The Interval Truth Model: A Cultural Consensus Model for Continuous Bounded Interval Responses

Matthias Kloft, Björn Siepe, & Daniel Heck

Workshop of the IBS-DR working groups
Non-clinical statistics and *Bayes Methods*,
Göttingen, December 5-6, 2024

Motivating Example: Prior Elicitation

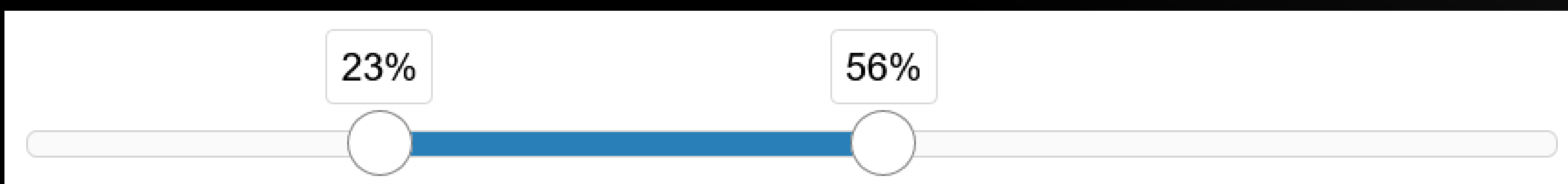
- Multiple experts
- Multiple drugs
- Expert ratings of 95% CIs for the response rate of the drugs
- Aggregation into one 95% CI per drug
  - Consensus

Interval Responses

Dual-range slider (**DRS**)

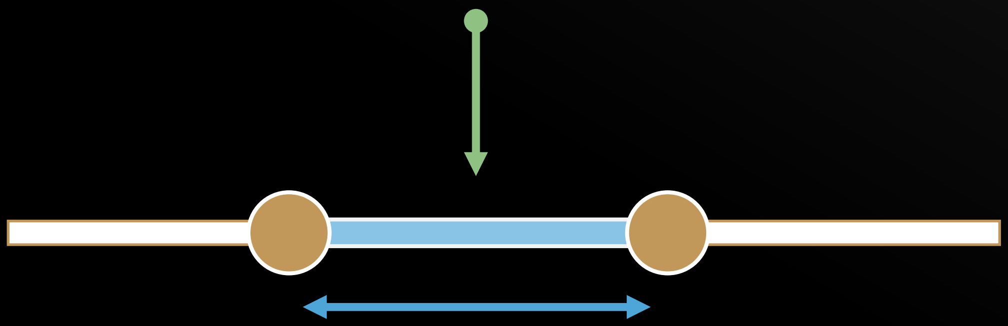
Interval Responses

Dual-range slider (**DRS**)



Interval Responses

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



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

Interval Responses: Applications

Variability of personality:

- "I am a well organized person."

Uncertainty in forecasting:

- "What will be the response rate for drug X?"

Ambiguity of verbal quantifiers:

- "What is the probability for an event that is described with the verbal quantifier ***seldom***?"

Topics of the Talk

What is an appropriate link function for interval responses?

- Smithson & Broomell (2024)

Application: consensus model

- Kloft et al. (2024, in preparation)

A Link Function for Interval Responses

Smithson & Broomell (2024)

Why Do we Need a Link Function?

Bounded Data

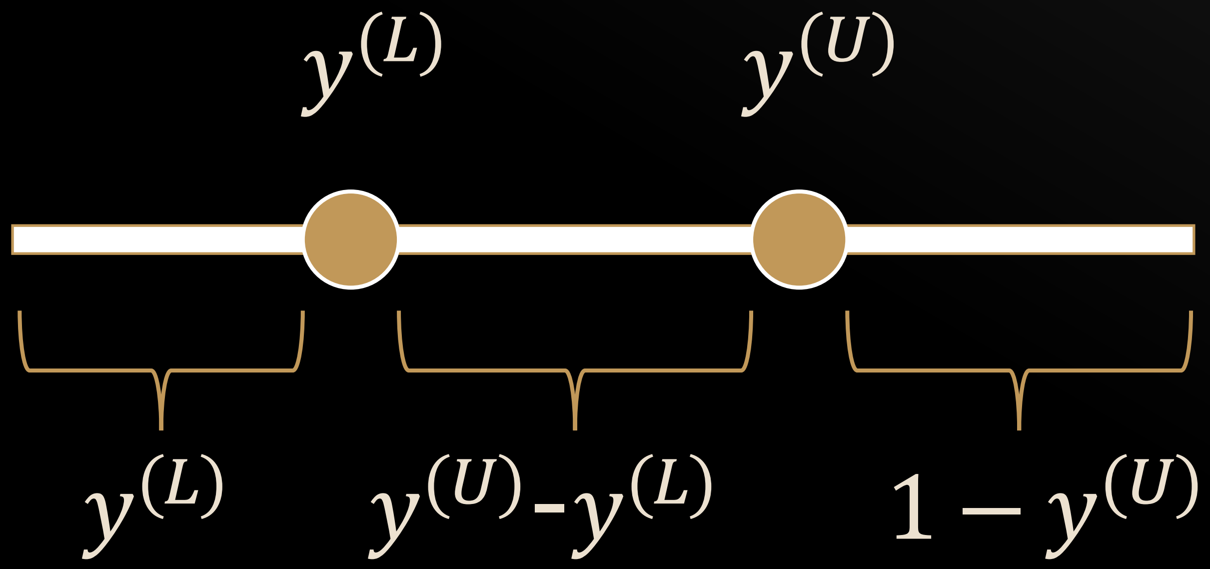
- Skew
- Dependencies between Location and Width

Aim

- Bivariate normal space for modeling
- Interpretable dimensions:
  - Location
  - Width

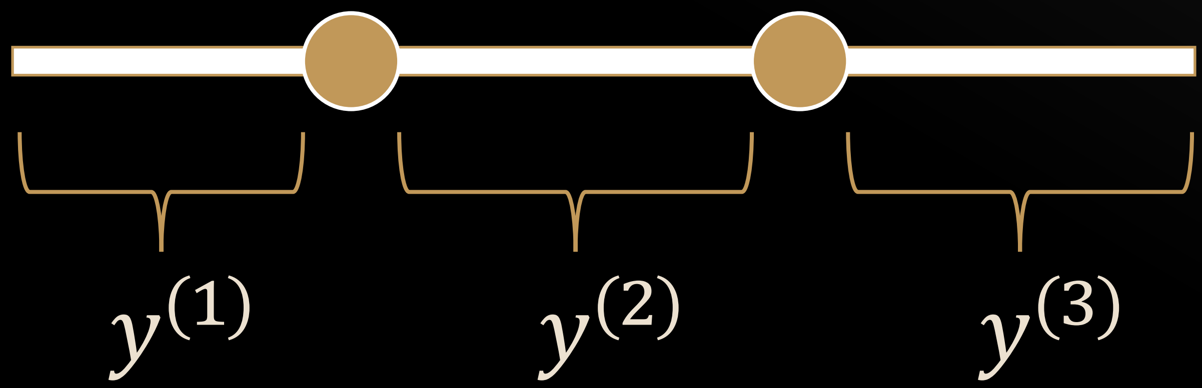
Compositional Data

- Components must sum to one: simplex



Compositional Data

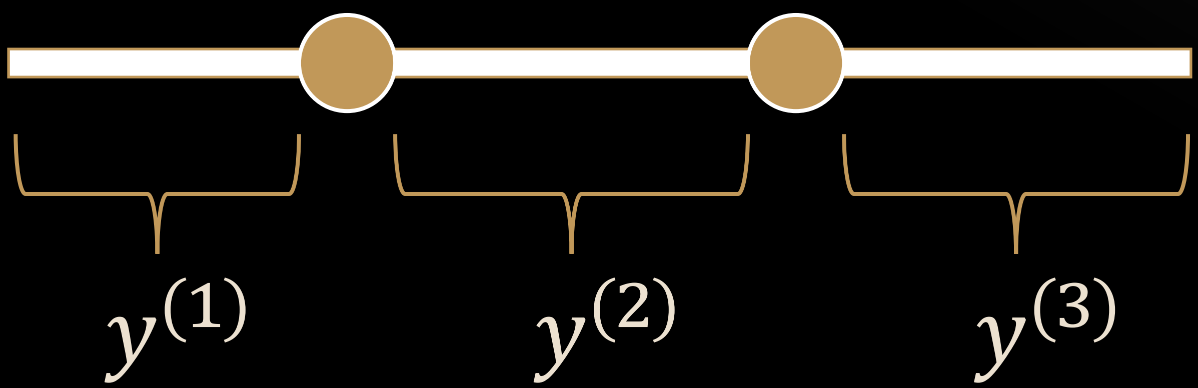
- Components must sum to one: simplex



Log-Ratios

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

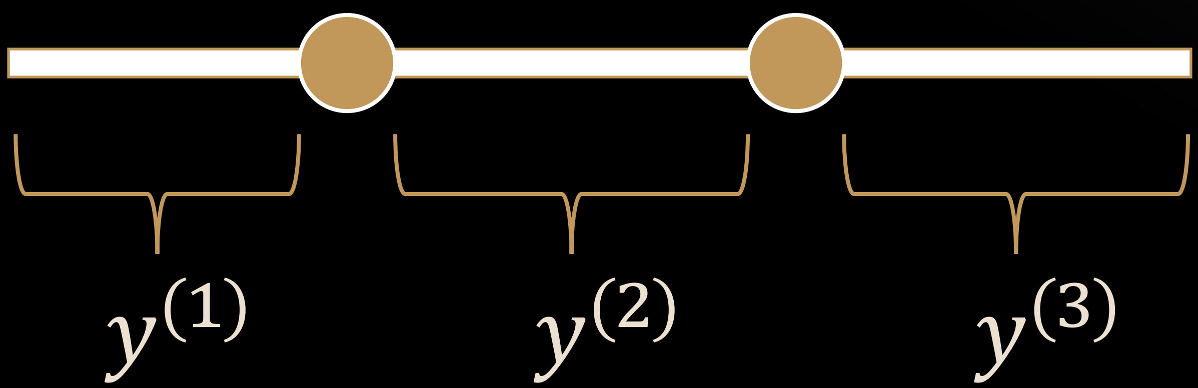
- Compares outer components



Log-Ratios

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

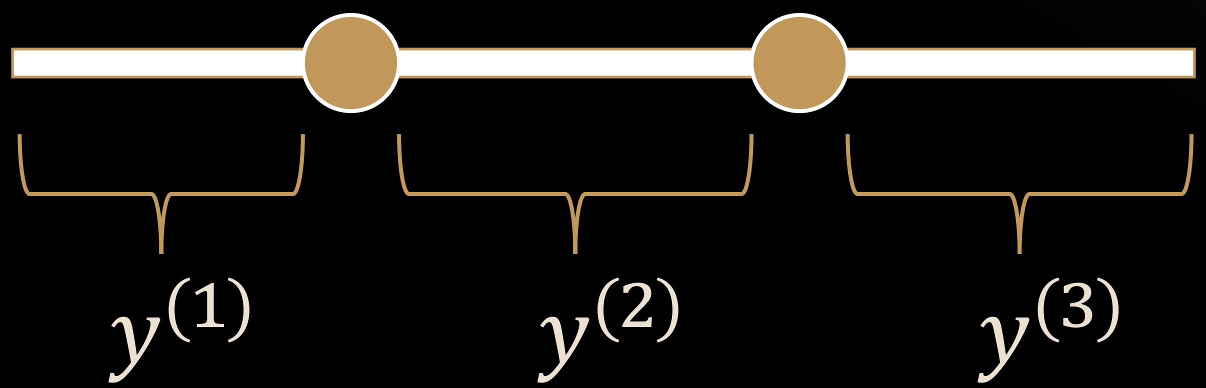
- $\boldsymbol{y} = [.33, .33, .33]$
- $\log\left(\frac{.33}{.33}\right) = \log(1) = 0$



