PSYCHOMETRIKA—VOL. 86, NO. 2, 378–403
JUNE 2021
https://doi.org/10.1007/s11336-021-09762-5

Psychometric Society

MAPPING UNOBSERVED ITEM–RESPONDENT INTERACTIONS: A LATENT SPACE ITEM RESPONSE MODEL WITH INTERACTION MAP




Minjeong Jeon
UNIVERSITY OF CALIFORNIA, LOS ANGELES

Ick Hoon Jin
YONSEI UNIVERSITY

Michael Schweinberger
RICE UNIVERSITY

Samuel Baugh
UNIVERSITY OF CALIFORNIA, LOS ANGELES
LAWRENCE BERKELEY NATIONAL LAB

Reporter: Yingshi Huang

Introduction

- Item response theory (IRT)
  - respondent j & item i

$$\operatorname{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j, \ \beta_i)) \ = \alpha_j + \beta_i$$

Introduction

- Item response theory (IRT)
  - respondent j & item i

    $$\text{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j, \ \beta_i)) = \alpha_j + \beta_i$$

1. conditional independence (between respondents, between items)
2. consistency of success probability (same ability, same easiness)

Introduction

- Item response theory (IRT)
  - respondent j & item i

$$\text{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j,\ \beta_i)) = \alpha_j + \beta_i$$

1. conditional independence (between respondents, between items)
2. consistency of success probability (same ability, same easiness)

unobserved interaction / dependence

(same ability, different cultures)

Introduction

- Item response theory (IRT)
  - respondent j & item i

$$\operatorname{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j,\ \beta_i)) = \alpha_j + \beta_i$$

1. conditional independence (between respondents, between items)
2. consistency of success probability (same ability, same easiness)

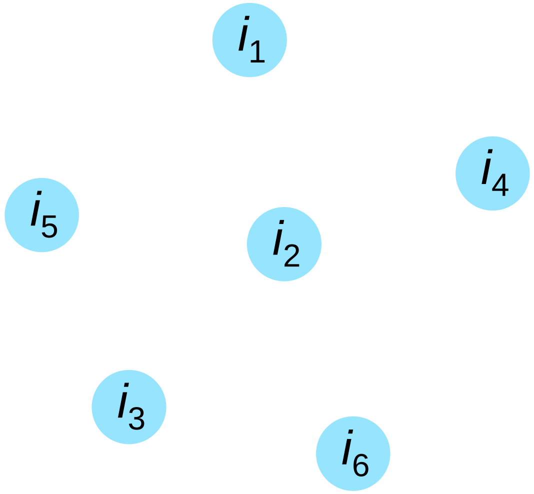
unobserved interaction / dependence

- Purpose:
  - introduce a novel latent space model

➡ represent interactions of respondents and items

Latent Space Model

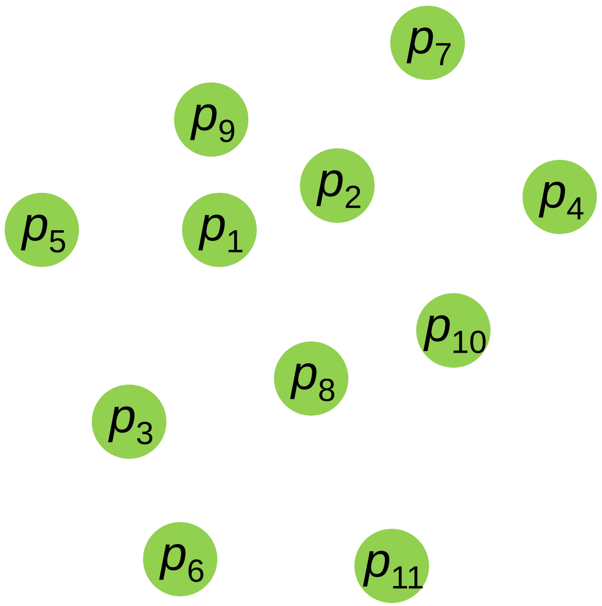
- How to interpret dependence?



