}{2}$



DRS Width: $y^{(U)} - y^{(L)}$

There are better representations for modeling!

INTERVAL RESPONSES

Variability:

- Self-ratings, stimuli

Uncertainty:

- Influenced by expertise and confidence
- Estimation (e.g., forecasting)

Ambiguity / Plausibility:

- No clear-cut true answer (e.g., verbal quantifiers like “seldom” or “likely”)
- Meta-uncertainty: plausible risks

TOPICS OF THE TALK

1. Convergence of DRS and VAS
2. Test-retest reliability of the DRS
3. Discriminant validity regarding the DRS Width

CONVERGENT VALIDITY: VAS & DRS LOCATION

Kloft et al. (2023), Kloft et al. (2024)

KLOFT ET AL., 2023: CROSS-SECTIONAL DESIGN

- 2 Response Formats: VAS & DRS
- 2 Extraversion Personality Scales, One per Response Format

KLOFT ET AL., 2023: CROSS-SECTIONAL DESIGN

- Modeling of extraversion scores via joint IRT models
 - **VAS**: Beta Response Model (Noel & Dauvier, 2007)
 - **DRS**: Dirichlet Dual Response Model
 - (Kloft et al., 2023)
 - **Correlation** of latent extraversion scores for VAS and DRS measures via Bayesian hierarchical model with MVN prior

Correlation between latent extraversion score
measured by VAS and DRS:

$$\rho = .87$$

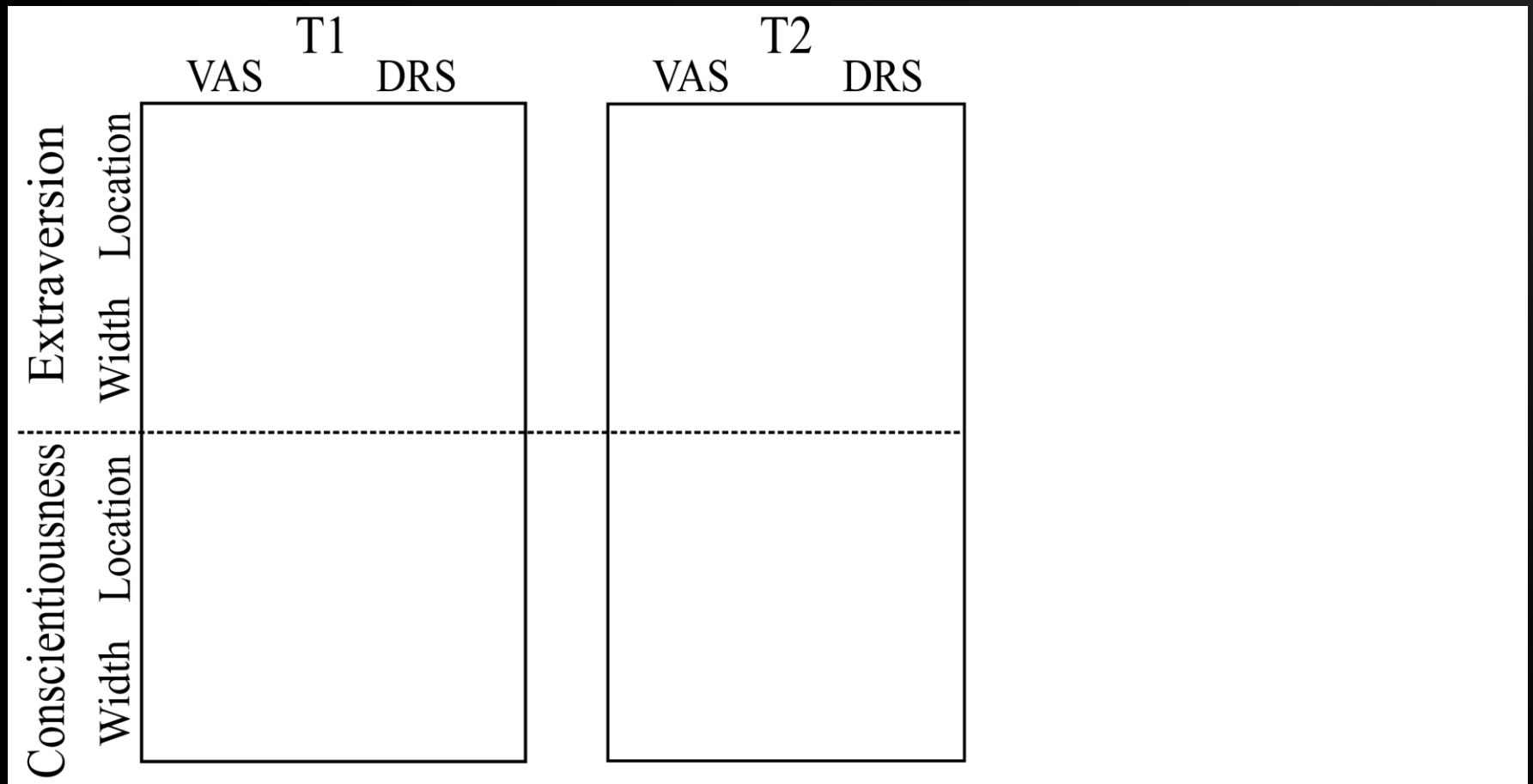
KLOFT ET AL., 2024: MTMM STUDY DESIGN

- 2 Time points (6-8 weeks apart)
- 2 Response Formats: VAS & DRS
- 2 Personality Scales: Extraversion & Conscientiousness person-descriptive adjectives
- Random split and assignment of items to response formats

KLOFT ET AL., 2024: MTMM STUDY DESIGN

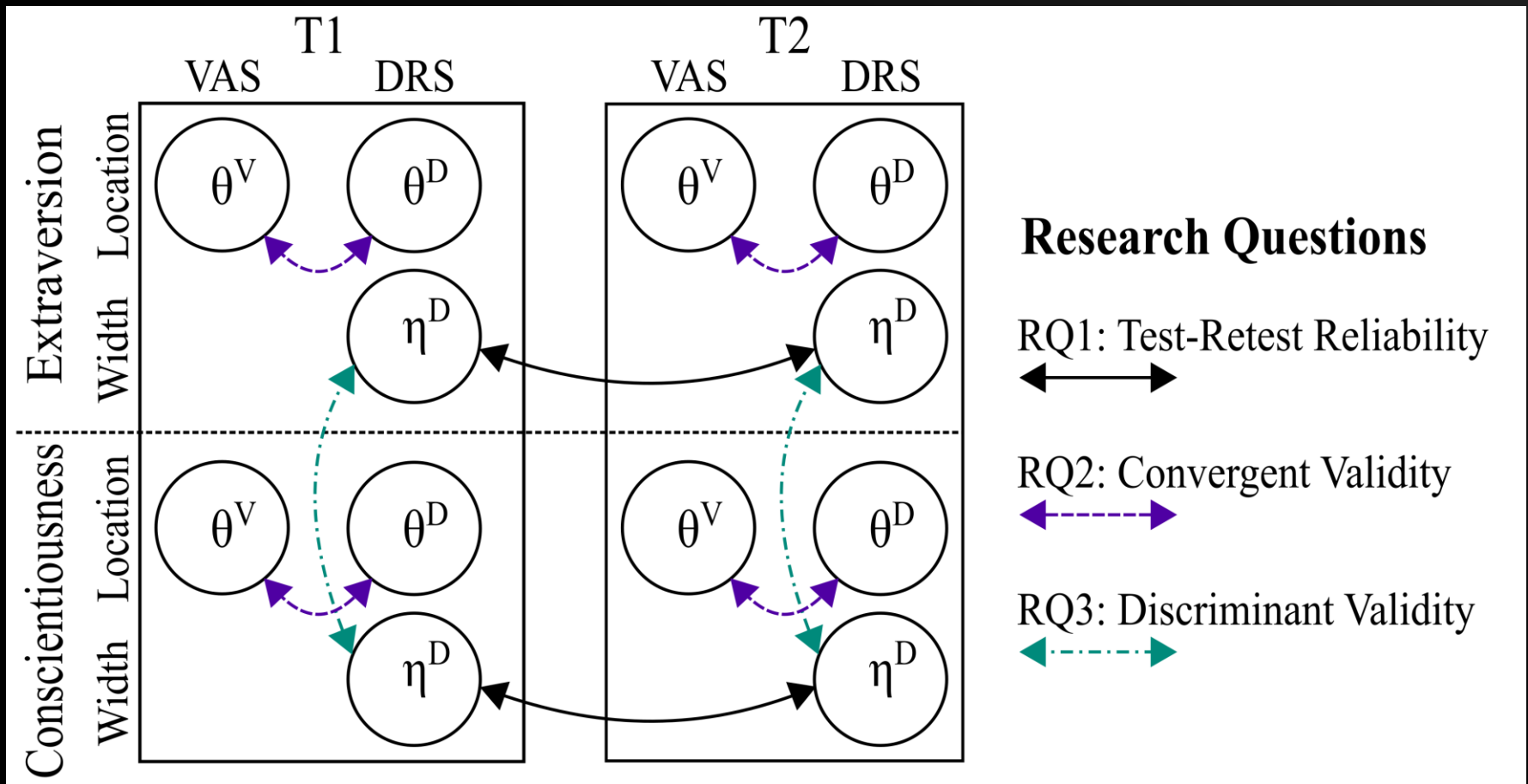
- Modeling of personality scores via joint IRT models
 - **VAS**: Beta Response Model (Noel & Dauvier, 2007)
 - **DRS**: Dirichlet Dual Response Model
 - (Kloft et al., 2023)
 - **Correlations** of latent person scores via Bayesian hierarchical model with MVN prior