Log-Ratios

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

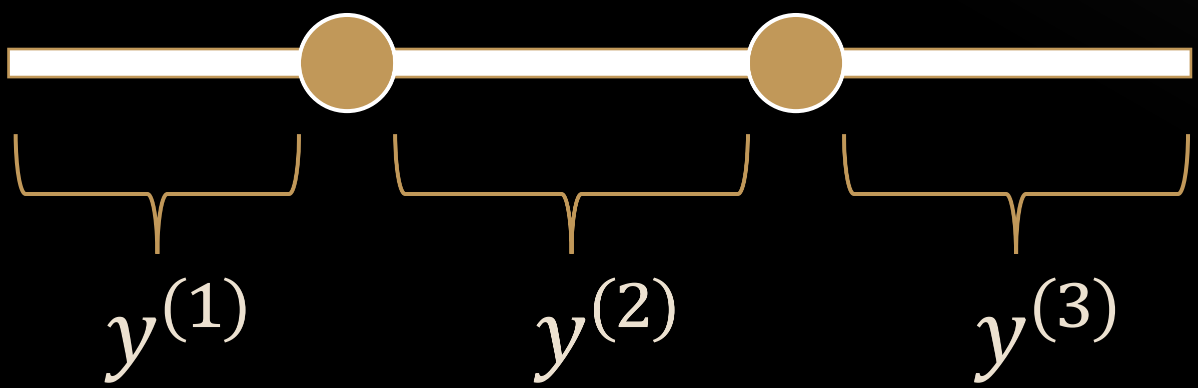
- Compares interval width to geometric mean of outer components



Log-Ratios

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

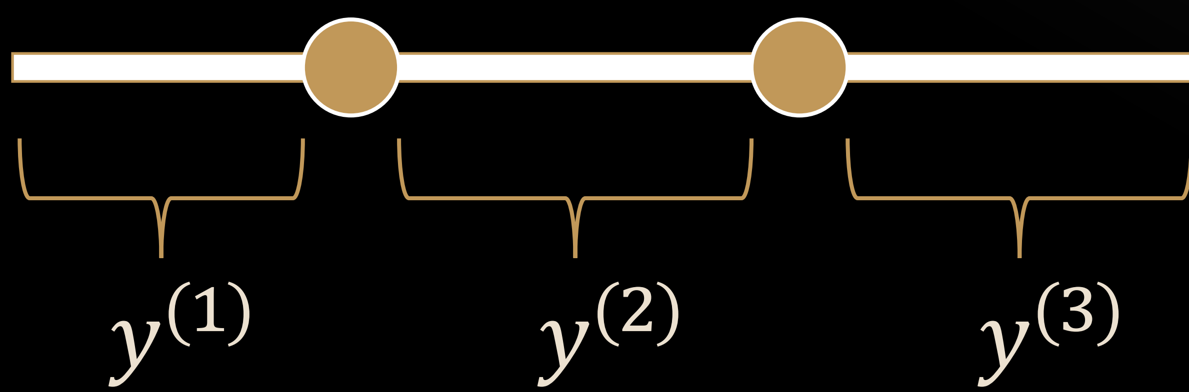
- $\boldsymbol{y} = [.33, .33, .33]$
- $\log\left(\frac{.33}{\sqrt{.33 \times .33}}\right) = \log(1) = 0$



Log-Ratios

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

- $\boldsymbol{y} = [.33, .33, .33]$
  - relates to the origin of the unbounded space



Isometric Log-Ratio Transformation

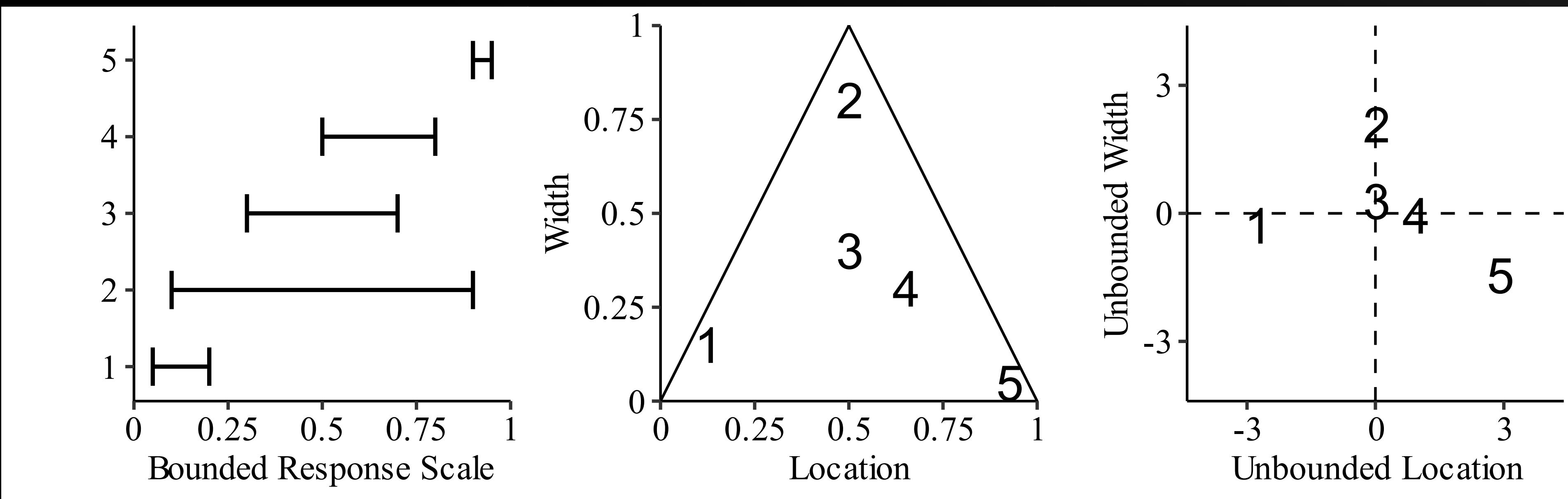
- Smithson & Broomel (2024)