Items answered by examinee *j*

Latent Space Model

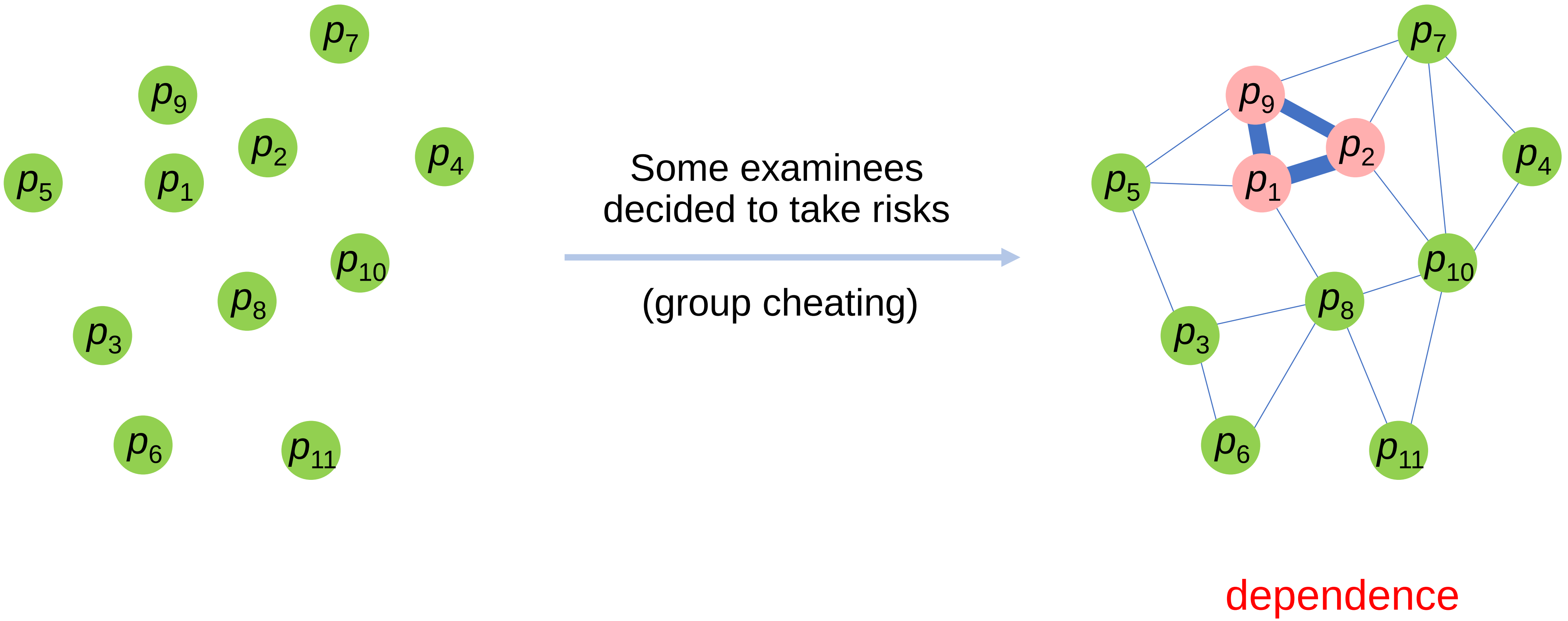
- How to interpret dependence?