KLOFT ET AL., 2024: MTMM STUDY DESIGN



2 Time points, 2 Response Formats (VAS, DRS),
2 Scales (Extraversion & Conscientiousness)

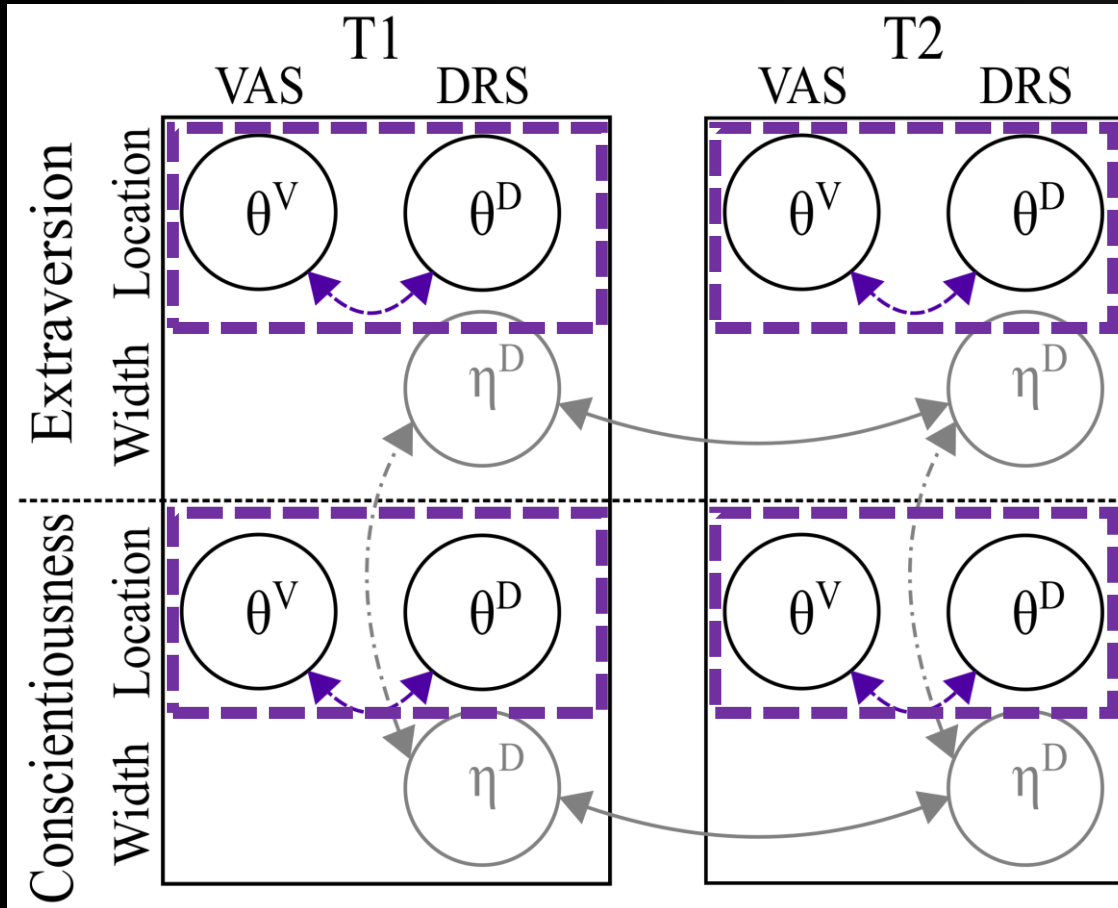
KLOFT ET AL., 2024: MTMM STUDY DESIGN



θ : scores for central tendency of personality

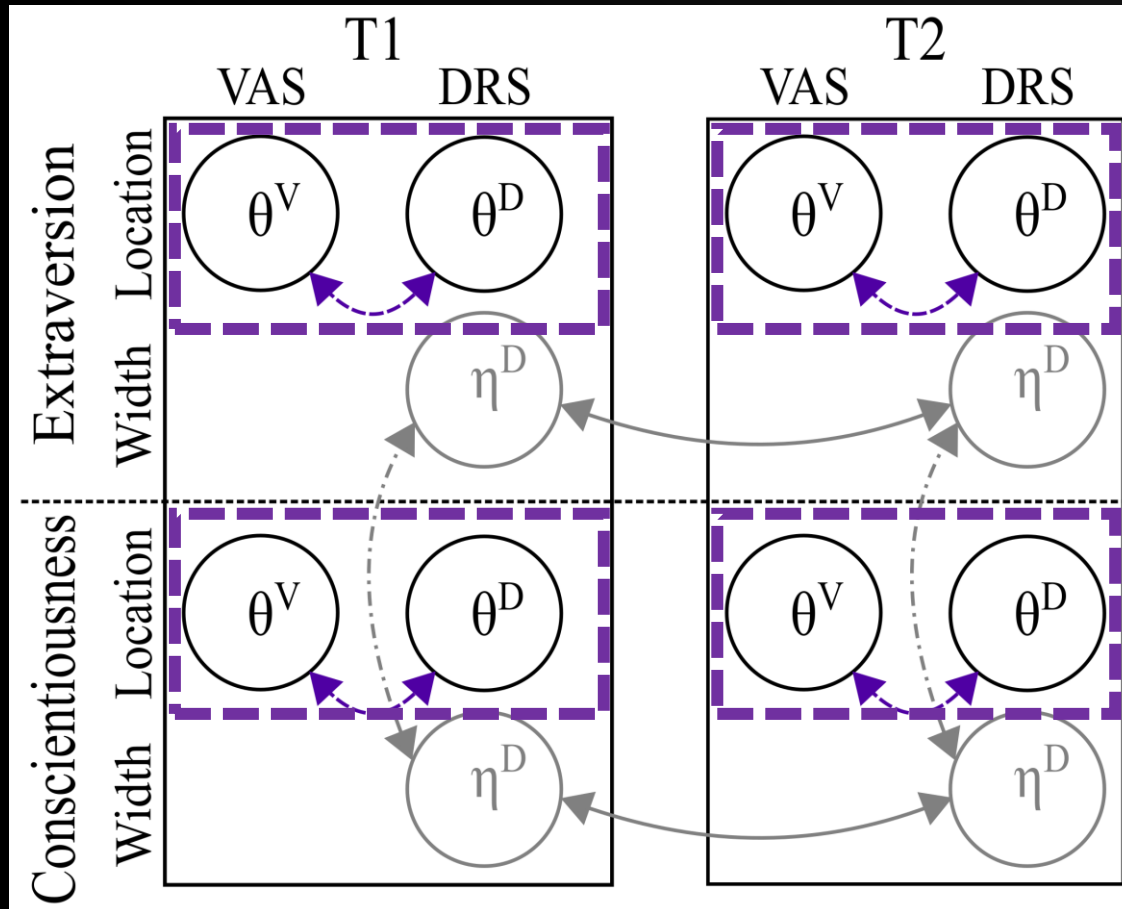
η : scores for perceived variability of personality

KLOFT ET AL., 2024: MTMM STUDY DESIGN



- Correlation between personality scores
- across response formats,
- for both scales,
- within each time point

KLOFT ET AL., 2024: CORRELATIONS



T1:

$$\hat{\rho} = .93$$

T2:

$$\hat{\rho} = .96$$

$$\hat{\rho} = .88$$

$$\hat{\rho} = .90$$

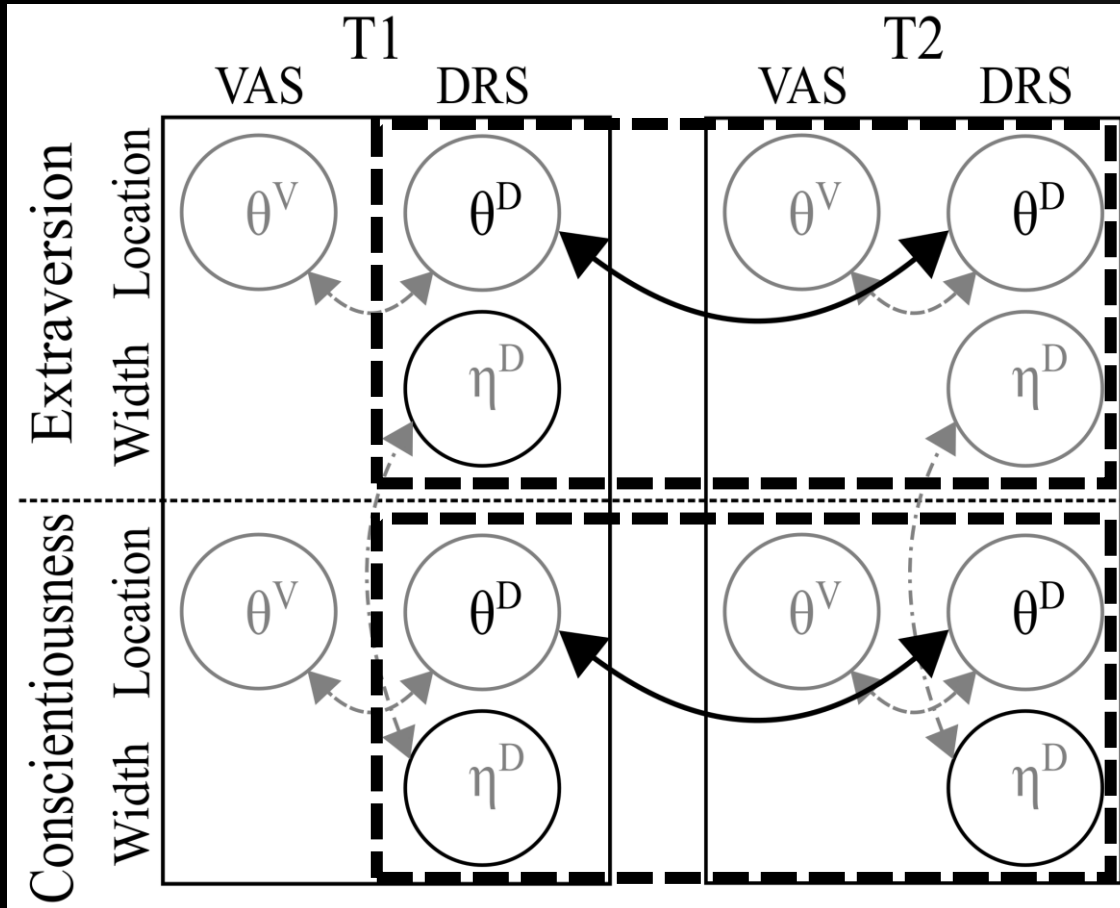
CONVERGENCE: VAS & DRS LOCATION

- Correlations between .87 and .93
- VAS and DRS Location measures may be used **equivalently** at least in this personality application

TEST-RETEST RELIABILITY

Kloft et al. (2024)