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

DATA EXAMPLE

Ternary

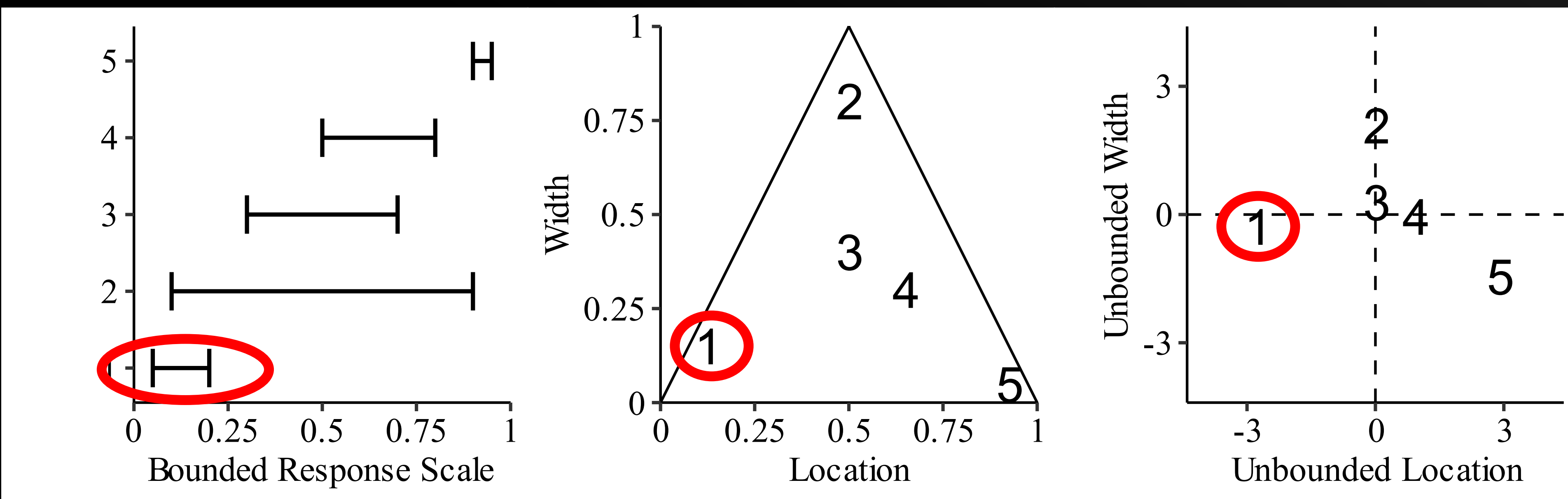
Bivariate Normal



Data Example

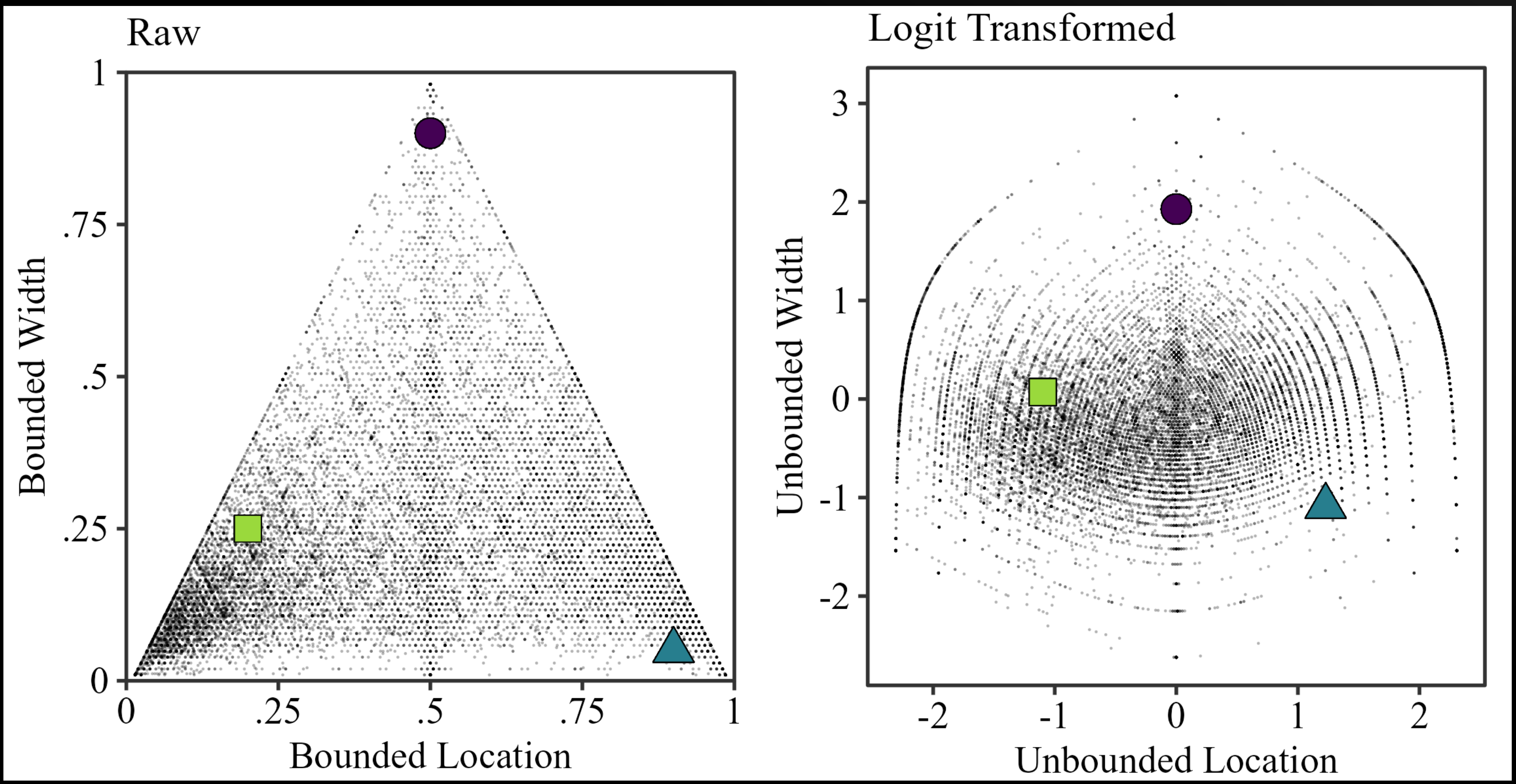
Ternary

Bivariate Normal



Data Example

- More suitable for models using a normal distribution



Application: Consensus Model

Kloft et al. (2024, in preparation)

Consensus Models

Setup:

- Multiple items,
  - e.g., different drugs for which we need ratings of 95% prior CIs
- Multiple experts / raters

Aim:

- Estimate averaged intervals
- A rater‵s influence is weighted by their expertise, i.e., consistency

Interval Truth Model: Overview

- Extension of a univarite logit-normal model (Anders et al., 2014)

➢ Bivariate logit-normal model
  - Link function: isometric log-ratio (ILR)

$$\mathrm{ILR}(\boldsymbol{z}_{ij}) \sim BVN\left(\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}\right)$$

  - Implementation in Stan

Assumptions & Prerequisites

- **Multiple raters**
- **Multiple** interval **ratings** per rater
- **True latent consensus interval** per item for the group of raters
- Raters make a **latent appraisal** of the **true consensus**

Assumptions & Prerequisites

- **Variance** in the level of **expertise** of raters
- Expertise can be estimated based on **consistency** of ratings in relation to the latent consensus intervals
- **Expertise** can be used to **weight** ratings in the aggregation of responses
  - Joint estimation of expertise and latent consensus

Assumptions & Prerequisites

- **Rater Biases** further distort the latent appraisal before we arrive at the observed ratings

Model Mechanics: Latent Appraisal

- **Latent appraisal**: latent consensus plus some error
- Error / Precision based on:
  - **Proficiency** of the rater
  - **Discernibility** of the item
- Bivariate normal distribution for appraisals
  - 2D: location and width
  - Expected value: latent consensus

Latent Appraisal

Rater *i* on item *j*

$$\begin{pmatrix} \mathrm{A}_{\mathrm{ij}}^{\mathrm{loc}} \\ \mathrm{A}_{\mathrm{ij}}^{\mathrm{wid}} \end{pmatrix} = \begin{pmatrix} \mathrm{T}_{\mathrm{j}}^{\mathrm{loc}} \\ \mathrm{T}_{\mathrm{j}}^{\mathrm{wid}} \end{pmatrix} + \begin{pmatrix} \epsilon_{ij}^{loc} \\ \epsilon_{\mathrm{ij}}^{\mathrm{wid}} \end{pmatrix}$$

$$\begin{pmatrix} \epsilon_{ij}^{loc} \\ \epsilon_{\mathrm{ij}}^{\mathrm{wid}} \end{pmatrix} \sim BVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}_{ij} \right)$$

Latent Appraisal

Latent Appraisal

$$\begin{pmatrix} \mathrm{A}_{\mathrm{ij}}^{\mathrm{loc}} \\ \mathrm{A}_{\mathrm{ij}}^{\mathrm{wid}} \end{pmatrix} = \begin{pmatrix} \mathrm{T}_{\mathrm{j}}^{\mathrm{loc}} \\ \mathrm{T}_{\mathrm{j}}^{\mathrm{wid}} \end{pmatrix} + \begin{pmatrix} \epsilon_{ij}^{loc} \\ \epsilon_{\mathrm{ij}}^{\mathrm{wid}} \end{pmatrix}$$

$$\begin{pmatrix} \epsilon_{ij}^{loc} \\ \epsilon_{\mathrm{ij}}^{\mathrm{wid}} \end{pmatrix} \sim BVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}_{ij} \right)$$

Latent Appraisal

Latent True Consensus

$$\begin{pmatrix} A_{ij}^{loc} \\ A_{ij}^{wid} \end{pmatrix} = \begin{pmatrix} T_{j}^{loc} \\ T_{j}^{wid} \end{pmatrix} + \begin{pmatrix} \epsilon_{ij}^{loc} \\ \epsilon_{ij}^{wid} \end{pmatrix}$$

$$\begin{pmatrix} \epsilon_{ij}^{loc} \\ \epsilon_{ij}^{wid} \end{pmatrix} \sim BVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}_{ij} \right)$$

Latent Appraisal

Error / Precision of appraisal

$$\begin{pmatrix} \mathrm{A}_{\mathrm{ij}}^{\mathrm{loc}} \\ \mathrm{A}_{\mathrm{ij}}^{\mathrm{wid}} \end{pmatrix} = \begin{pmatrix} \mathrm{T}_{\mathrm{j}}^{\mathrm{loc}} \\ \mathrm{T}_{\mathrm{j}}^{\mathrm{wid}} \end{pmatrix} + \begin{pmatrix} \epsilon_{ij}^{loc} \\ \epsilon_{\mathrm{ij}}^{\mathrm{wid}} \end{pmatrix}$$

$$\begin{pmatrix} \epsilon_{ij}^{loc} \\ \epsilon_{\mathrm{ij}}^{\mathrm{wid}} \end{pmatrix} \sim BVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{\Sigma}_{ij} \right)$$

Latent Appraisal

Error / Precision of appraisal

$$\boldsymbol{\Sigma}_{ij} = \begin{pmatrix} \left(\frac{1}{E_i^{loc}}\frac{1}{\lambda_j^{loc}}\right)^2 & \rho_{ij} \\ \rho_{ij} & \left(\frac{1}{E_i^{wid}}\frac{1}{\lambda_j^{wid}}\right)^2 \end{pmatrix}$$

$$\rho_{ij} = \sqrt{\omega_j \left(\frac{1}{E_i^{loc}}\frac{1}{\lambda_j^{loc}}\right)^2 \left(\frac{1}{E_i^{wid}}\frac{1}{\lambda_j^{wid}}\right)^2}$$

LATENT APPRAISAL

Rater proficiency to detect consensus

$$\boldsymbol{\Sigma}_{ij} = \begin{pmatrix} \left(\frac{1}{E_i^{loc}}\frac{1}{\lambda_j^{loc}}\right)^2 & \rho_{ij} \\ \rho_{ij} & \left(\frac{1}{E_i^{wid}}\frac{1}{\lambda_j^{wid}}\right)^2 \end{pmatrix}$$

$$\rho_{ij} = \sqrt{\omega_j \left(\frac{1}{E_i^{loc}}\frac{1}{\lambda_j^{loc}}\right)^2 \left(\frac{1}{E_i^{wid}}\frac{1}{\lambda_j^{wid}}\right)^2}$$

Latent Appraisal

Item discernibility / difficulty

$$\boldsymbol{\Sigma}_{ij} = \begin{pmatrix} \left(\frac{1}{E_i^{loc}} \frac{1}{\lambda_j^{loc}}\right)^2 & \rho_{ij} \\ \rho_{ij} & \left(\frac{1}{E_i^{wid}} \frac{1}{\lambda_j^{wid}}\right)^2 \end{pmatrix}$$

$$\rho_{ij} = \sqrt{\omega_j \left(\frac{1}{E_i^{loc}} \frac{1}{\lambda_j^{loc}}\right)^2 \left(\frac{1}{E_i^{wid}} \frac{1}{\lambda_j^{wid}}\right)^2}$$

LATENT APPRAISAL

Residual covariance between location and width

$$\boldsymbol{\Sigma}_{ij} = \begin{pmatrix} \left(\frac{1}{E_i^{loc}} \frac{1}{\lambda_j^{loc}}\right)^2 & \rho_{ij} \\ \rho_{ij} & \left(\frac{1}{E_i^{wid}} \frac{1}{\lambda_j^{wid}}\right)^2 \end{pmatrix}$$

$$\rho_{ij} = \sqrt{\omega_j \left(\frac{1}{E_i^{loc}} \frac{1}{\lambda_j^{loc}}\right)^2 \left(\frac{1}{E_i^{wid}} \frac{1}{\lambda_j^{wid}}\right)^2}$$

Latent Appraisal

Residual correlation between location and width

$$\boldsymbol{\Sigma}_{ij} = \begin{pmatrix} \left(\frac{1}{E_i^{loc}}\frac{1}{\lambda_j^{loc}}\right)^2 & \rho_{ij} \\ \rho_{ij} & \left(\frac{1}{E_i^{wid}}\frac{1}{\lambda_j^{wid}}\right)^2 \end{pmatrix}$$

$$\rho_{ij} = \sqrt{\omega_j \left(\frac{1}{E_i^{loc}}\frac{1}{\lambda_j^{loc}}\right)^2 \left(\frac{1}{E_i^{wid}}\frac{1}{\lambda_j^{wid}}\right)^2}$$

Model Mechanics: Biases

- Latent appraisal is scaled and shifted by the rater‘s biases
  - These affect all responses of a particular rater

- Shifting Biases
  - Interval **locations** shift to the **left / right** on the response scale
  - Interval **widths** get **wider / narrower**