Responses of different examinees on item i

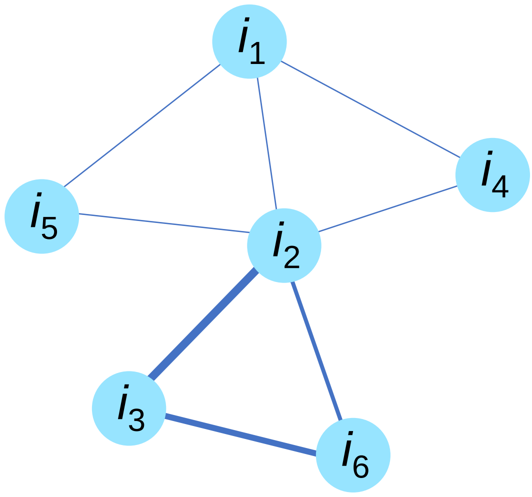
Latent Space Model

- How to interpret dependence: under a network structure

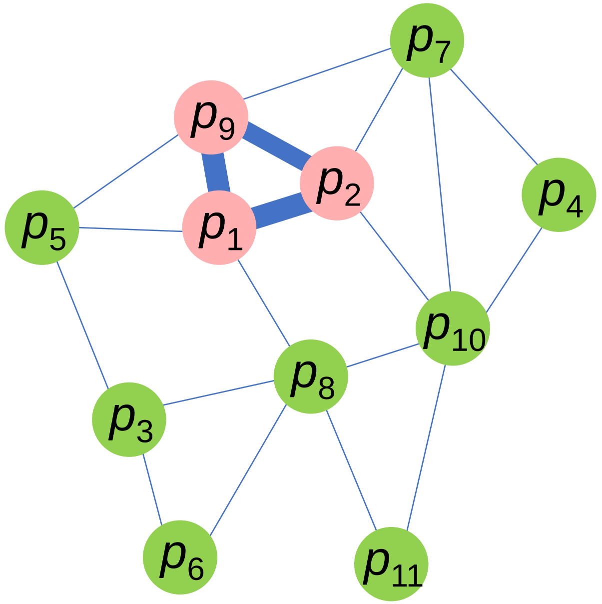


Doubly Latent Space Model

- Capture local item and person dependence: doubly LSM



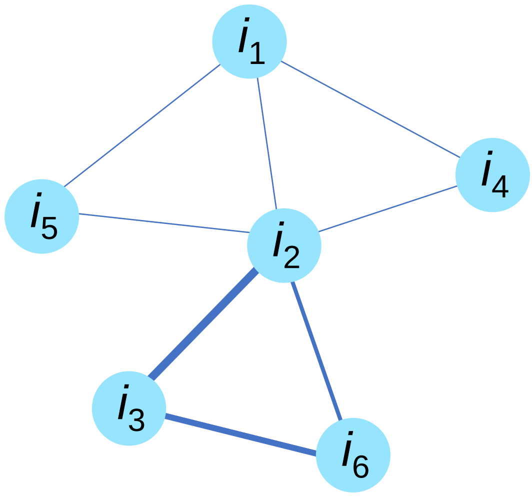
Items answered by examinee j



Responses of different examinees on item i

Doubly Latent Space Model

- Capture local item and person dependence: doubly LSM



Items answered by examinee j

item pairs

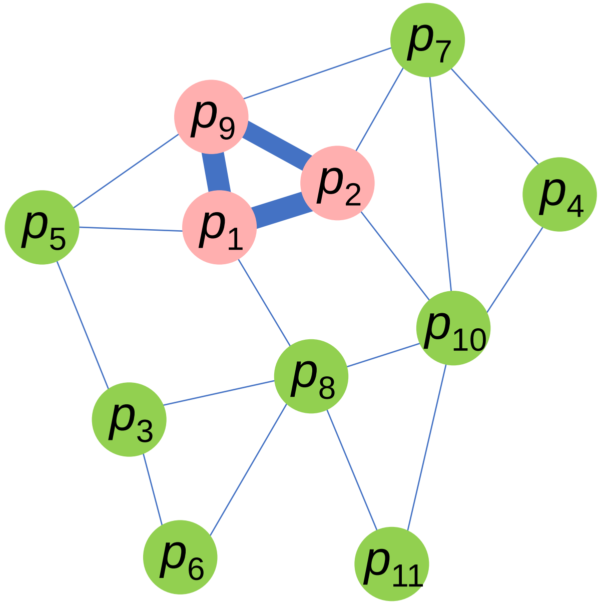
	1	2	3	4	5	6
1	0	0	0	0	0	0
2	0	0	**1**	0	0	**1**
3	0	**1**	0	0	0	**1**
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	0	**1**	**1**	0	0	0

examinee j who answers item 2, 3, and 6 correctly

$\boldsymbol{U}_{j,I\times I} = \{x_{ji}x_{jk}\}$ (item i and item k)

Doubly Latent Space Model

- Capture local item and person dependence: doubly LSM



Responses of different examinees on item i

examinee pairs

	1	2	…	9	…	11
1	0	**1**		**1**		0
2	**1**	0	…	**1**		0
…		…		**1**	…	0
9	**1**	**1**	**1**	0		0
…			…			
11	0	0	0	0		0

item i being answered correctly only by examinee 1, 2, and 9

$\boldsymbol{Y}_{i,N\times N} = \{x_{ji}x_{li}\}$ (examinee j and examinee l)

Ick Hoon Jin, Minjeong Jeon, 2019 Psychometrika

Doubly Latent Space Model

- How to model the probability of a relation between nodes: embedding

the concept of “social space”

a space of unobserved latent characteristics that represent potential transitive tendencies in network relations

Peter D. Hoff, Adrian E. Raftery, Mark S. Handcock, 2002 JASA
Ick Hoon Jin, Minjeong Jeon, 2019 Psychometrika

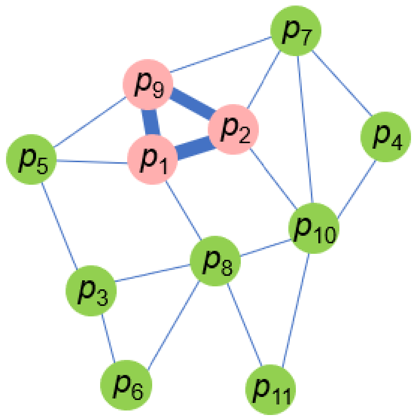
Doubly Latent Space Model

- How to model the probability of a relation between nodes: embedding

the concept of “social space”

a space of unobserved latent characteristics that represent potential transitive tendencies in network relations

− an (maybe unappropriated) illustration:



These students came from the same school

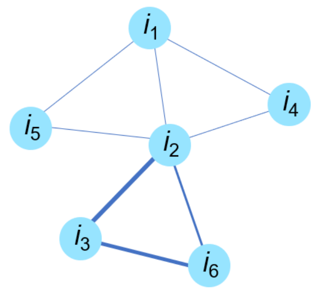
(social space: school)

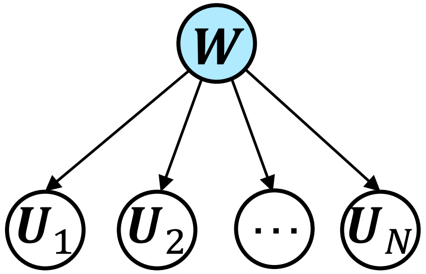
Peter D. Hoff, Adrian E. Raftery, Mark S. Handcock, 2002 JASA
Ick Hoon Jin, Minjeong Jeon, 2019 Psychometrika