KLOFT ET AL., 2024: MTMM STUDY DESIGN

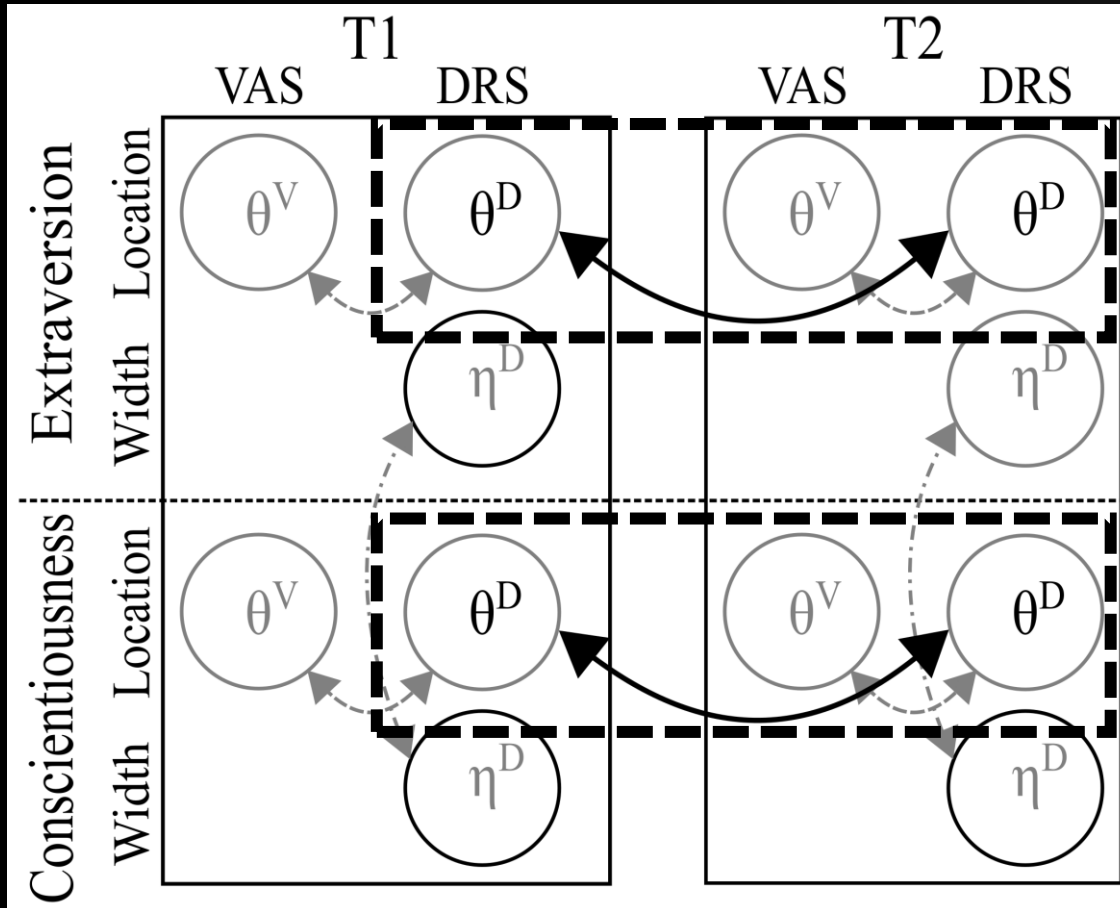


- Correlations of personality scores
- measured by DRS Location & Width
- across **time**

θ : scores for central tendency of personality

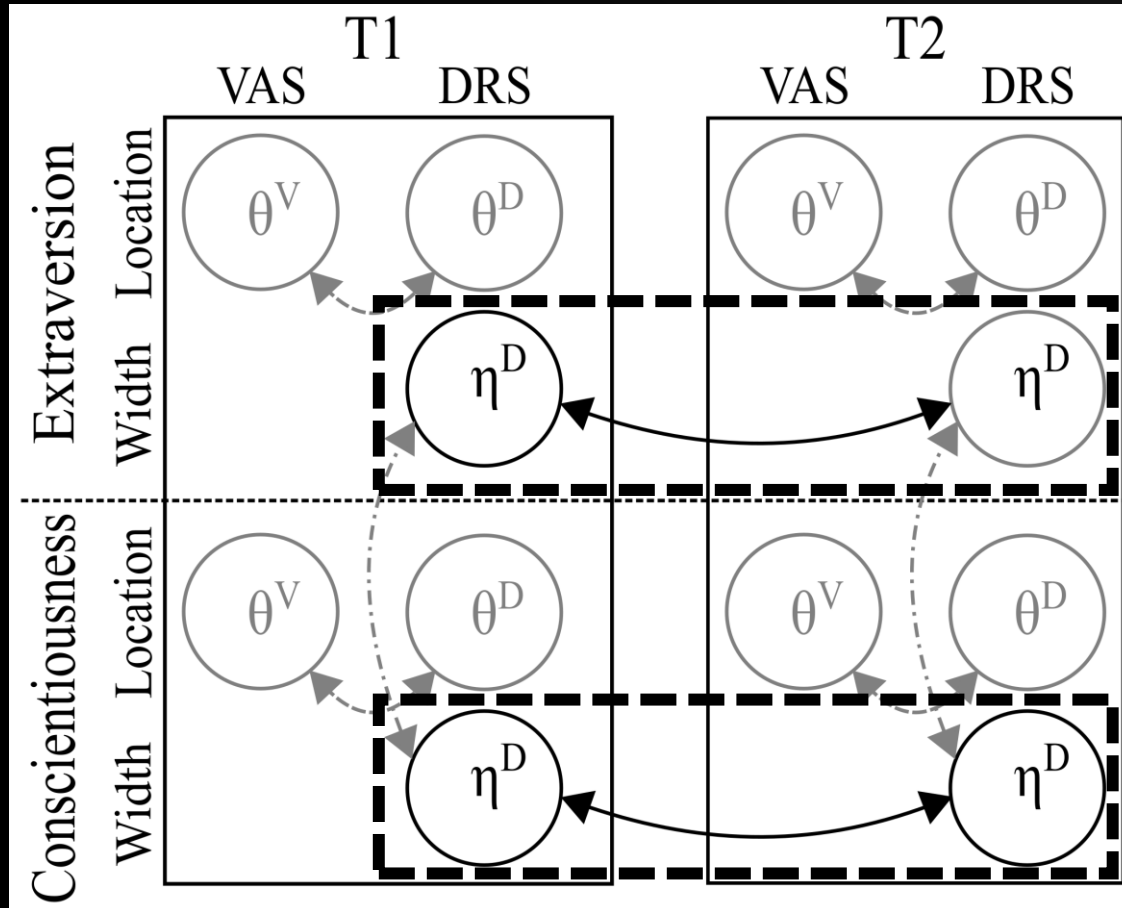
η : scores for perceived variability of personality

AUTOCORRELATION: DRS LOCATION



High reliability for DRS **Location** scores

AUTOCORRELATION: DRS WIDTH



Location:
 $\hat{\rho} = .92$

Width:
 $\hat{\rho} = .81$

$\hat{\rho} = .87$

$\hat{\rho} = .73$

Good reliability for DRS **Width** scores

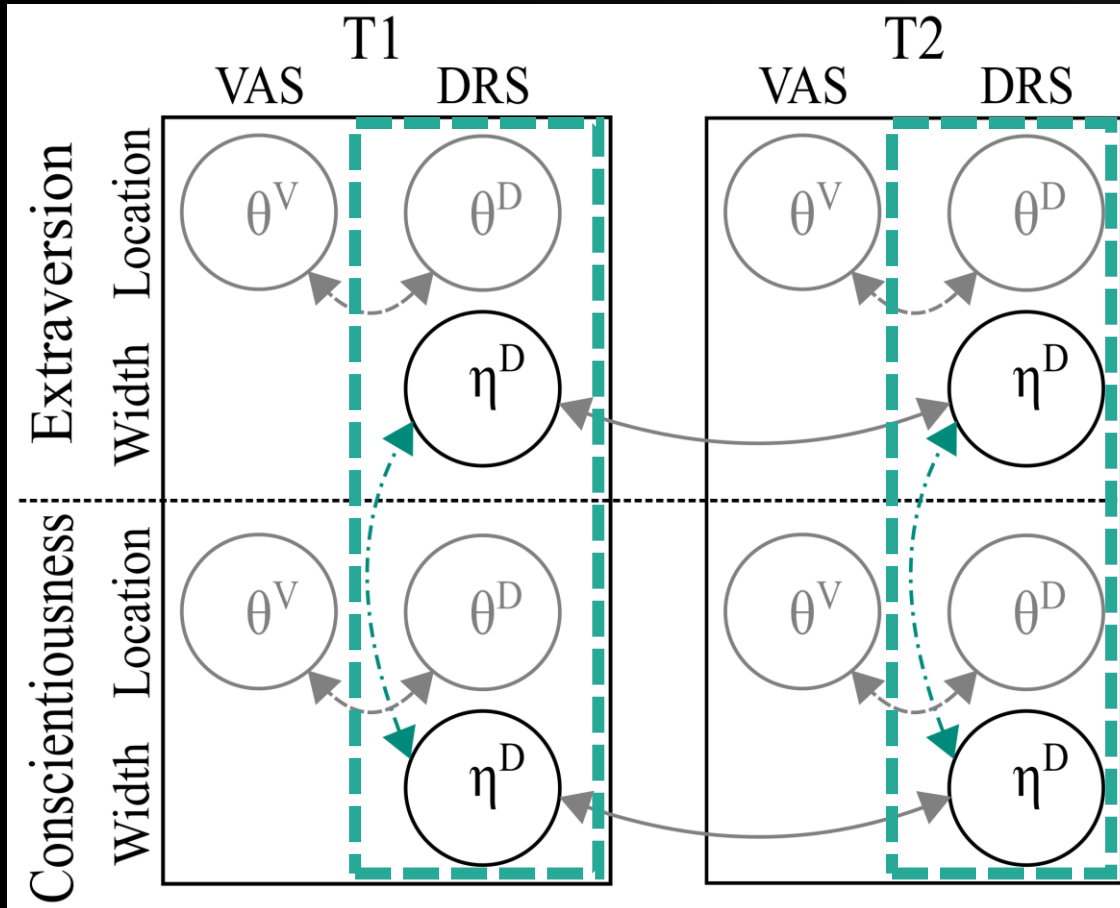
RELIABILITY: DRS LOCATION & WIDTH

- Test-retest reliability:
 - Good for DRS Location scores
 - Seems okay for DRS Width scores
- Caveat: this might be due to strong response biases
 - Investigate discriminant validity between different constructs

DISCRIMINANT VALIDITY: DRS WIDTH

Kloft et al. (2024), Kloft & Heck (2024)