- Scaling Bias (extremity bias):
  - Pushes / pulls the **location outwards / inwards** with respect to the response scale‘s center

Biases

A rater`s biases shift and scale the latent appraisal

$$\mathrm{ILR}(\mathbf{y}_{\mathrm{ij}}) = \left(\mathrm{A}_{\mathrm{ij}}^{\mathrm{loc}} a_i^{loc} + b_i^{loc},\ \mathrm{A}_{\mathrm{ij}}^{\mathrm{wid}} + b_i^{wid}\right)^{\top}$$

Biases

Shifting biases (like a random intercept)

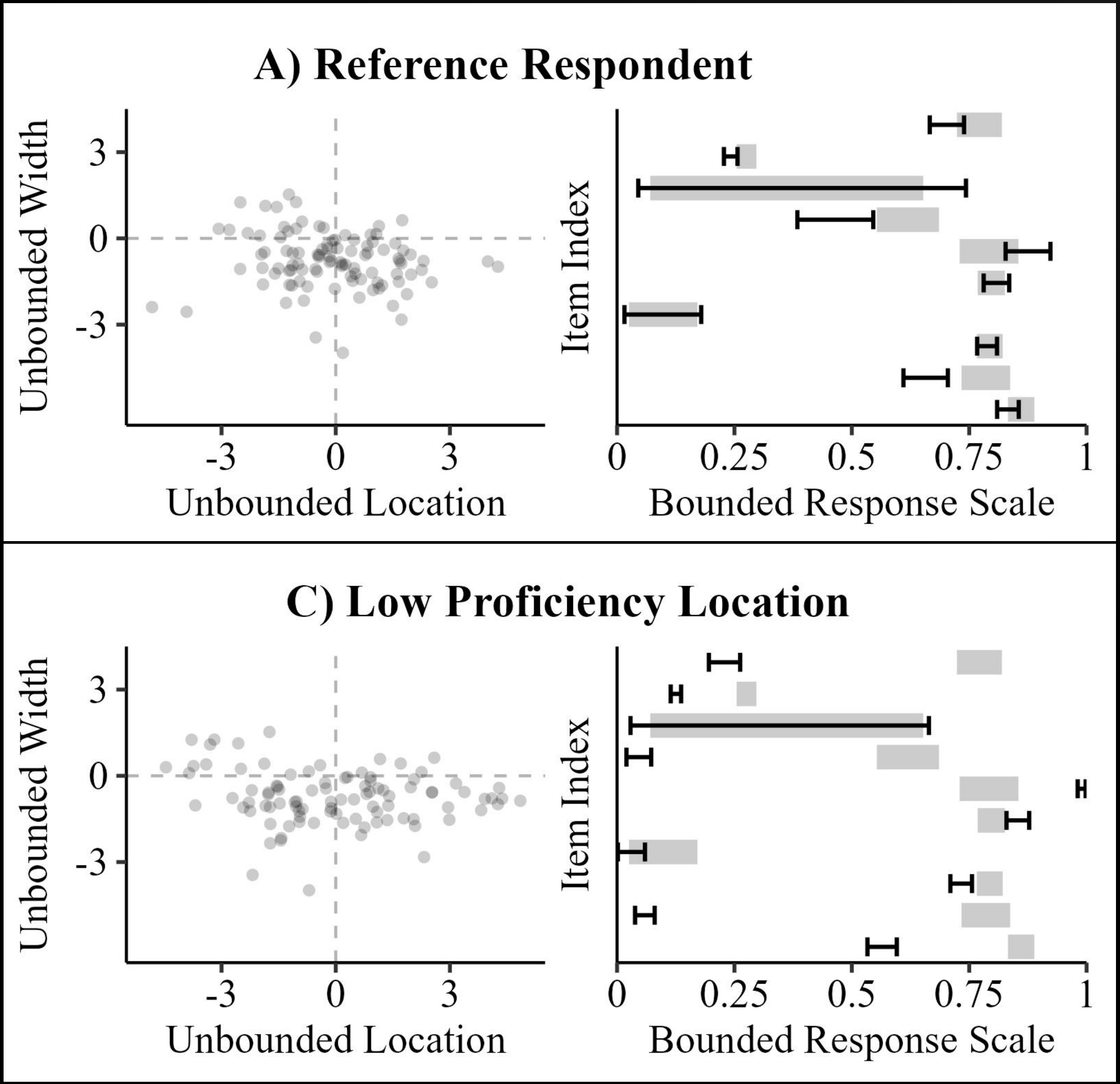
$$\mathrm{ILR}(\mathbf{y}_{\mathrm{ij}}) = \left(\mathrm{A}_{\mathrm{ij}}^{\mathrm{loc}} a_i^{loc} + b_i^{loc},\ \mathrm{A}_{\mathrm{ij}}^{\mathrm{wid}} + b_i^{wid}\right)^{\top}$$

BIASES

Scaling bias (extremety bias)

$$\mathrm{ILR}(\mathbf{y}_{\mathrm{ij}}) = \left(\mathrm{A}_{\mathrm{ij}}^{\mathrm{loc}} a_i^{loc} + b_i^{loc},\ \mathrm{A}_{\mathrm{ij}}^{\mathrm{wid}} + b_i^{wid}\right)^{\top}$$

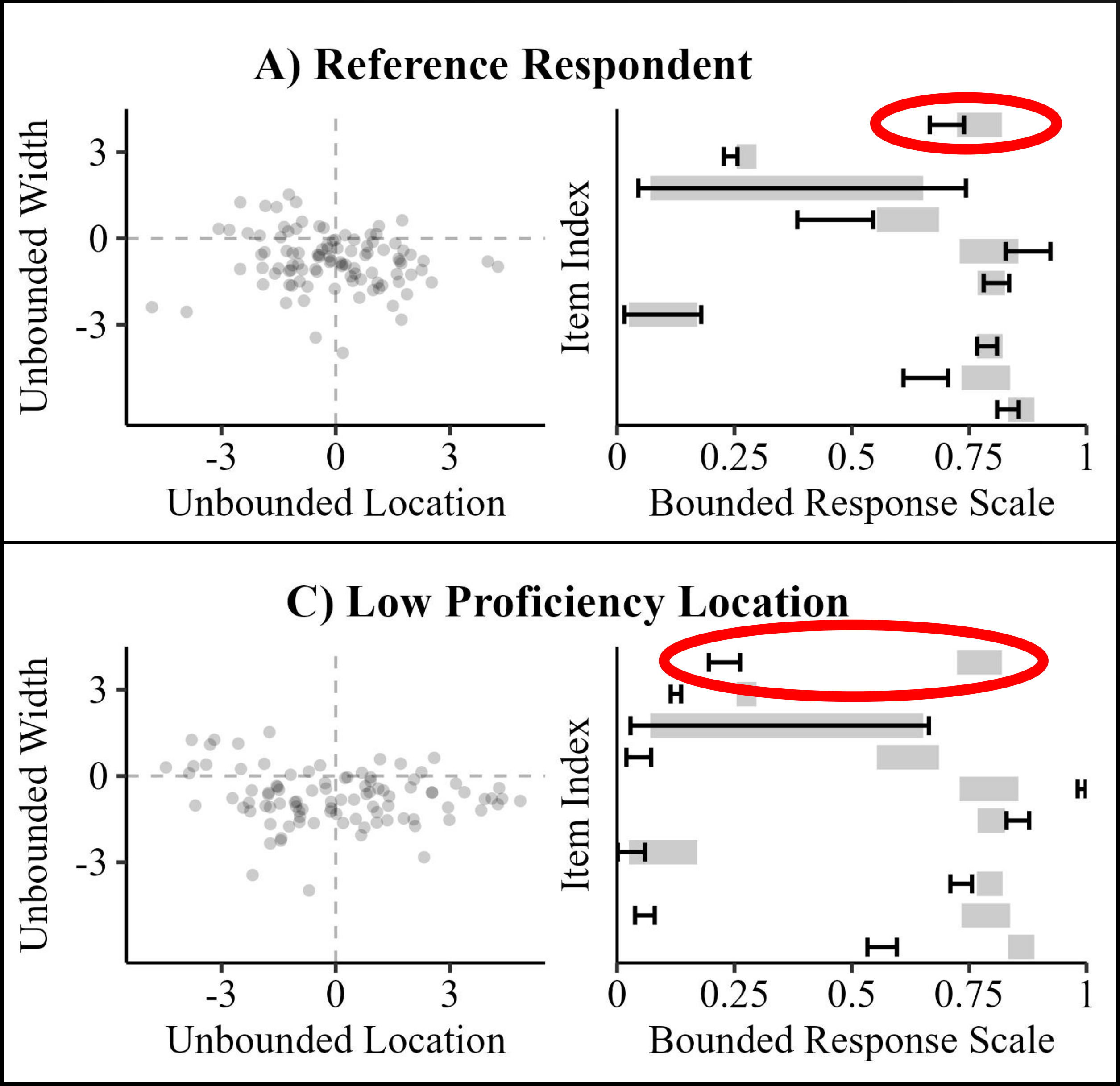
Latent Appraisal (No Biases)