Doubly Latent Space Model

- *p*-dimensional latent space (typically 2-3 dimensions)

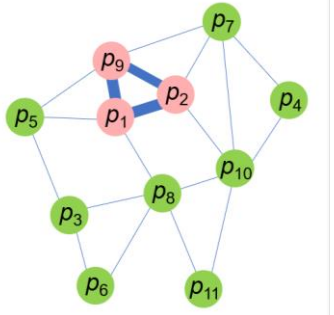
$\boldsymbol{U}_{j,I\times I} = \{x_{ji}x_{jk}\}$

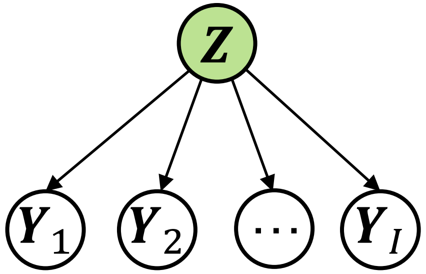




multiple networks

$\boldsymbol{Y}_{i,N\times N} = \{x_{ji}x_{li}\}$



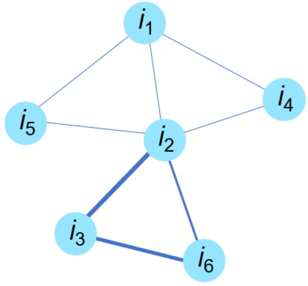


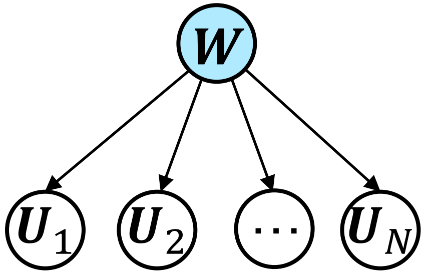
multiple networks

Doubly Latent Space Model

- *p*-dimensional latent space (typically 2-3 dimensions)

$\boldsymbol{U}_{j,I\times I} = \{x_{ji}x_{jk}\}$

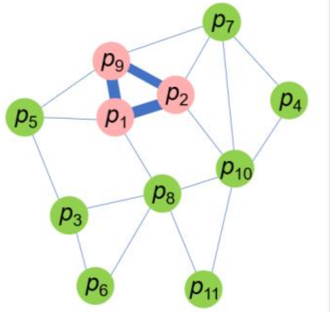


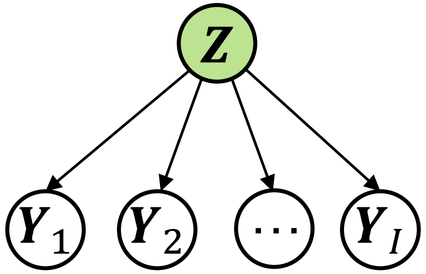


multiple networks

$$P(\boldsymbol{U}|\boldsymbol{W},\boldsymbol{\alpha}) = \prod_{j=1}^{N} P(\boldsymbol{U}_j|\boldsymbol{W};\alpha_j) = \prod_{j=1}^{N}\prod_{i\neq k} \frac{\exp(\alpha_j - \|\boldsymbol{w}_i - \boldsymbol{w}_k\|)^{u_{j,ik}}}{1+\exp(\alpha_j - \|\boldsymbol{w}_i - \boldsymbol{w}_k\|)}$$

$\boldsymbol{Y}_{i,N\times N} = \{x_{ji}x_{li}\}$





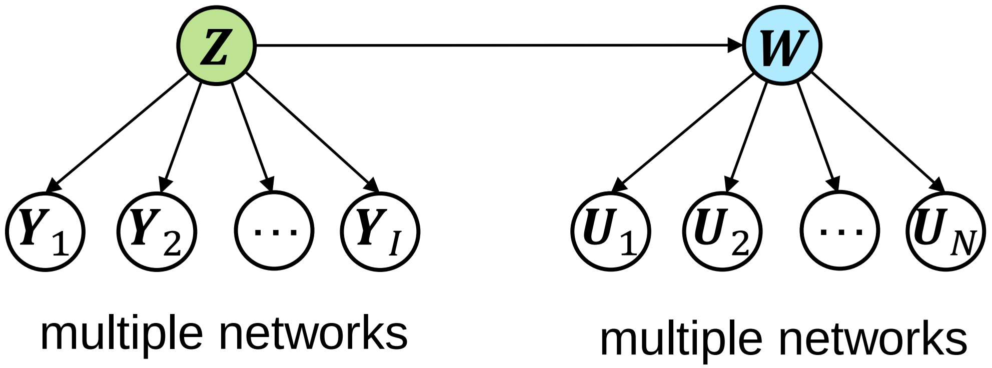
multiple networks

$$P(\boldsymbol{Y}|\boldsymbol{Z},\boldsymbol{\beta}) = \prod_{i=1}^{I} P(\boldsymbol{Y}_i|\boldsymbol{Z};\beta_i) = \prod_{i=1}^{I}\prod_{j\neq l} \frac{\exp(\beta_i - \|\boldsymbol{z}_j - \boldsymbol{z}_l\|)^{y_{i,jl}}}{1+\exp(\beta_i - \|\boldsymbol{z}_j - \boldsymbol{z}_l\|)}$$