KLOFT ET AL., 2024: MTMM STUDY DESIGN

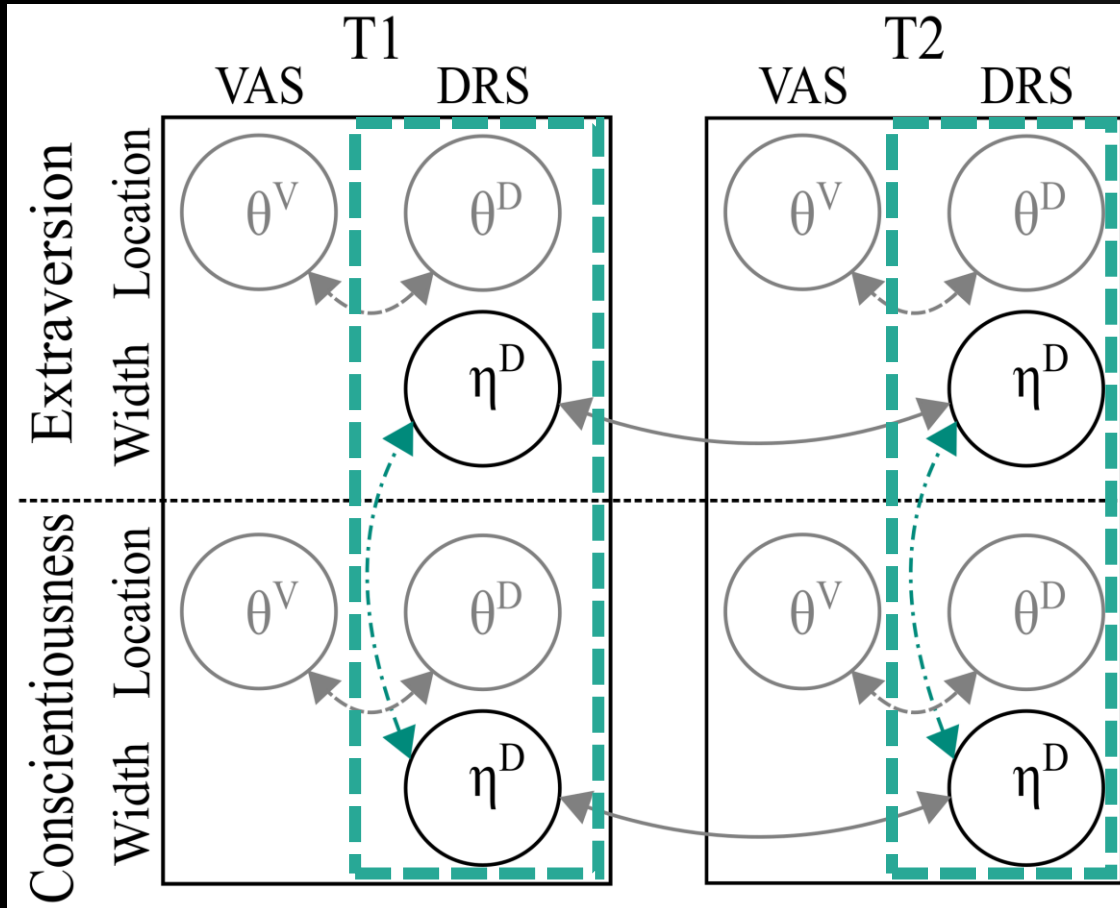


- Correlations of personality scores
- measured by DRS Location & Width
- across **personality scales**

θ : scores for central tendency of personality

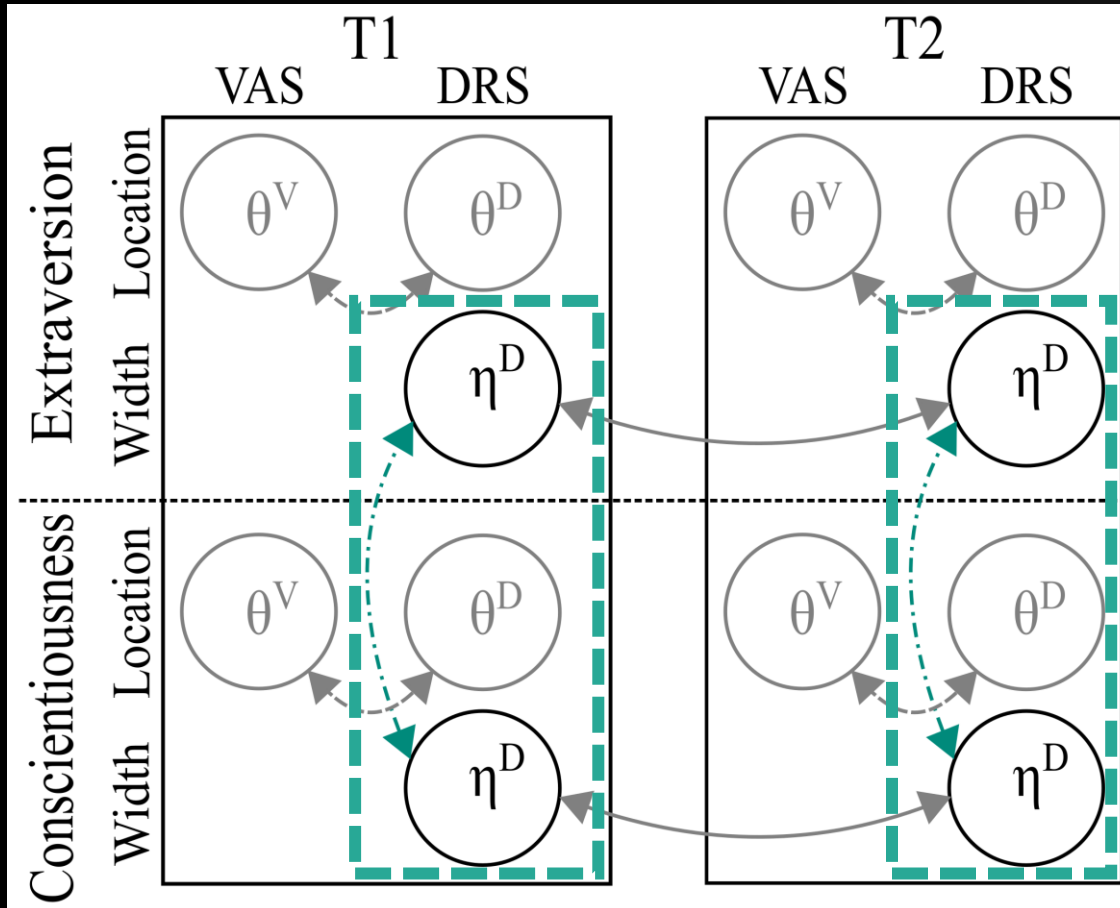
η : scores for perceived variability of personality

KLOFT ET AL., 2024: MTMM STUDY DESIGN



Different perceived variability for **different** traits?

KLOFT ET AL., 2024: CORRELATIONS

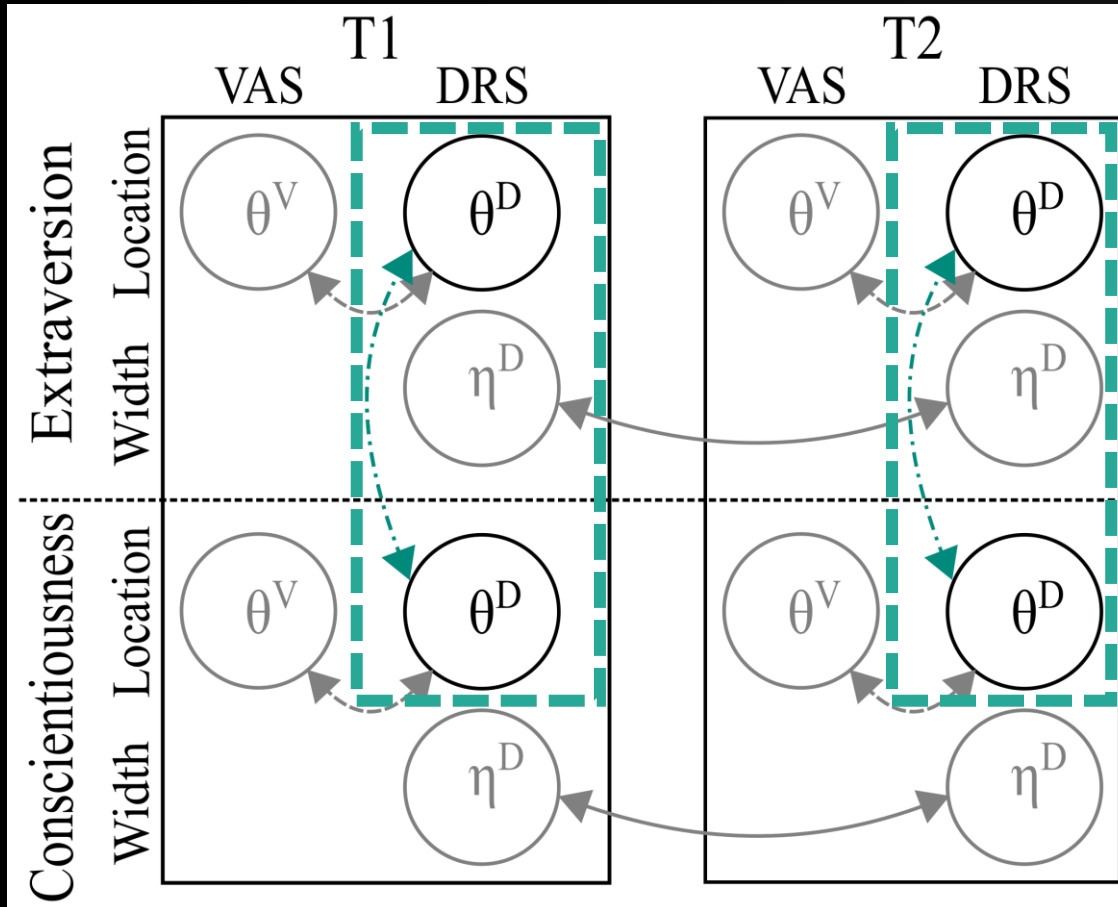


Width:
 $\hat{\rho} = .94$

$\hat{\rho} = .96$

Poor discrimination of DRS **Width** scores

KLOFT ET AL., 2024: CORRELATIONS



Location:

$$\hat{\rho} = .30$$

$$\hat{\rho} = .31$$

Width:

$$\hat{\rho} = .94$$

$$\hat{\rho} = .96$$

Reasonable correlation of DRS **Location** scores

KLOFT ET AL., 2024: CONCLUSIONS

What do DRS **Widths** measure?

- Personal **preference** for a particular width
 - Preference response style
- **Global trait** of perceived variability in personality?
- Limitation: only two constructs, both from the personality domain
- Follow-up Study: Kloft & Heck, 2024

KLOFT & HECK, 2024: AIM OF STUDY

- Focus on DRS **Widths**:
 - **Sensitivity** to different tasks?
 - Just a preference response style?
 - Factorial **dimensionality** across different tasks?