High proficiency, **location**

Low proficiency, **location**

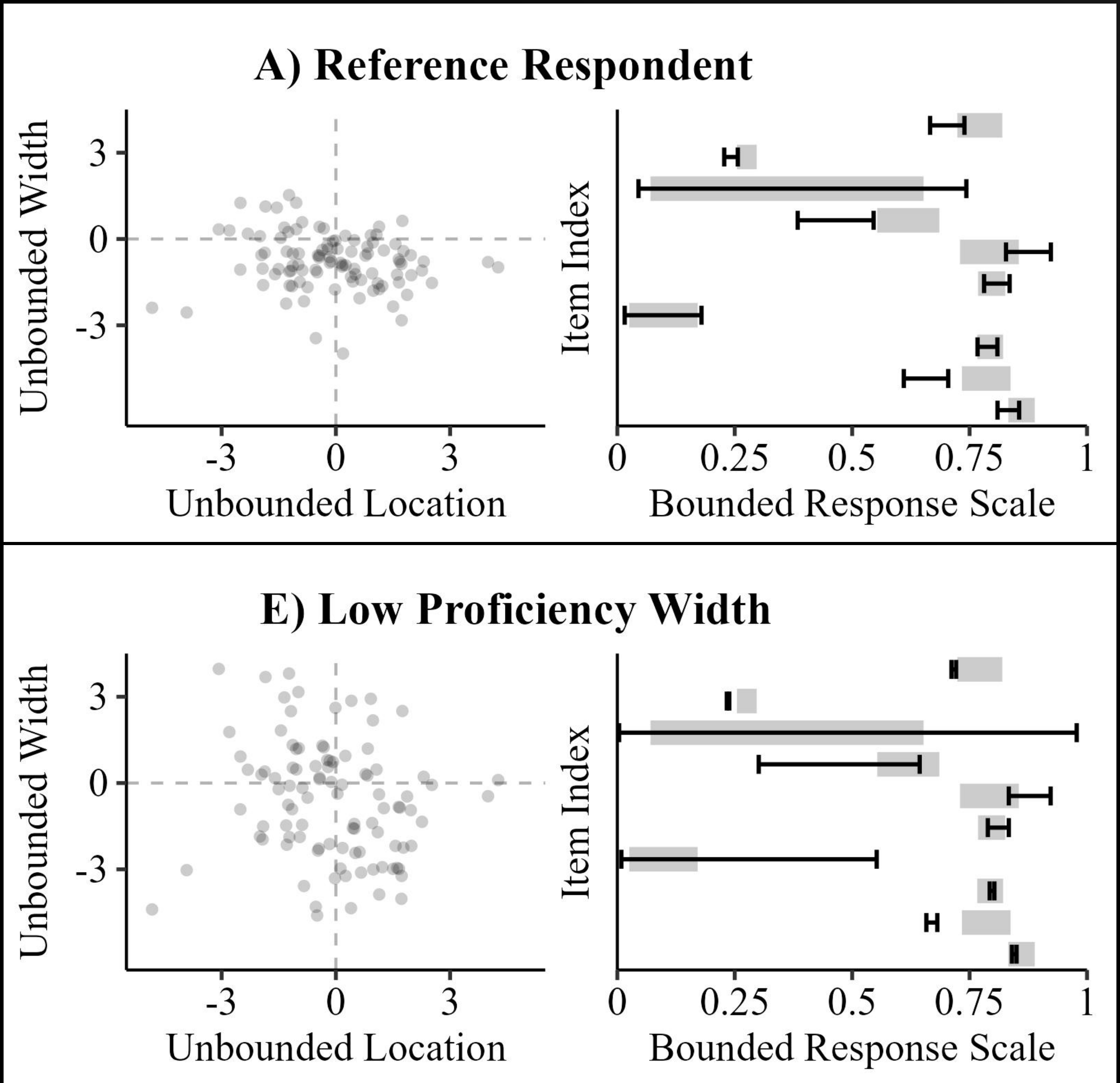
LATENT APPRAISAL (NO BIASES)



High proficiency, **location**

Low proficiency, **location**

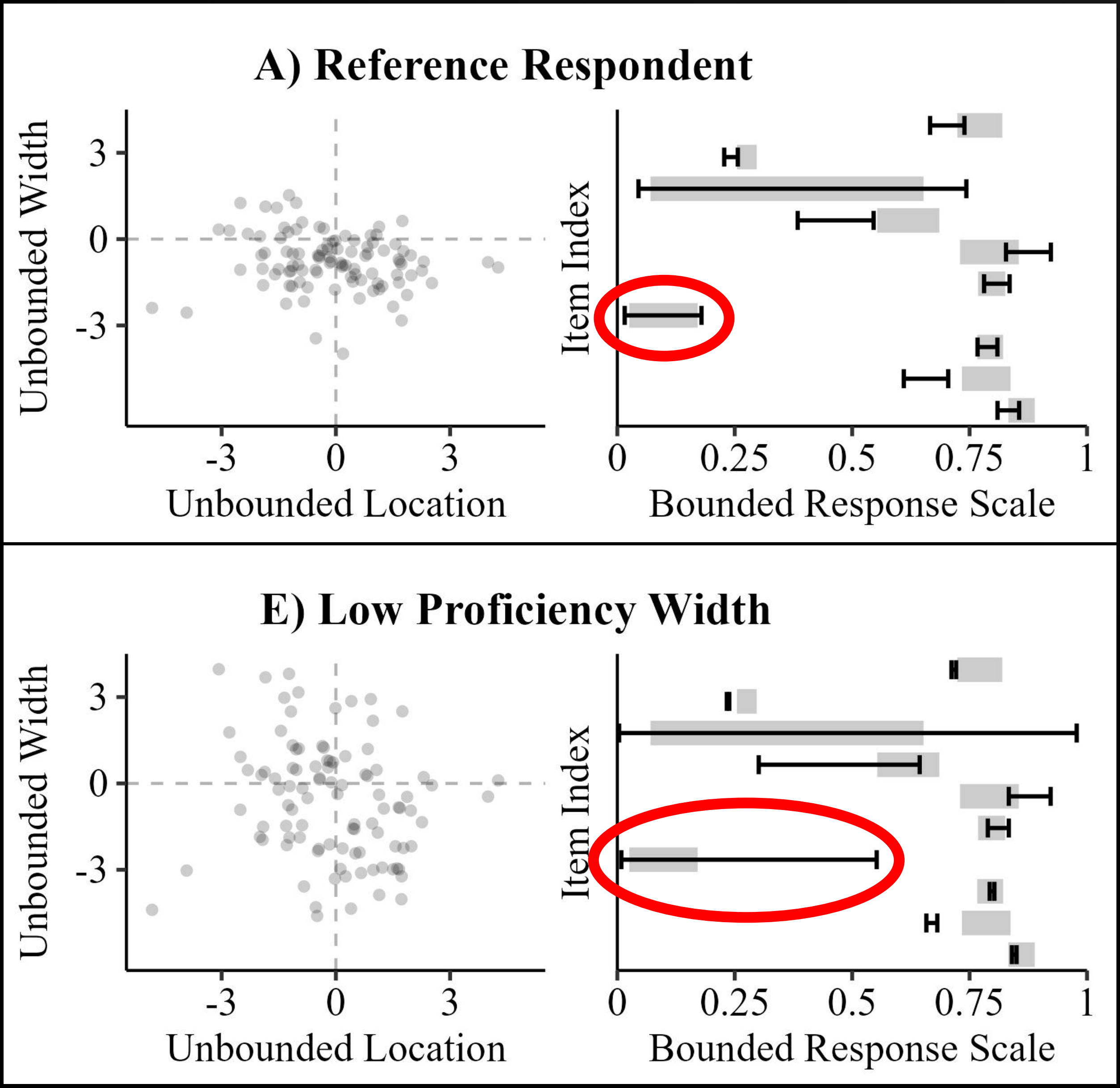
Latent Appraisal (No Biases)



High proficiency, **width**

Low proficiency, **width**

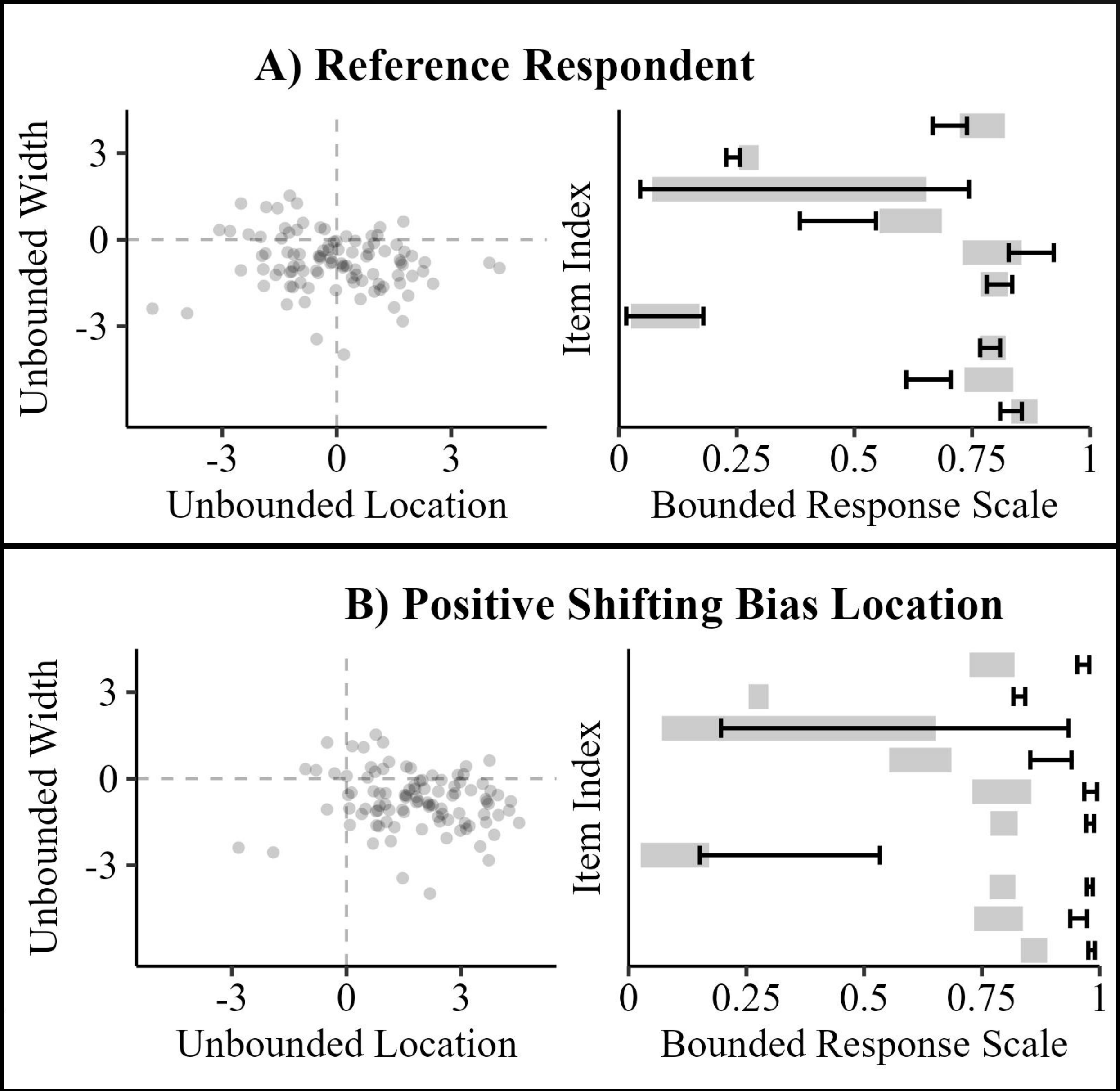
Latent Appraisal (No Biases)