Ick Hoon Jin, Minjeong Jeon, 2019 Psychometrika

Latent Space Item Response Model

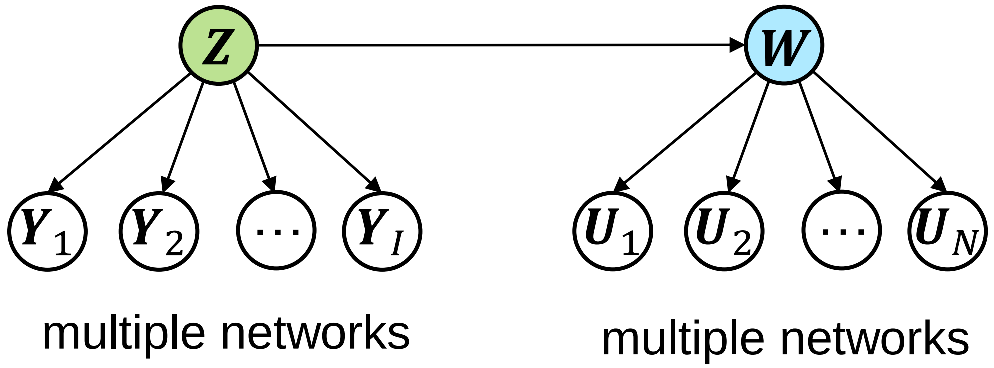
- A concise version is needed



1. work with *functions* of item response data
2. deal with *multiple* networks
3. must *combine* two LSMs for simultaneous estimation: $\boldsymbol{w}_j = f_j(\boldsymbol{Z})$

Latent Space Item Response Model

- A concise version is needed



1. work with *functions* of item response data
2. deal with *multiple* networks
3. must *combine* two LSMs for simultaneous estimation: $\boldsymbol{w}_j = f_j(\boldsymbol{Z})$

work with original response data & single network

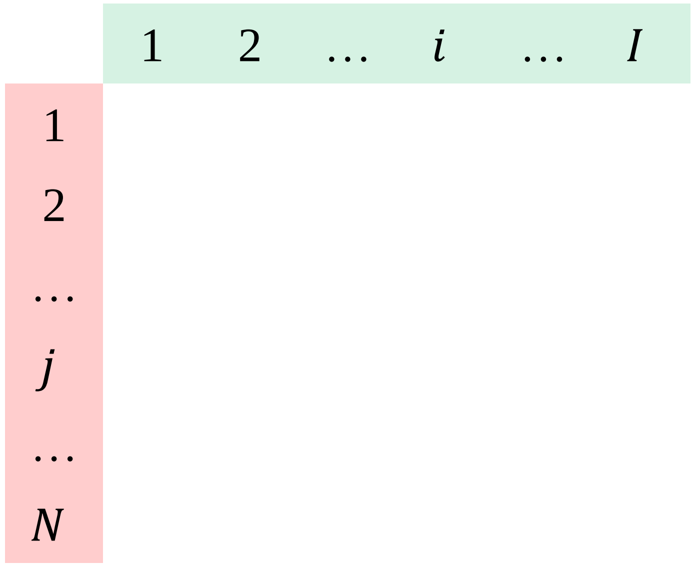
Latent Space Item Response Model

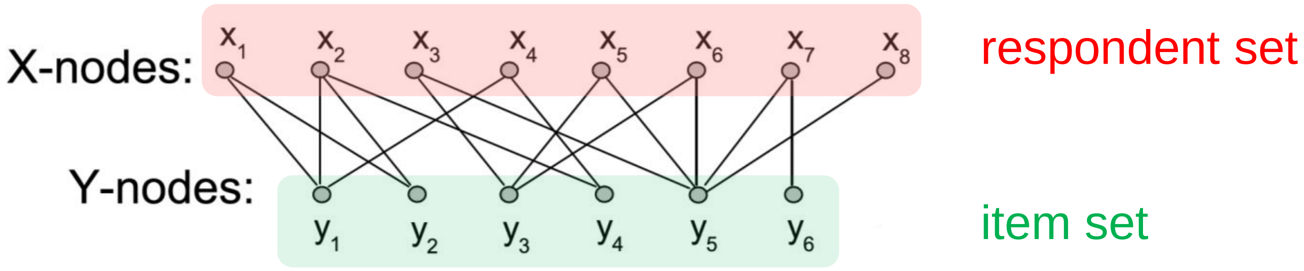
- Work with original response data & single network
  - original response data: $\boldsymbol{Y} \in \{0, 1\}^{N \times I}$
  - a single network?