KLOFT & HECK, 2024: STUDY DESIGN

- Seven tasks of varying similarity
- Logit transformation of interval responses (Smithson & Broomel, 2024; see also additional slides)
- Factor Analyses: EFA, CFA
 - We use **only** the transformed **widths**
 - Transformation acts as a **link function**
 - Accounts for boundedness
 - Adjusts the width for the location

KLOFT & HECK, 2024: APPLICATIONS / TASKS

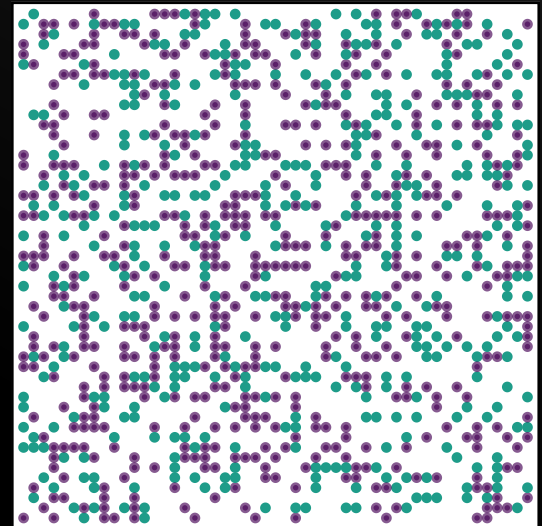
Extraversion & Conscientiousness:

Talkative = [75% – 93%] applicable

Color Dot Estimation:

% purple dots = [46% – 73%]

(true = 60%)



Election Forecasting: outcomes for 6 parties

- *Green party* = [10% – 19%]

KLOFT & HECK, 2024: APPLICATIONS / TASKS

Estimation of percentages:

- Death Causes: *heart diseases*
- Reasons for Incapability for work: *mental health*

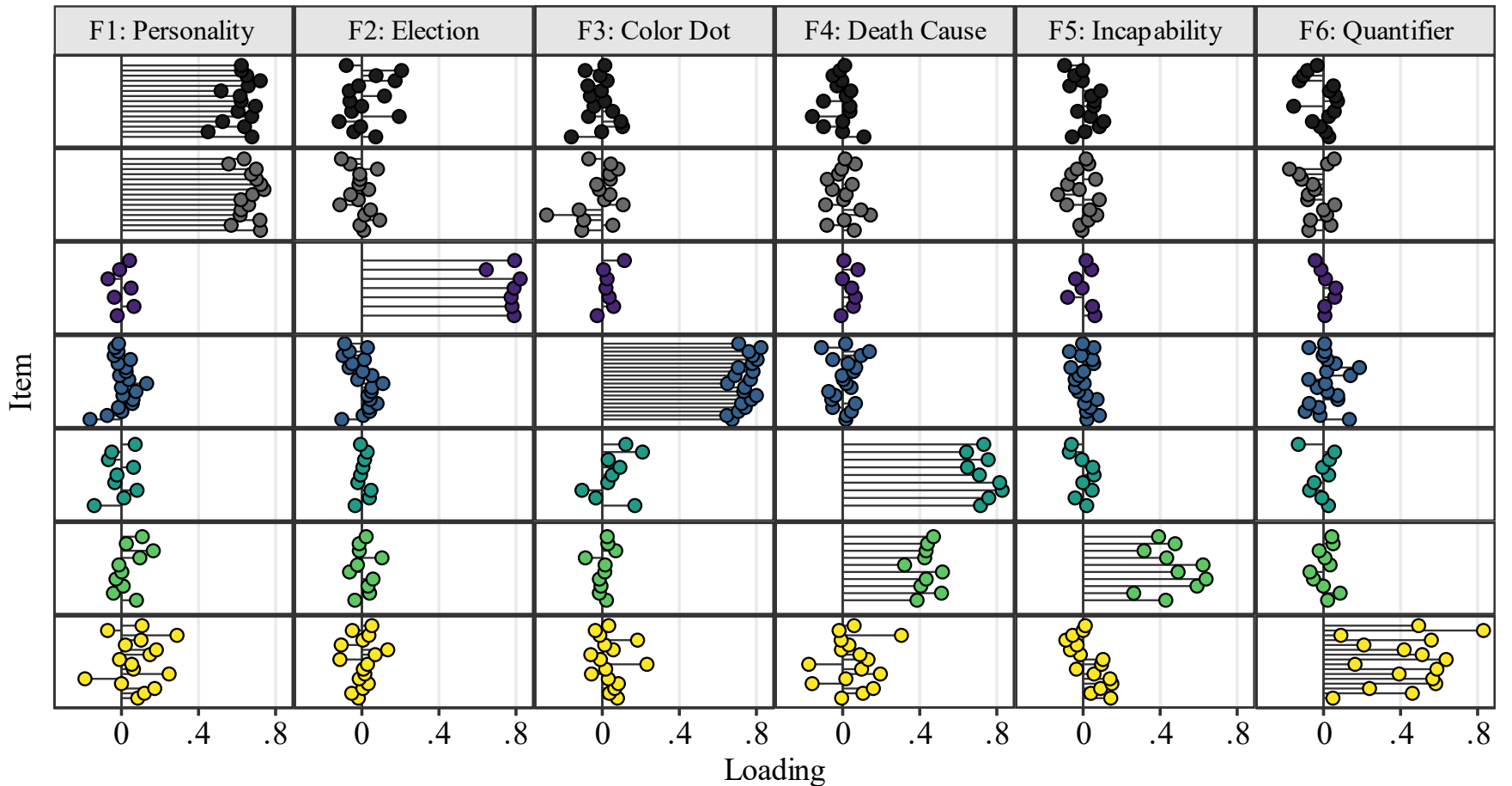
Judgment of verbal quantifiers as probabilities that a so described event would occur:

- *Seldom* = [5% – 30%]

INTERVAL WIDTH: EFA LOADINGS

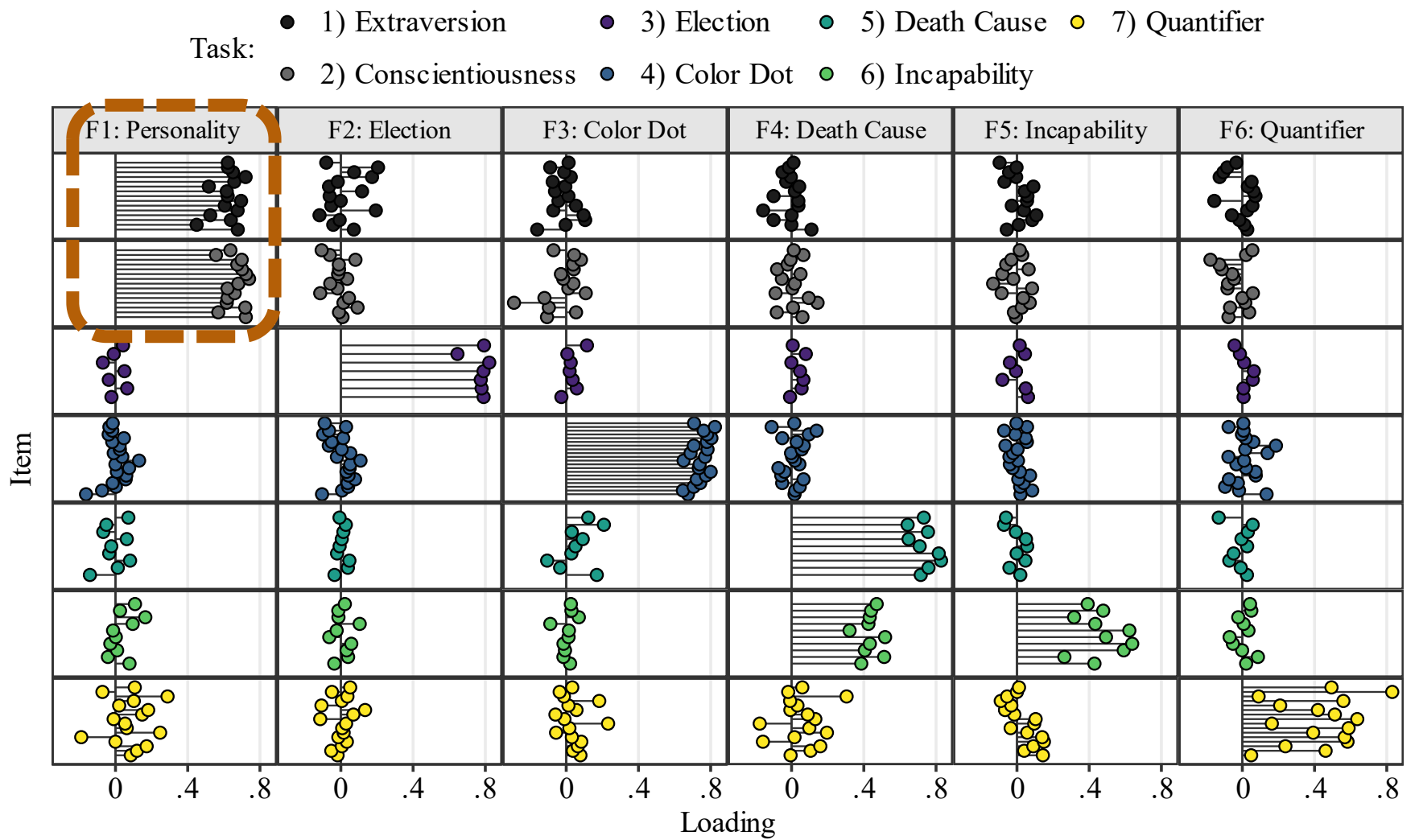
Task:

- 1) Extraversion
- 2) Conscientiousness
- 3) Election
- 4) Color Dot
- 5) Death Cause
- 6) Incapability
- 7) Quantifier



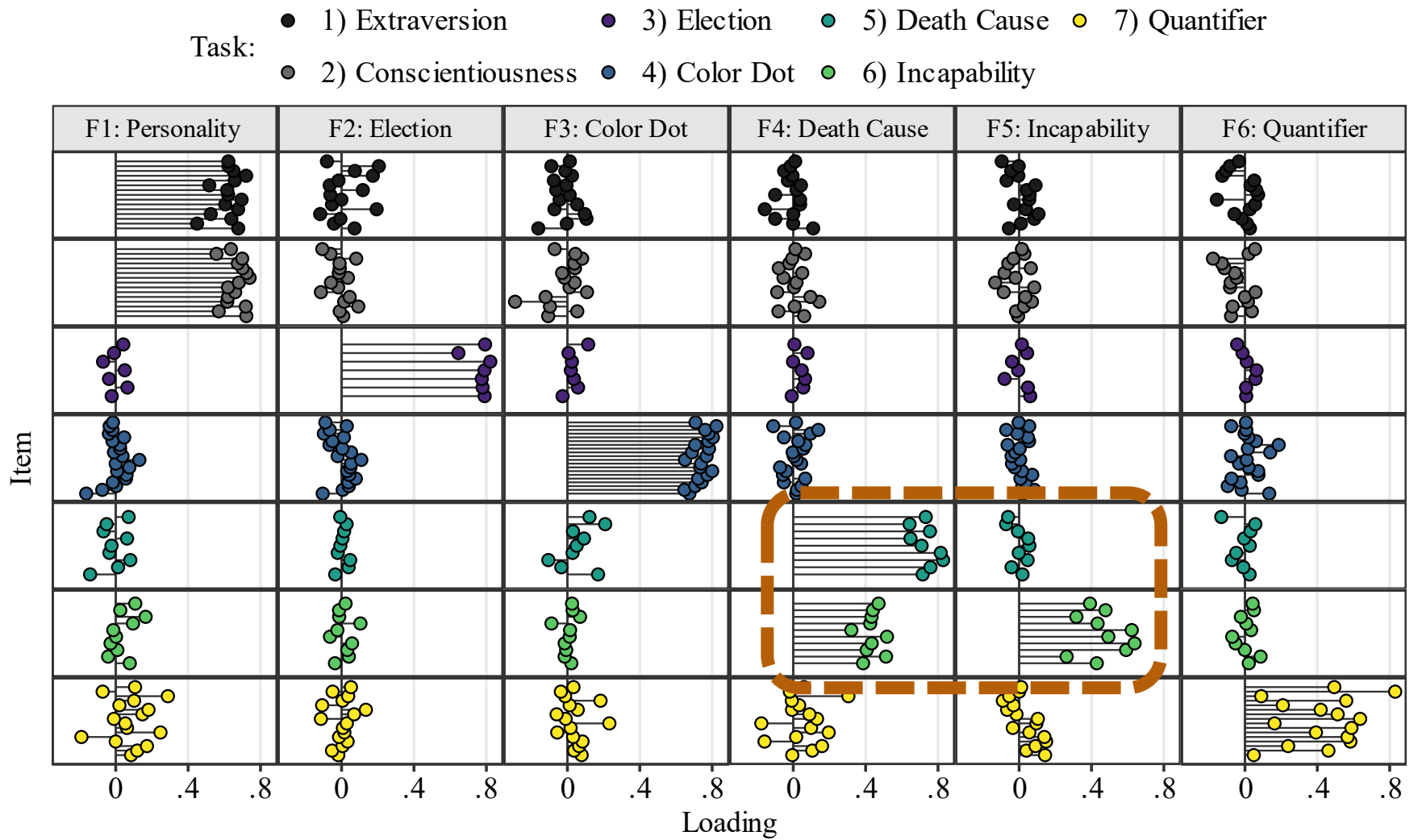
DRS **Width** only!