High proficiency, **width**

Low proficiency, **width**

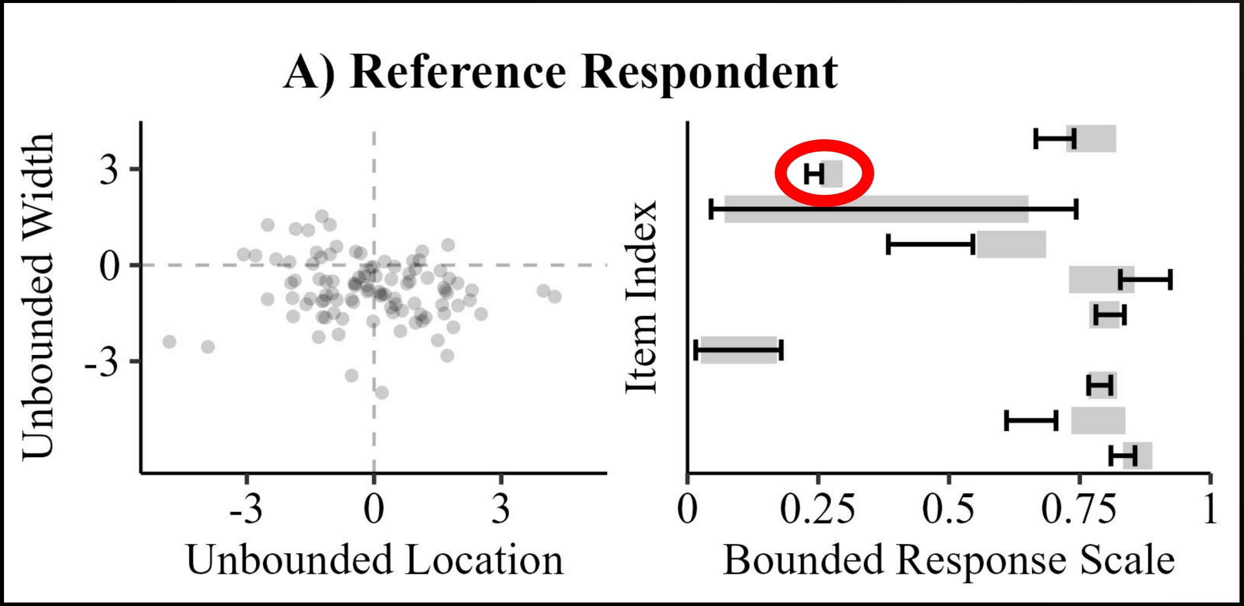
BIASES



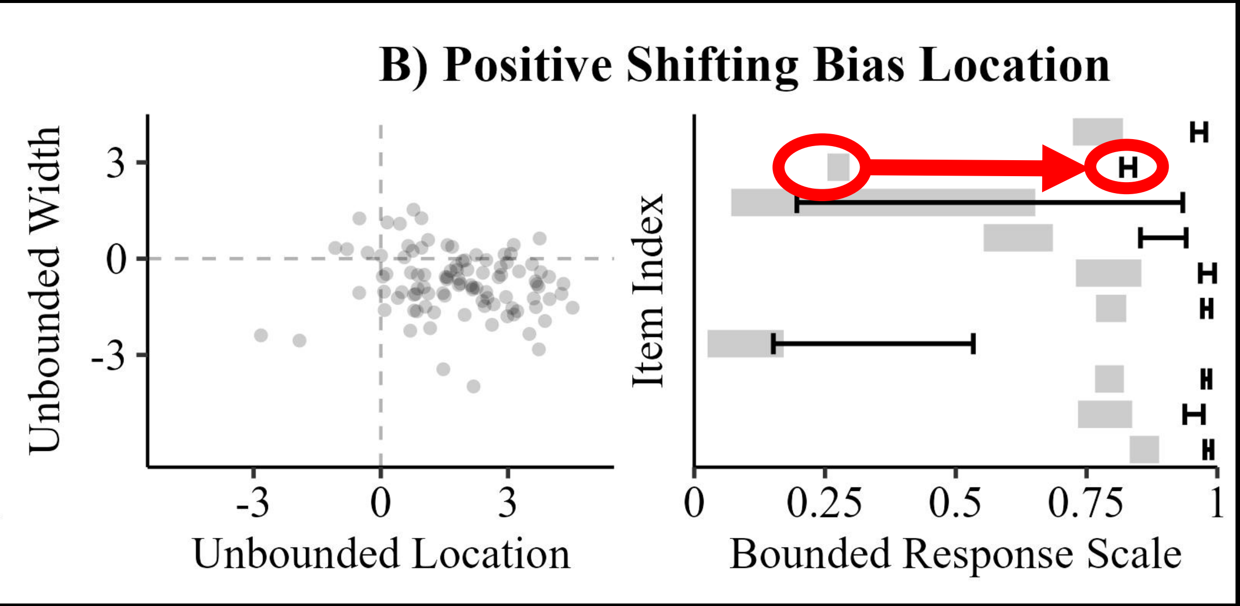
No biases,
High proficiencies

Positive **location** shifting bias

Biases

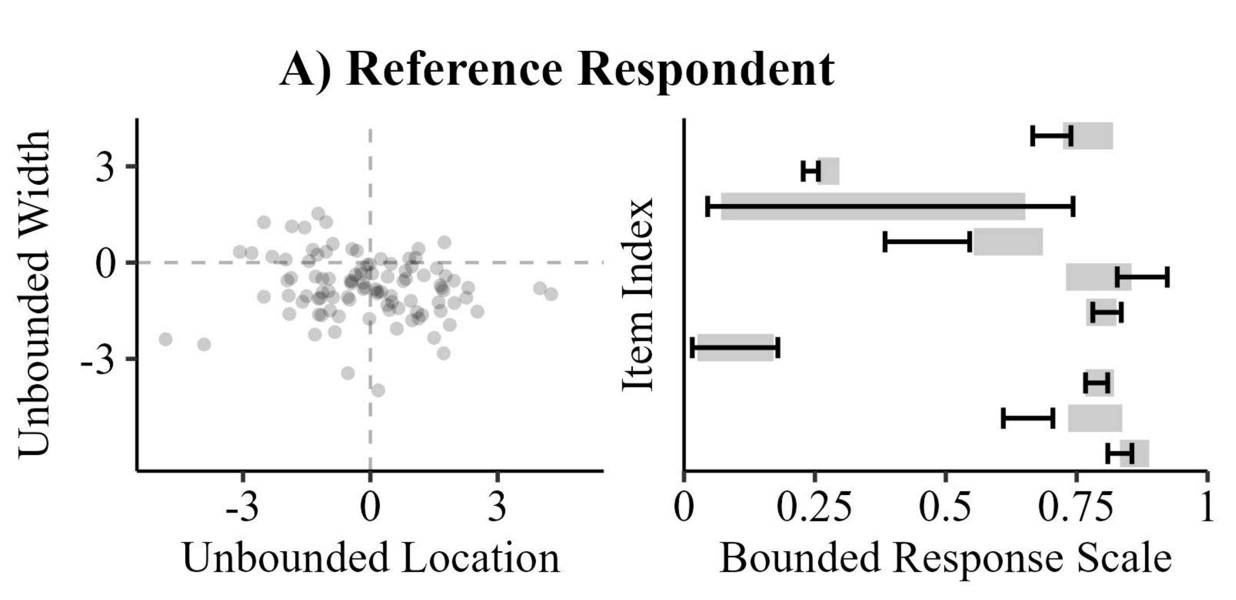


No biases,
High proficiencies



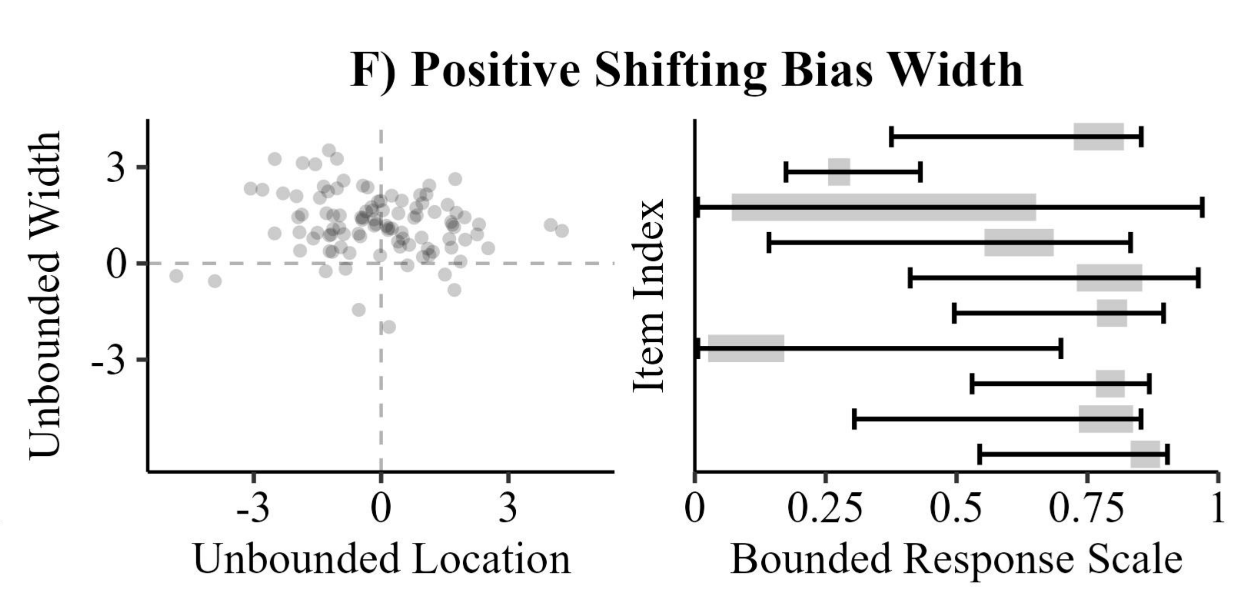
Positive **location** shifting bias

Model Mechanics: Biases



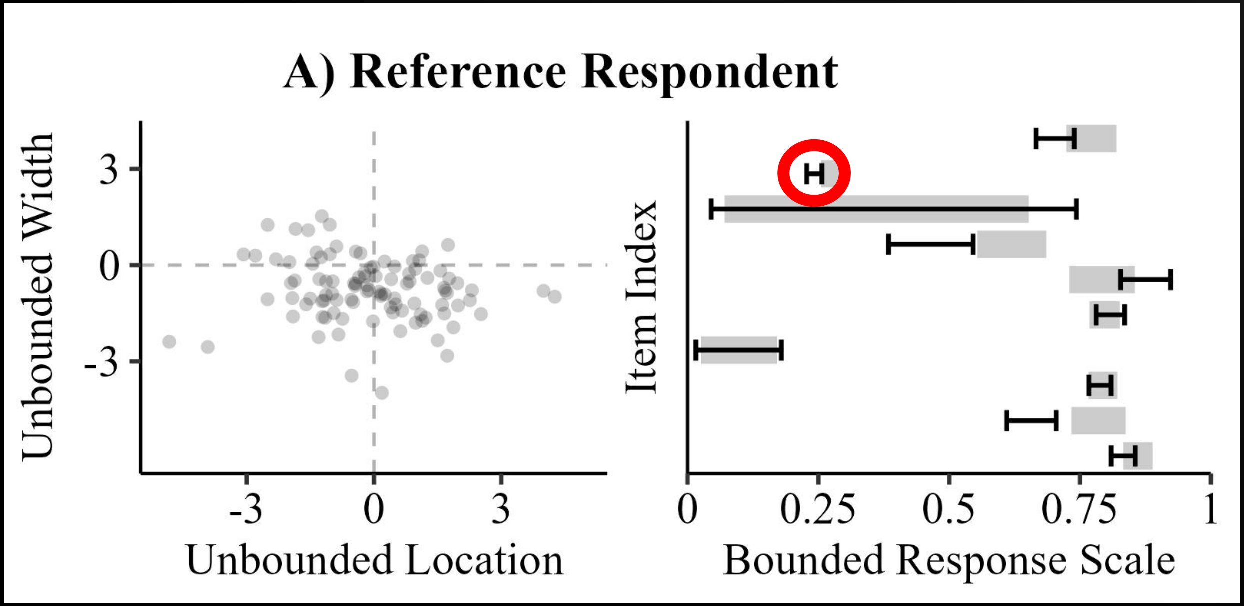
No biases,

High proficiencies

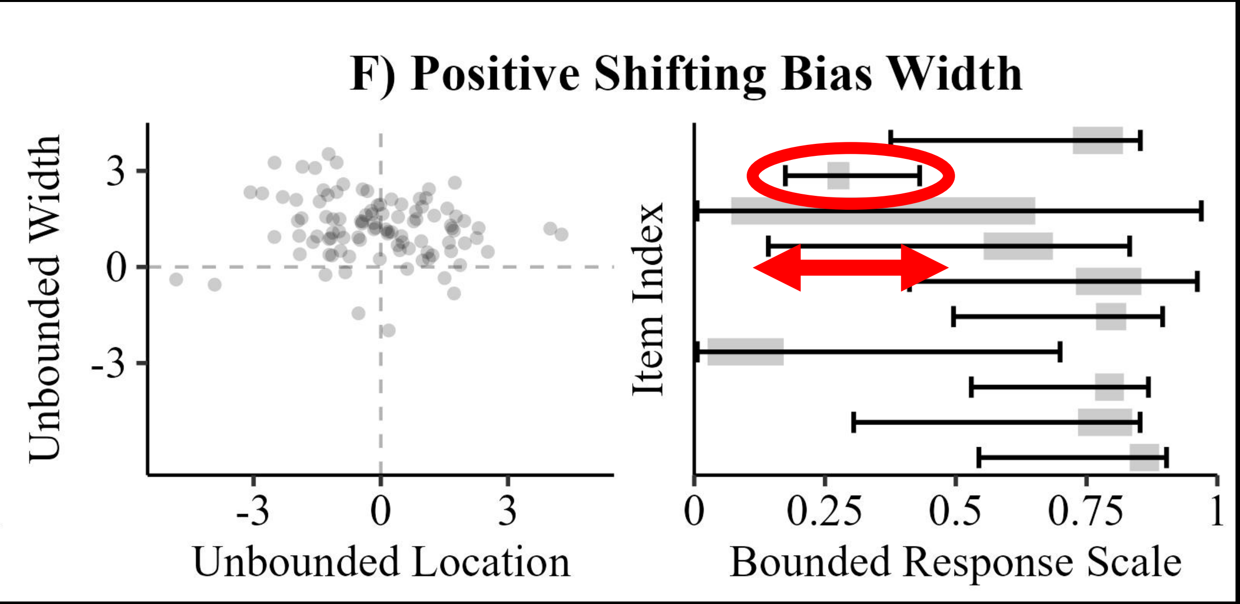


Positive **width** shifting bias

BIASES

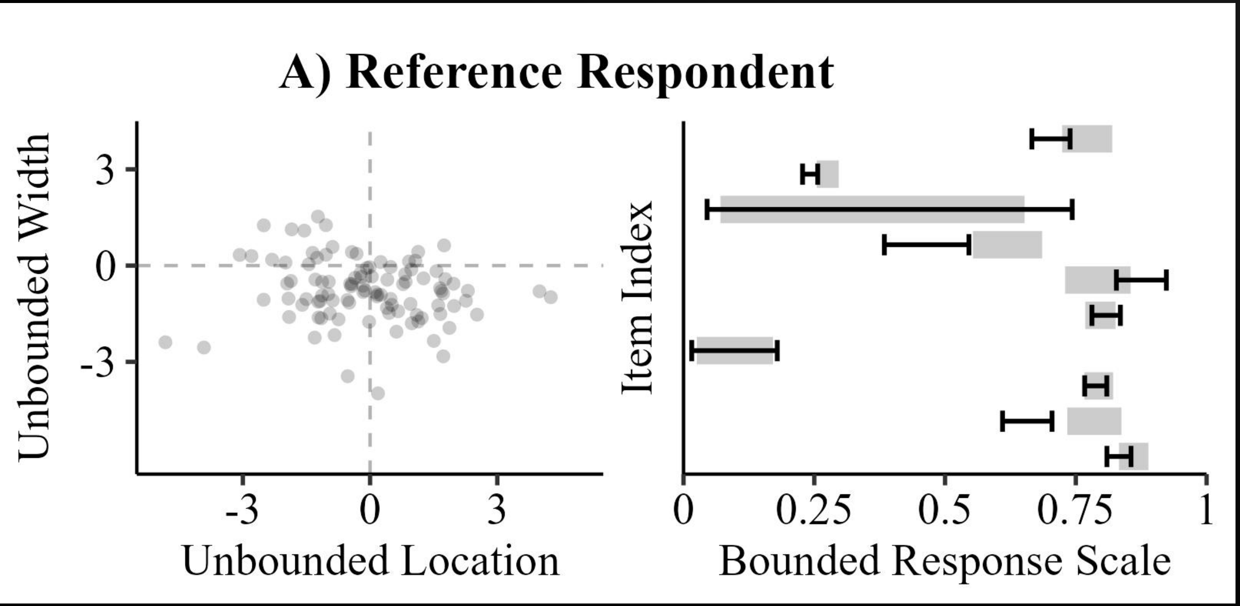


No biases,
High proficiencies



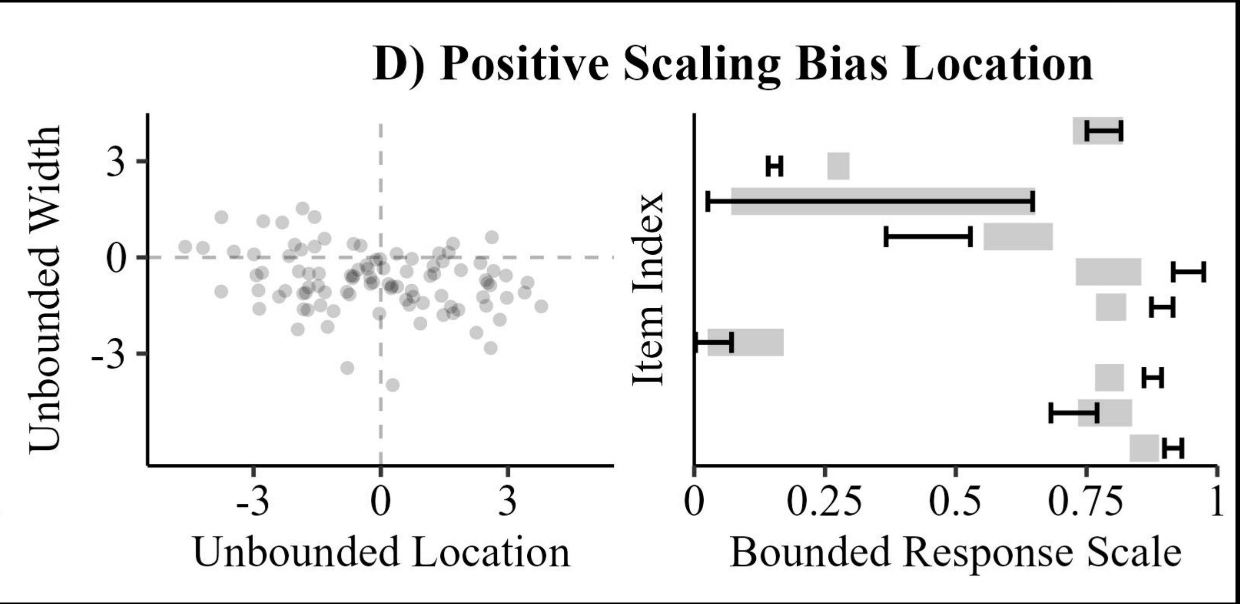
Positive **width** shifting bias

BIASES



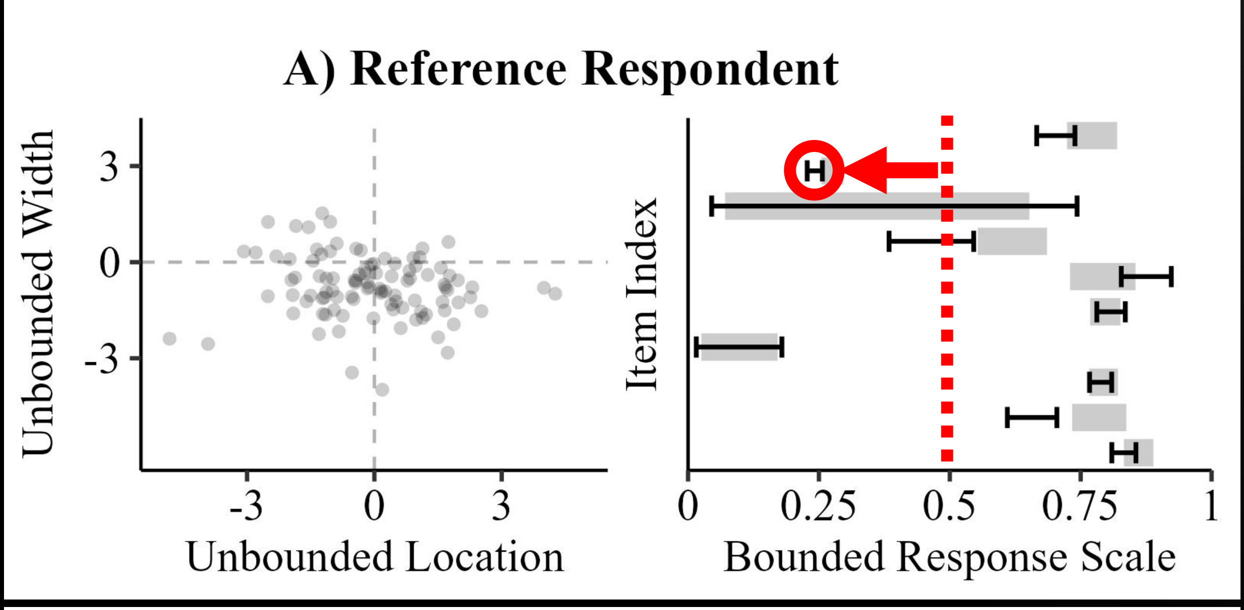
No biases,

High proficiencies

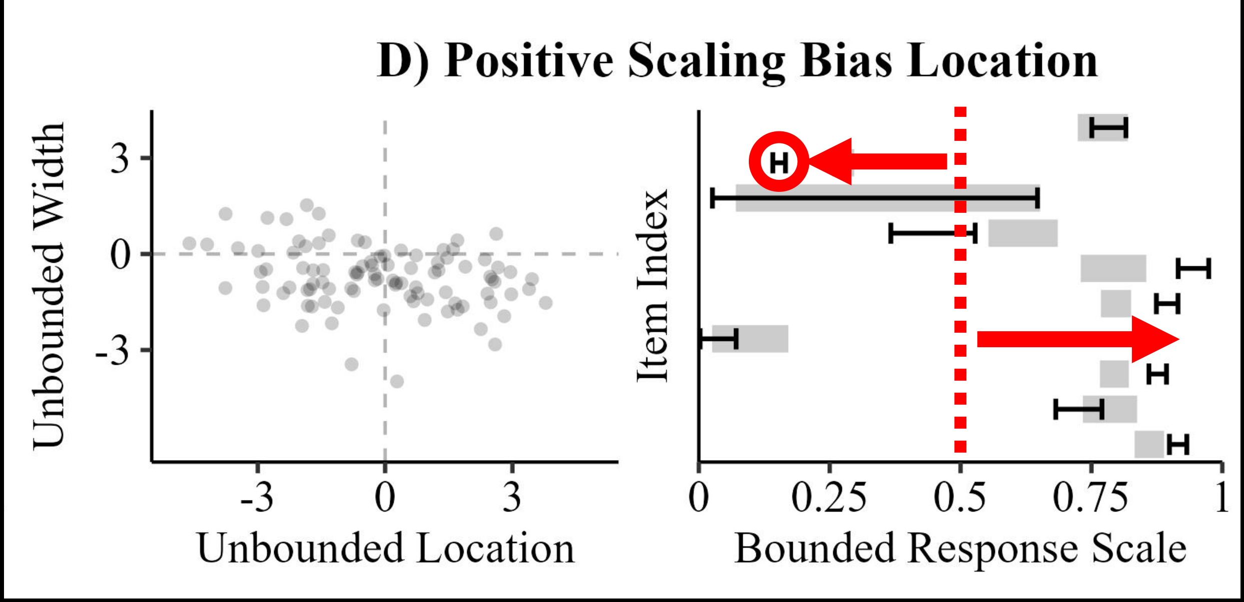


Positive **location** scaling bias

Biases



No biases, High proficiencies

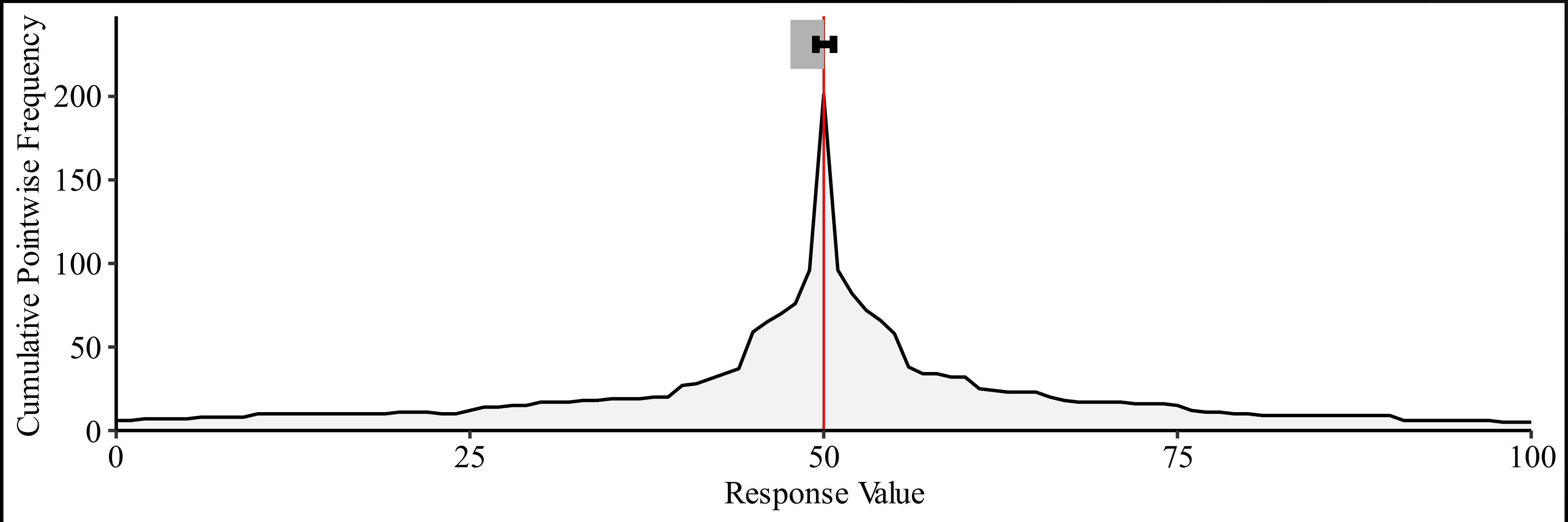


Positive **location** scaling bias

Empirical Example: Verbal Quantifiers