	1	2	…	i	…	I
1						
2						
…						
j						
…						
N						

Latent Space Item Response Model

- Work with original response data & single network
  - original response data: $\boldsymbol{Y} \in \{0, 1\}^{N \times I}$
  - a single network?



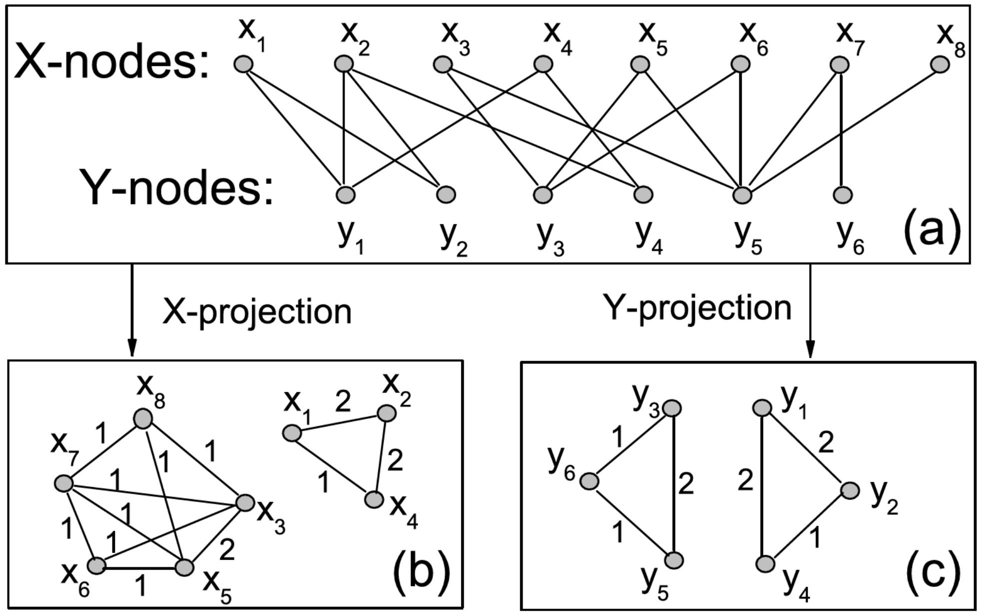


Latent Space Item Response Model

- Work with original response data & single network
  - original response data: $\boldsymbol{Y} \in \{0, 1\}^{N \times I}$
  - a single network?

	1	2	…	*i*	…	*I*
1						
2						
…						
j						
…						
N						

bipartite network



Tao Zhou, Jie Ren, Matus Medo, Yi-Cheng Zhang, 2007 Physical Review E

Latent Space Item Response Model

- Work with original response data & single network
  - original response data: $\boldsymbol{Y} \in \{0, 1\}^{N \times I}$
  - a single network: $\operatorname{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j, \beta_i, \boldsymbol{a}_j, \boldsymbol{b}_i)) = \alpha_j + \beta_i + g(\boldsymbol{a}_j, \boldsymbol{b}_i)$

the position of
respondent j and item i

Latent Space Item Response Model

- Work with original response data & single network
  - original response data: $\boldsymbol{Y} \in \{0, 1\}^{N \times I}$
  - a single network: $\text{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j, \beta_i, \boldsymbol{a}_j, \boldsymbol{b}_i)) = \alpha_j + \beta_i + g(\boldsymbol{a}_j, \boldsymbol{b}_i)$

the position of respondent j and item i

1. multiplicative effect $g(\boldsymbol{a}_j, \boldsymbol{b}_i) = \boldsymbol{a}_j^\top \boldsymbol{b}_i$
2. distance effect $g(\boldsymbol{a}_j, \boldsymbol{b}_i) = -\gamma \, d(\boldsymbol{a}_j, \boldsymbol{b}_i)$

weight (≥ 0) distance

Latent Space Item Response Model

- Work with original response data & single network
  - original response data: $\boldsymbol{Y} \in \{0, 1\}^{N \times I}$
  - a single network: $\text{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j, \beta_i, \boldsymbol{a}_j, \boldsymbol{b}_i)) = \alpha_j + \beta_i + g(\boldsymbol{a}_j, \boldsymbol{b}_i)$

the position of respondent j and item i

1. multiplicative effect $g(\boldsymbol{a}_j, \boldsymbol{b}_i) = \boldsymbol{a}_j^\top \boldsymbol{b}_i$
2. distance effect $g(\boldsymbol{a}_j, \boldsymbol{b}_i) = -\gamma \, d(\boldsymbol{a}_j, \boldsymbol{b}_i)$