INTERVAL WIDTH: EFA LOADINGS



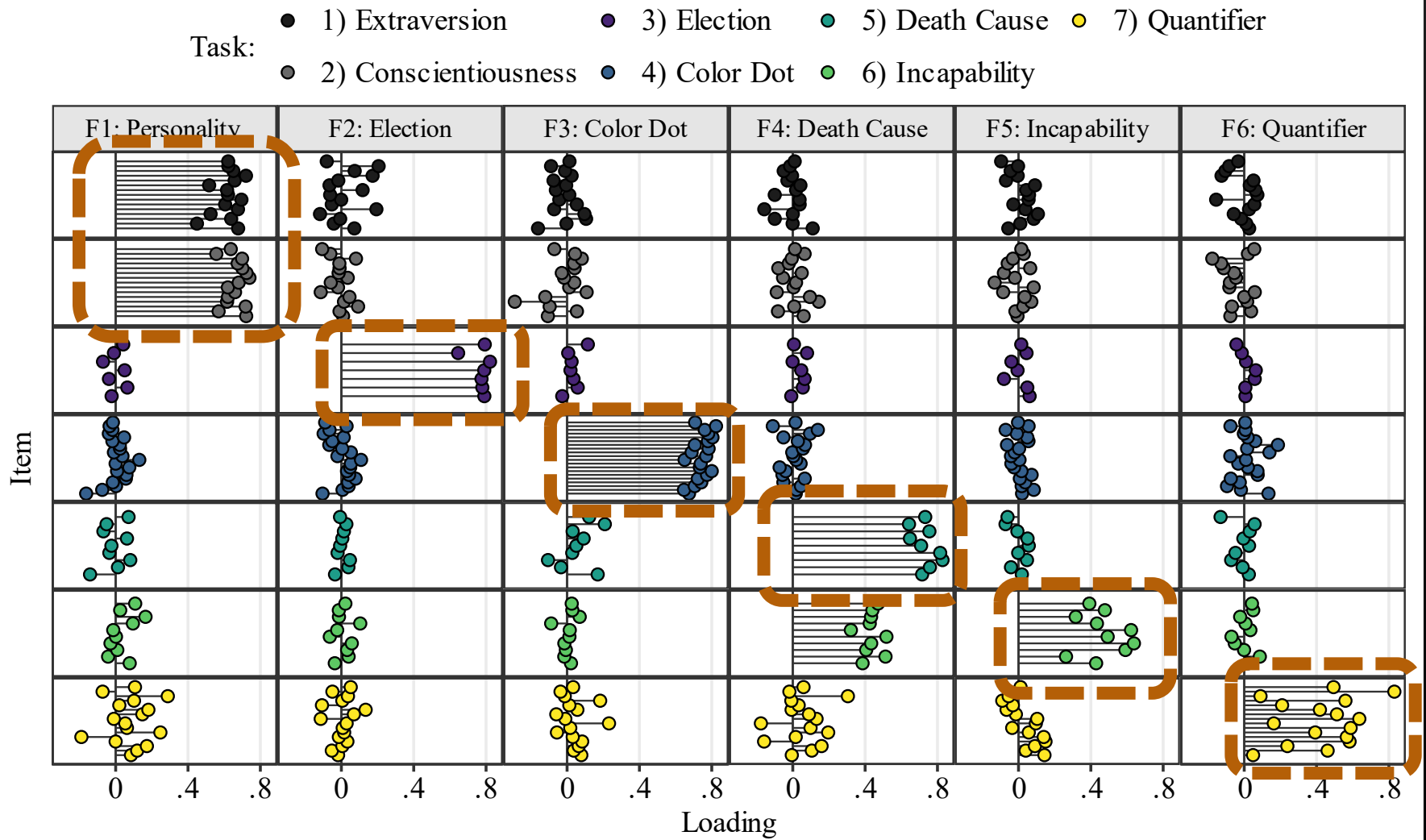
Common factor for Extraversion & Conscientiousness