„50-50 Chance“

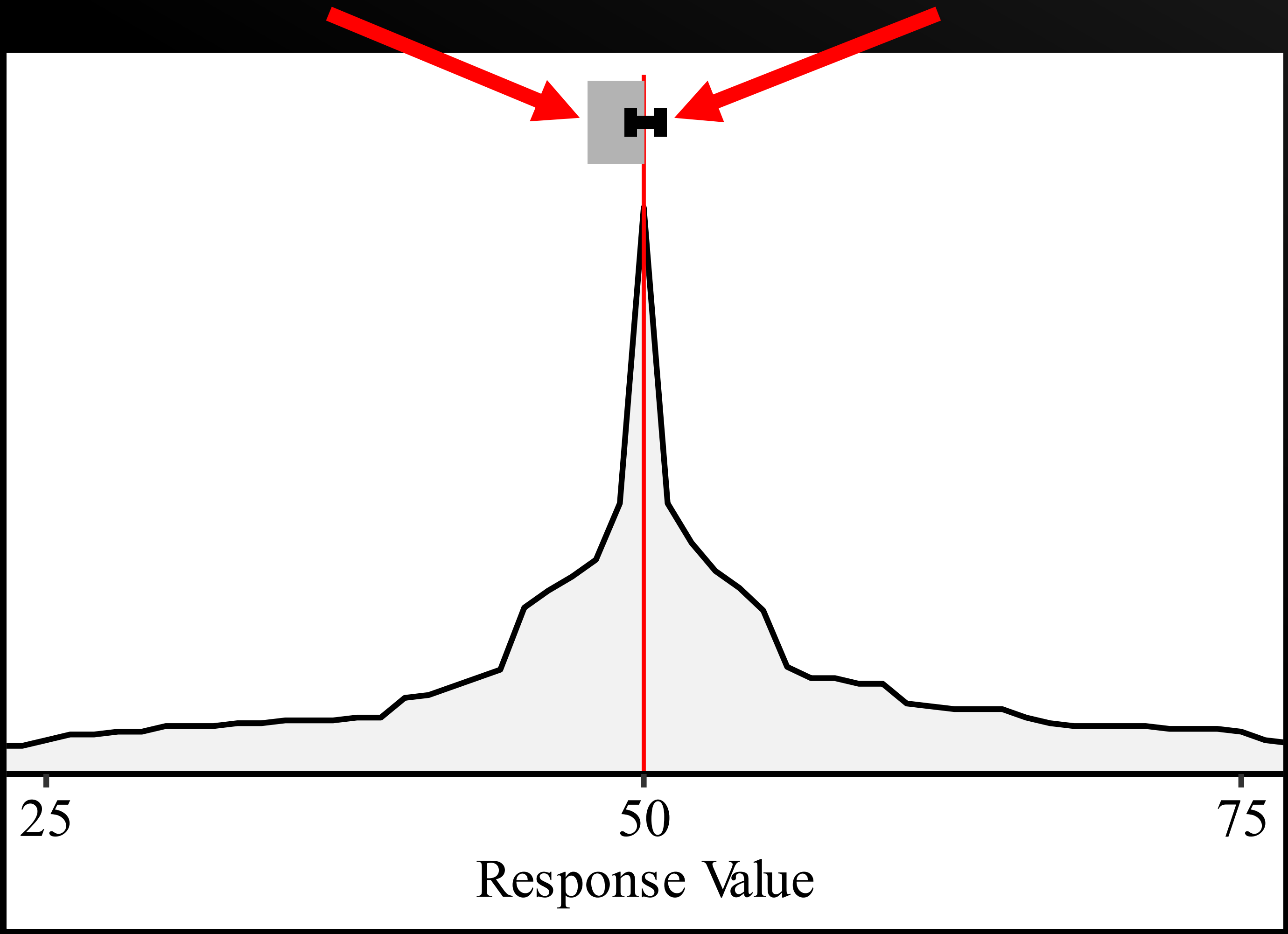
- Control item in the study
- Should be centered on x = 50 and narrow



Empirical Example: Verbal Quantifiers

Mean of logit-transformed location and width

Estimated consensus interval



Take-Home Points

- Isometric log-ratio transformation makes interval responses **more suitable** for modeling frameworks using **normal** distributions

- We can estimate weighted consensus intervals to smooth out extreme ratings

Thanks to:

- Prof. Dr. Daniel W. Heck

- Björn Siepe

Contact:

kloft@uni-marburg.de

Slides:

https://github.com/matthiaskloft/

References

- Anders, R., Oravecz, Z., & Batchelder, W. H. (2014). Cultural consensus theory for continuous responses: A latent appraisal model for information pooling. *Journal of Mathematical Psychology, 61*, 1–13. https://doi.org/10.1016/j.jmp.2014.06.001
- **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
- **Kloft, M.**, Siepe, B.S., & Heck, D.W. (2024). *The interval truth model: A consensus model for continuous bounded interval responses*. PsyArXiv. https://doi.org/10.31234/osf.io/dzvw2
- 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

Model Estimation

Priors: Latent Consensus

Concept:

- Specify **priors** (i.e. sample parameters) on the original **bounded** scale
- **Transform** the sampled parameters to the unbounded space to insert them in the bivariate normal model

Priors: Latent Consensus

Weakly informative prior on marginal distribution of interval width:

$$T_j^{wid(0,1)} \sim \mathrm{Beta}(1.2, 3)$$

Priors: Latent Consensus

Uninformative prior on distribution of location shift, conditional on a particular width:

$$s_j \sim \text{Beta}(1, 1)$$

$$T_j^{loc(0,1)} = s_j \left(1 - T_j^{wid(0,1)}\right) + \frac{T_j^{wid(0,1)}}{2}$$

Priors: Latent Consensus

Transform the bounded intervals to the unbounded space:

$$\begin{pmatrix} \mathrm{T_j^{loc}} \\ \mathrm{T_j^{wid}} \end{pmatrix} = \mathrm{ILR}\begin{pmatrix} T_j^{loc(0,1)} - \dfrac{T_j^{wid(0,1)}}{2} \\ T_j^{loc(0,1)} + \dfrac{T_j^{wid(0,1)}}{2} \end{pmatrix}$$