weight (≥ 0) distance

a) ℓ_1-distance: $d(\boldsymbol{a}, \boldsymbol{b}) = ||\boldsymbol{a} - \boldsymbol{b}||_1 = \sum_{i=1}^{p} |a_i - b_i|$

b) ℓ_2-distance: $d(\boldsymbol{a}, \boldsymbol{b}) = ||\boldsymbol{a} - \boldsymbol{b}||_2 = \sqrt{\sum_{i=1}^{p} (a_i - b_i)^2}$

c) ℓ_∞-distance: $d(\boldsymbol{a}, \boldsymbol{b}) = ||\boldsymbol{a} - \boldsymbol{b}||_\infty = \max_{1 \leq i \leq p} |a_i - b_i|$

- reflexivity: $d(\boldsymbol{a}, \boldsymbol{b}) = 0$ if and only if $\boldsymbol{a} = \boldsymbol{b} \in \mathbb{M}$;
- symmetry: $d(\boldsymbol{a}, \boldsymbol{b}) = d(\boldsymbol{b}, \boldsymbol{a})$ for all $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{M}$;
- triangle inequality: $d(\boldsymbol{a}, \boldsymbol{b}) \leq d(\boldsymbol{a}, \boldsymbol{c}) + d(\boldsymbol{b}, \boldsymbol{c})$ for all $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{M}$.

Latent Space Item Response Model

- Work with original response data & single network
  - original response data: $\boldsymbol{Y} \in \{0, 1\}^{N \times I}$
  - a single network: $\text{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j, \beta_i, \boldsymbol{a}_j, \boldsymbol{b}_i)) = \alpha_j + \beta_i + g(\boldsymbol{a}_j, \boldsymbol{b}_i)$

the position of respondent j and item i

1. multiplicative effect $g(\boldsymbol{a}_j, \boldsymbol{b}_i) = \boldsymbol{a}_j^\top \boldsymbol{b}_i$
2. distance effect $g(\boldsymbol{a}_j, \boldsymbol{b}_i) = -\gamma \, d(\boldsymbol{a}_j, \boldsymbol{b}_i)$

weight (≥ 0) distance

$\gamma = 0$: Rasch model
$\gamma > 0$: capture deviations from the main effects
(embed respondents into a shared metric space)

Latent Space Item Response Model

- The distance effect is easier to interpret than the multiplicative effect

$$\boldsymbol{a}_j = (0,\ 1/100) \qquad \boldsymbol{a}_j = (0,\ 100)$$

$$\boldsymbol{b}_i = (1/100,\ 0) \qquad \boldsymbol{b}_i = (100,\ 0)$$

Latent Space Item Response Model

- The distance effect is easier to interpret than the multiplicative effect

$$\boldsymbol{a}_j = (0,\ 1/100) \qquad \boldsymbol{a}_j = (0,\ 100)$$

$$\boldsymbol{b}_i = (1/100,\ 0) \qquad \boldsymbol{b}_i = (100,\ 0)$$

1. multiplicative effect $\boldsymbol{a}_j^\top \boldsymbol{b}_i = 0$ $\qquad$ $\boldsymbol{a}_j^\top \boldsymbol{b}_i = 0$

2. distance effect (ℓ_2-distance) $d(\boldsymbol{a}_j, \boldsymbol{b}_i) = 0.01$ $\qquad$ $d(\boldsymbol{a}_j, \boldsymbol{b}_i) = 141.42$

angle & length!

Properties

- Two main lines of research involving networks

1. Graphical models (model structure):

   vertices: the variables of interest

   edge: conditional dependencies among variables

2. Random graph models (data structure):

   vertices: individuals

   edge: friendships among individuals / correct responses

Properties

- Practical advantages
  - provides a geometric representation of interactions

Properties

- Practical advantages
  - provides a geometric representation of interactions

$$\text{logit}(\mathbb{P}(y_{j,i} = 1 \mid \alpha_j, \beta_i, \zeta_{j,i})) = \alpha_j + \beta_i + \zeta_{j,i}$$