INTERVAL WIDTH: EFA LOADINGS



High cross loadings: similar tasks

INTERVAL WIDTH: EFA LOADINGS

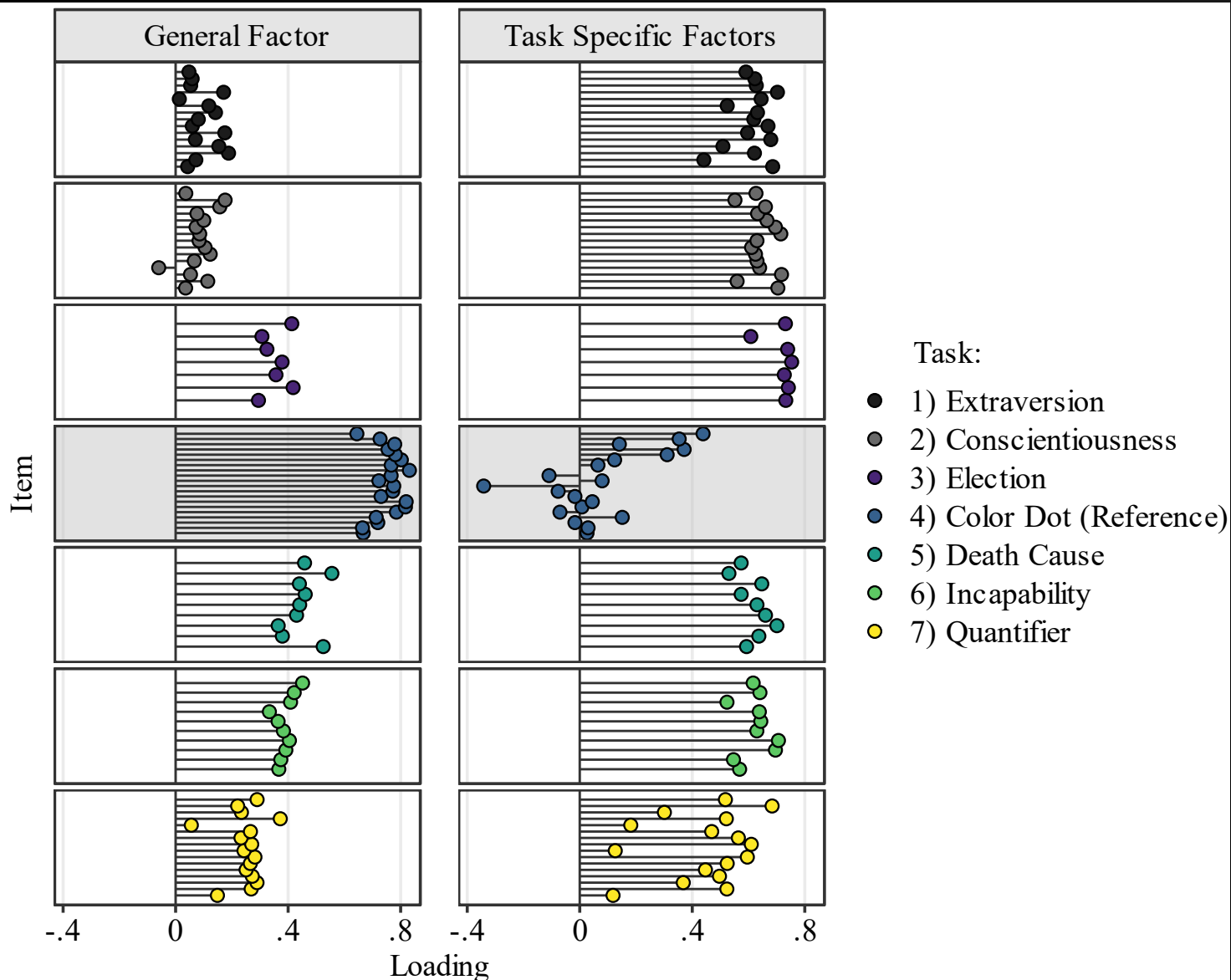


Factor structure **follows** task structure

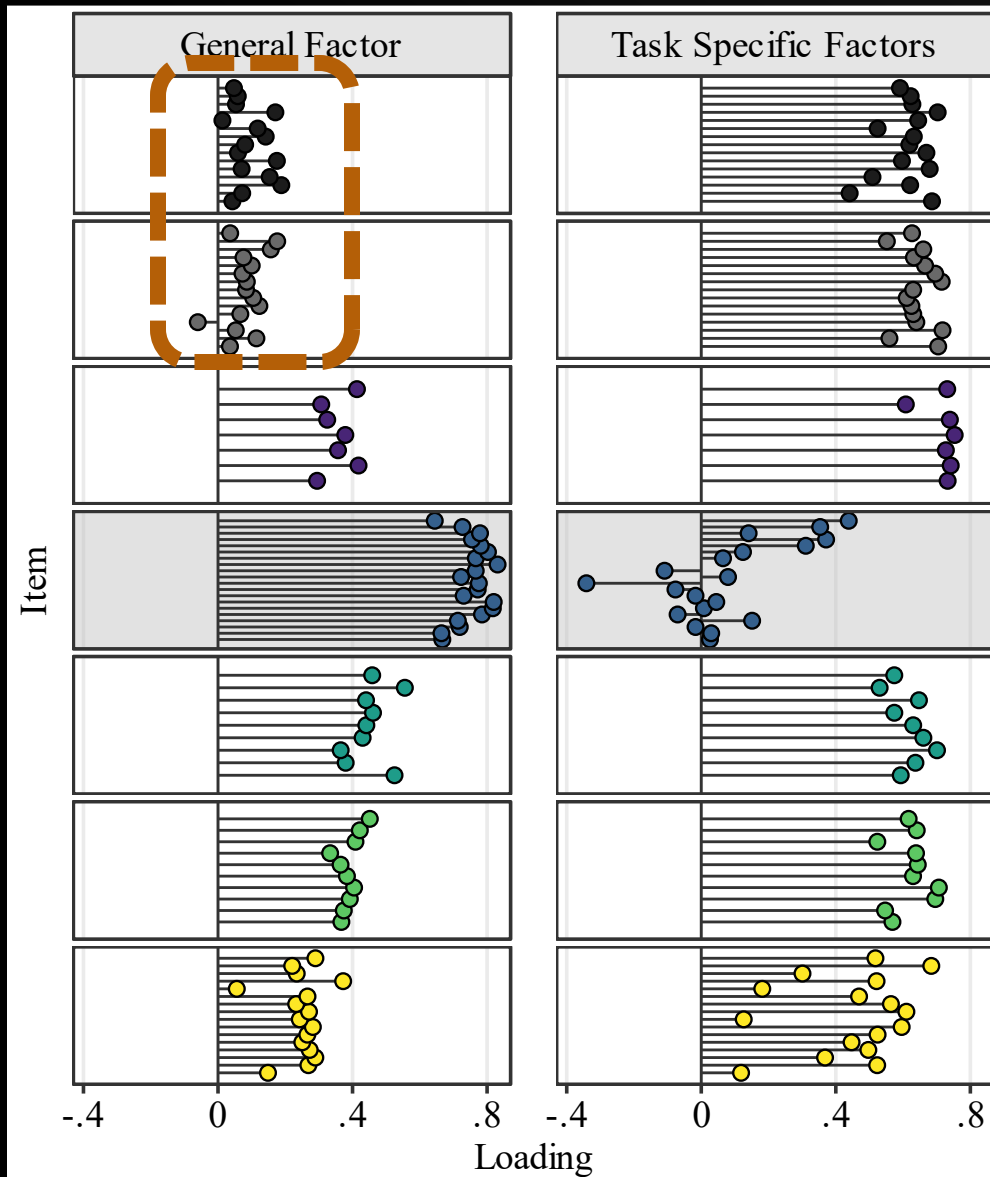
KLOFT & HECK, 2024: BIFACTOR MODEL

- Again only transformed DRS **Width**
- Bifactor- $(S.I-1)$ model (Eid et al., 2017)
- 6 **Specific** Factors
 - Assignment by theoretical task
- **General** Factor: preferred width / response style
 - Color dot estimation as reference task

LOADINGS: GENERAL VS. SPECIFIC FACTOR(S)

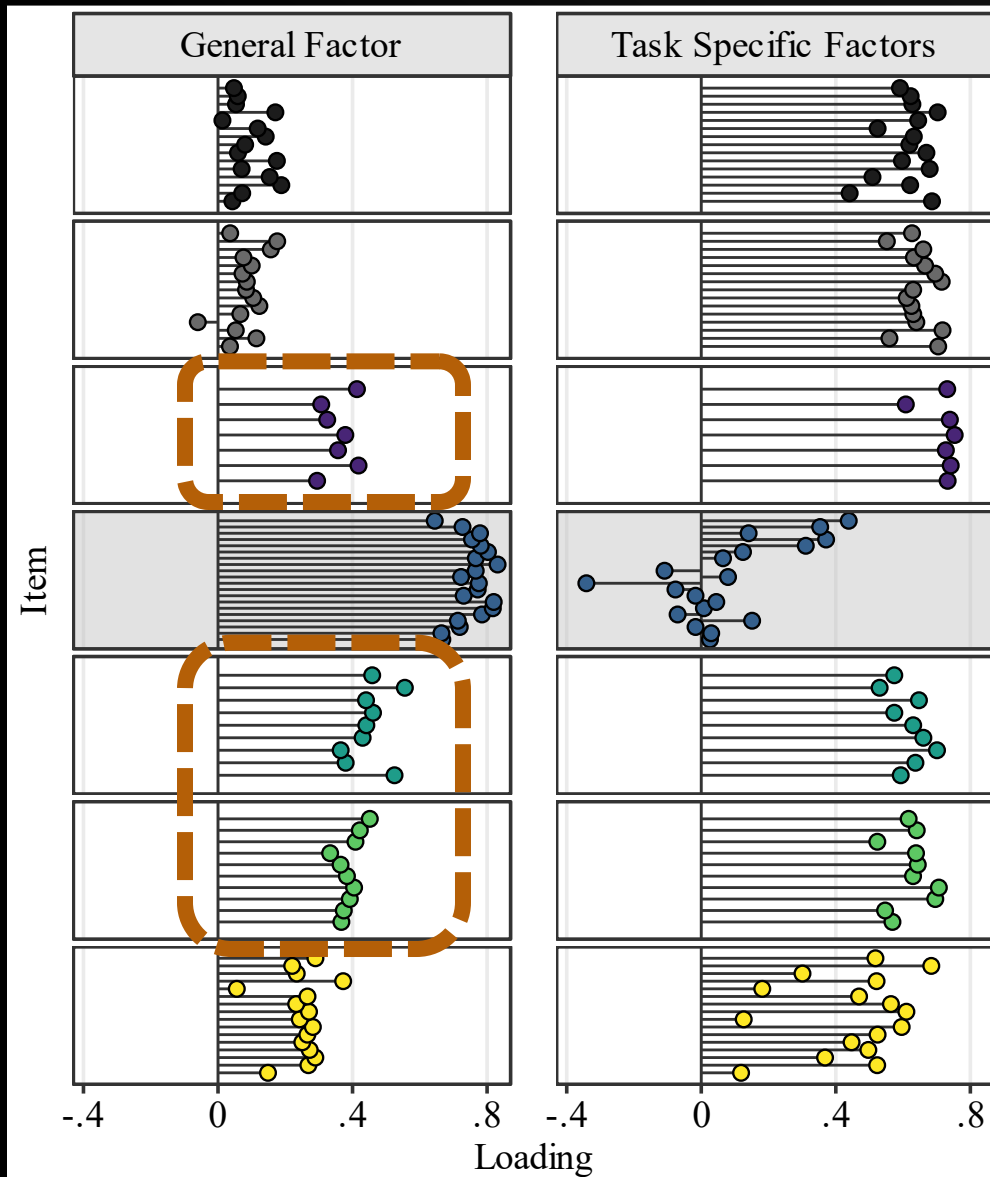


LOADINGS: GENERAL VS. SPECIFIC FACTOR(S)



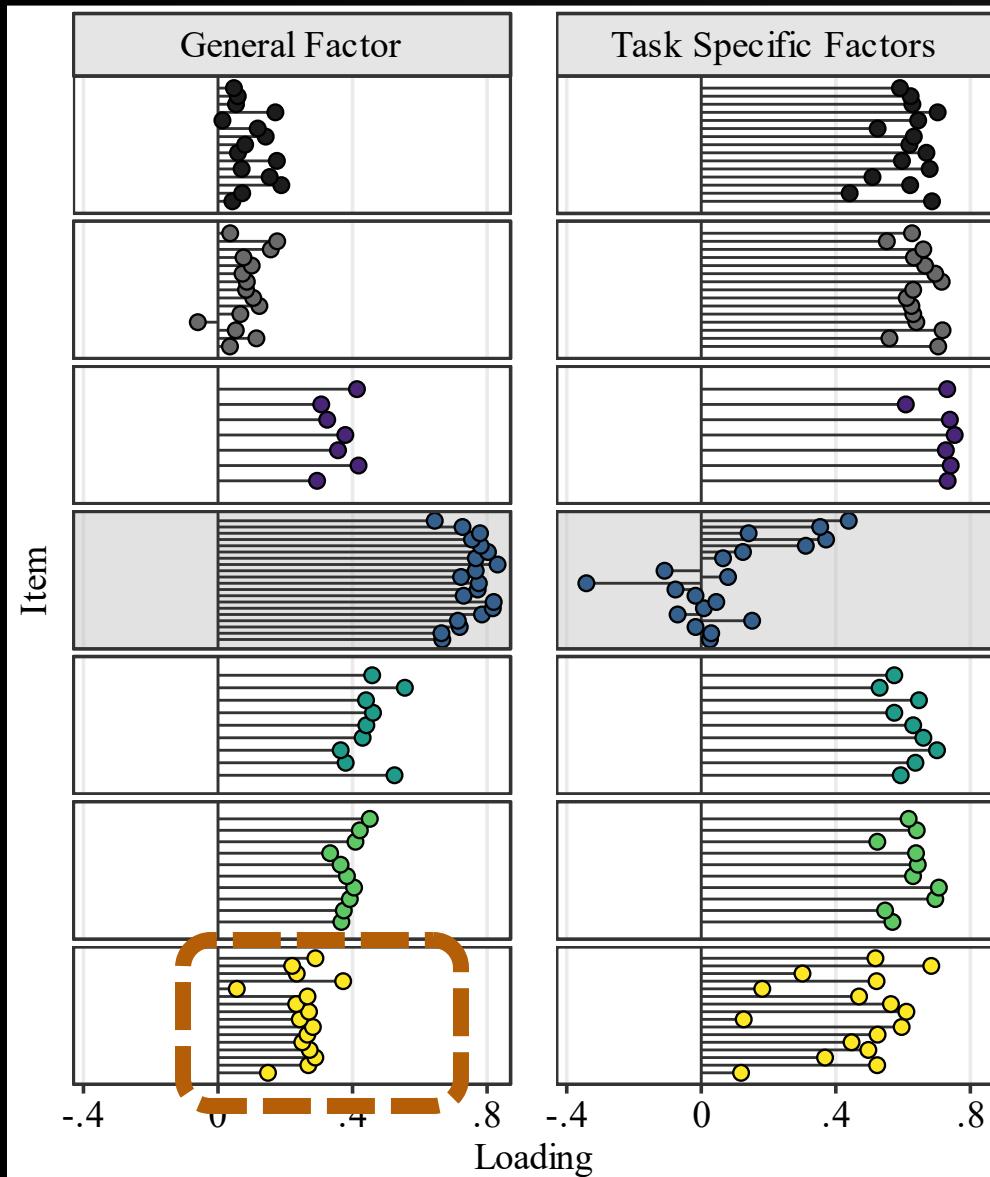
Personality:
mostly small to zero
loadings on General
Factor

LOADINGS: GENERAL VS. SPECIFIC FACTOR(S)



Estimation tasks:
mostly small to
medium loadings on
General Factor

LOADINGS: GENERAL VS. SPECIFIC FACTOR(S)



Verbal quantifiers
judgment task:
mostly small
loadings on General
Factor

KLOFT & HECK, 2024: BIFACTOR MODEL

- Across all tasks: low influence of the general factor compared to the specific factor

BIFACTOR MODEL: CORRELATIONS BETWEEN SPECIAL FACTORS

	Personality	Election	Color Dot (Reference)	Death Cause	Incapability
Election	0.11 [-0.03,0.24]				
Color Dot (Reference)	-0.16 [-0.29,-0.02]	-0.15 [-0.28,-0.02]			
Death Cause	0.07 [-0.07,0.2]	0.26 [0.13,0.38]	-0.02 [-0.16,0.11]		
Incapability	0.1 [0.06,0.32]	0.24 [0.11,0.37]	-0.03 [-0.16,0.11]	0.66 [0.58,0.73]	
Quantifier	0.28 [0.15,0.4]	0.19 [0.06,0.32]	-0.06 [-0.2,0.07]	0.3 [0.18,0.42]	0.36 [0.24,0.48]

BIFACTOR MODEL: CORRELATIONS BETWEEN SPECIAL FACTORS

	Personality	Election	Color Dot (Reference)	Death Cause	Incapability
Election	0.11 [-0.03,0.24]				
Color Dot (Reference)	-0.16 [-0.29,-0.02]	-0.15 [-0.28,-0.02]			
Death Cause	0.07 [-0.07,0.2]	0.26 [0.13,0.38]	-0.02 [-0.16,0.11]		
Incapability	0.1 [0.06,0.32]	0.24 [0.11,0.37]	-0.03 [-0.16,0.11]	0.66 [0.58,0.73]	
Quantifier	0.28 [0.15,0.4]	0.19 [0.06,0.32]	-0.06 [-0.2,0.07]	0.3 [0.18,0.42]	0.36 [0.24,0.48]

CONCLUSIONS

- Perceived variability of personality: DRS measures will probably not be valid
- Central tendency: DRS yields similar measurements to the VAS
- DRS Width: respondents seem to be sensitive to requirements of different tasks
- Similar tasks show less discriminant validity
 - Shared expertise vs. shared response bias?

FUTURE RESEARCH

- More work to disentangle expertise and uninformed confidence
- Test-retest reliability in domains other than personality
- Any ideas how to better model the data from different tasks?

THANKS TO:



Prof. Dr. Daniel W. Heck



Björn Siepe



Dr. Jean-Paul Snijder



Dr. Raphael Hartmann



Prof. Dr. Andreas Voss

Contact:

kloft@uni-marburg.de/

Slides:

<https://github.com/matthiaskloft>

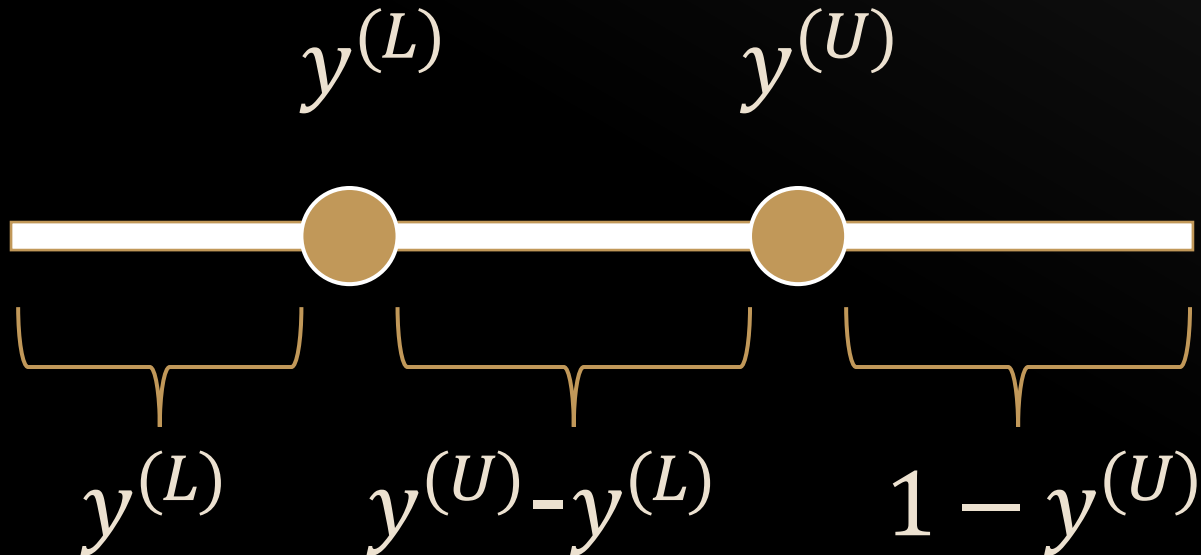
REFERENCES

- Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, 22(3), 541–562. <https://doi.org/10.1037/met0000083>
- **Kloft, M.**, Hartmann, R., Voss, A., & Heck, D. W. (2023). The Dirichlet dual response model: An item response model for continuous bounded interval responses. *Psychometrika*. <https://doi.org/10.1007/s11336-023-09924-7>
- **Kloft, M.**, Snijder, J.-P., & Heck, D.W. (2024). Measuring the variability of personality traits with interval responses: Psychometric properties of the dual-range slider response format. *Behavior Research Methods*, 56 (4), 3469–3486. <https://doi.org/10.3758/s13428-024-02394-4>
- **Kloft, M.**, & Heck, D.W. (2024). Discriminant validity of interval response formats: Investigating the dimensional structure of interval widths. *Educational and Psychological Measurement* 0 (0). <https://doi.org/10.1177/00131644241283400>
- Gersen, L. (2024). Leongersen/noUiSlider [Software]. <https://github.com/leongersen/noUiSlider>
- Smithson, M., & Broomell, S. B. (2024). Compositional data analysis tutorial: Psychological Methods. *Psychological Methods*, 29 (2), 362–378. <https://doi.org/10.1037/met0000464>

ADDITIONAL SLIDES

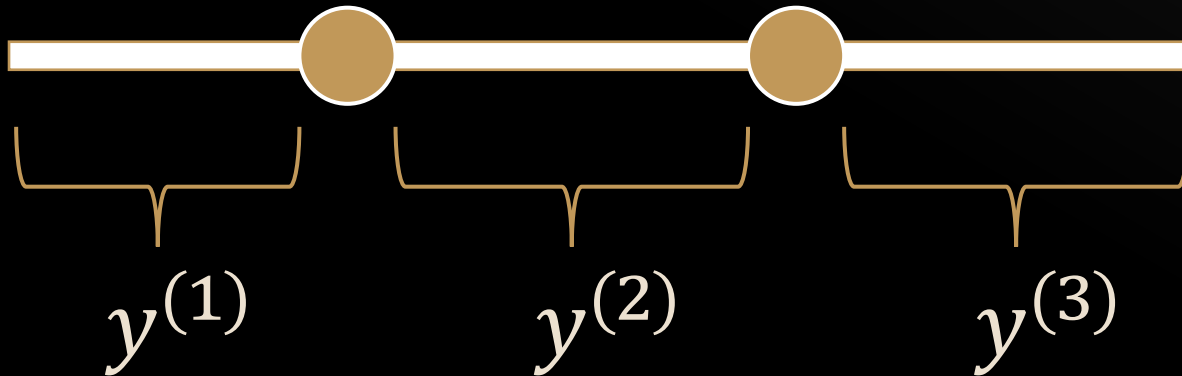
COMPOSITIONAL DATA

- Components must sum to one: simplex



COMPOSITIONAL DATA

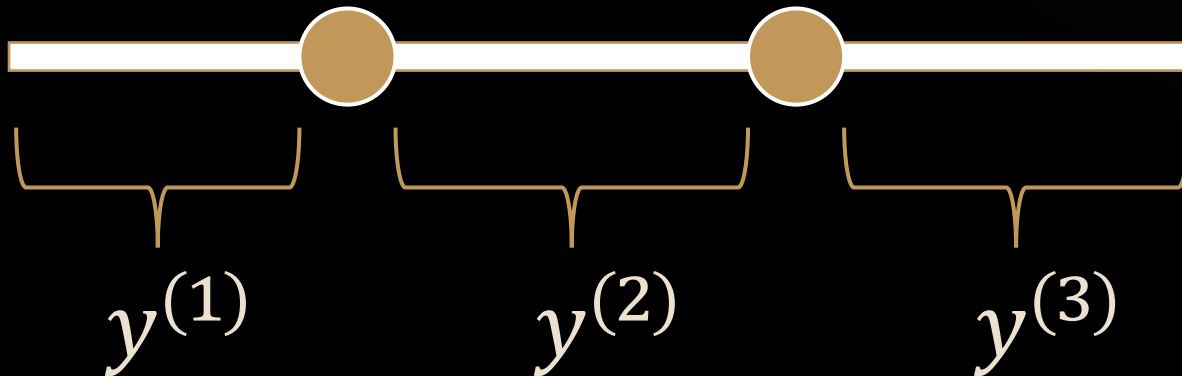
- Components must sum to one: simplex



LOG-RATIOS

Unbounded **Location**: $\log \left(\frac{y^{(1)}}{y^{(3)}} \right)$

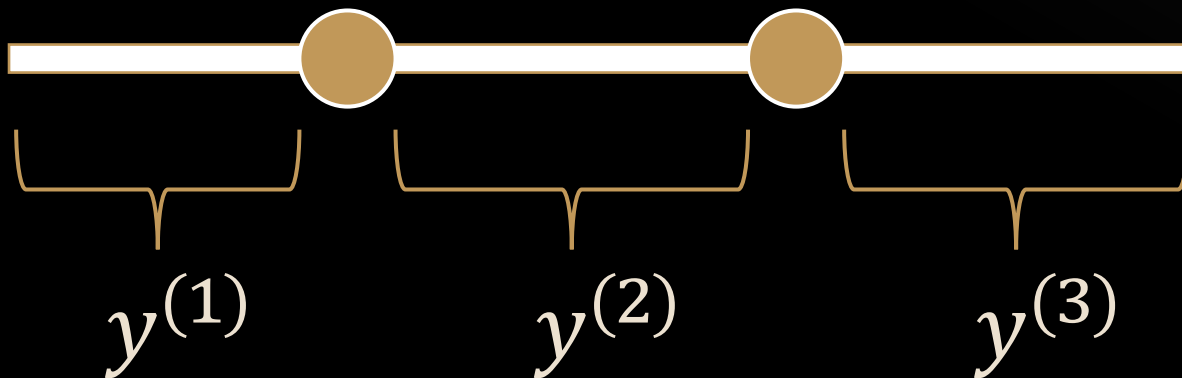
- Compares outer components



LOG-RATIOS

Unbounded **Width**: $\log \left(\frac{y^{(2)}}{\sqrt{y^{(1)} \times y^{(3)}}} \right)$

- Compares interval width to geometric mean of outer components



ISOMETRIC LOG-RATIO TRANSFORMATION

- Smithson & Broome (2024)

$$\mathbf{z} = \begin{pmatrix} z^{loc} \\ z^{wid} \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{1}{2}} \log \left(\frac{y^{(1)}}{y^{(3)}} \right) \\ \sqrt{\frac{2}{3}} \log \left(\frac{y^{(2)}}{\sqrt{y^{(1)} \times y^{(3)}}} \right) \end{pmatrix}$$

DATA EXAMPLE

- More suitable for models using a normal distribution