1. Rasch:

   14 items & 200 respondents $\zeta_{j,i} = 0$

2. Rasch with local dependence:

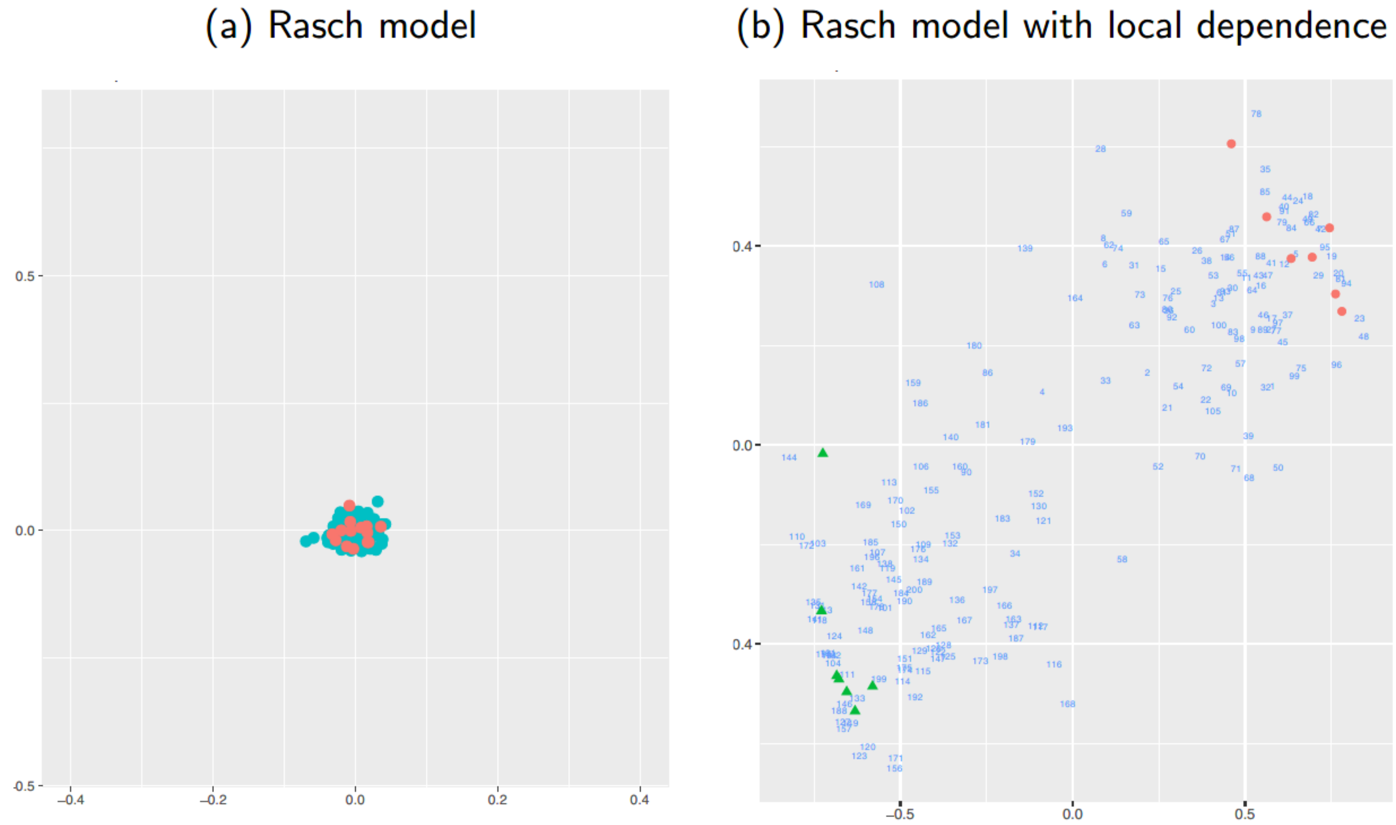
   first 100 respondents $\rightarrow$ item 1-7 $\zeta_{j,i} \sim N(2, 0.2)$ item 8-14 $\zeta_{j,i} = 0$

   last 100 respondents $\rightarrow$ item 8-14 $\zeta_{j,i} \sim N(2, 0.2)$ item 1-7 $\zeta_{j,i} = 0$

latent space dimension: $\mathbb{M} = \mathbb{R}^2$

Properties

- Practical advantages
  - provides a geometric representation of interactions



red: item 1-7
green: item 8-14
blue: respondents

Properties

- Theoretical advantages
  - weaker conditional independence assumption

$$\mathbb{P}(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{\alpha},\ \boldsymbol{\beta},\ \gamma,\ \boldsymbol{A},\ \boldsymbol{B}) \ = \prod_{j=1}^{N} \prod_{i=1}^{I} \mathbb{P}(Y_{j,i} = y_{j,i} \mid \alpha_j,\ \beta_i,\ \gamma,\ \boldsymbol{a}_j,\ \boldsymbol{b}_i)$$

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N), \boldsymbol{\beta} = (\beta_1, \ldots, \beta_I), \boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N), \text{and } \boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_I)$$

allow for respondent–item interactions:

✓ testlets (e.g., items similar in content)

✓ learning and practice effects

✓ repeated measurements

✓ nested respondents (shared school or family memberships)

Properties

- Theoretical advantages
  - drop some of the homogeneity assumptions

    same abilities: respondents j_1 and j_2

    different distances from item i: $d(\boldsymbol{a}_{j_1}, \boldsymbol{b}_i) < d(\boldsymbol{a}_{j_2}, \boldsymbol{b}_i)$

Properties

- Theoretical advantages
  - drop some of the homogeneity assumptions

    same abilities: respondents j_1 and j_2

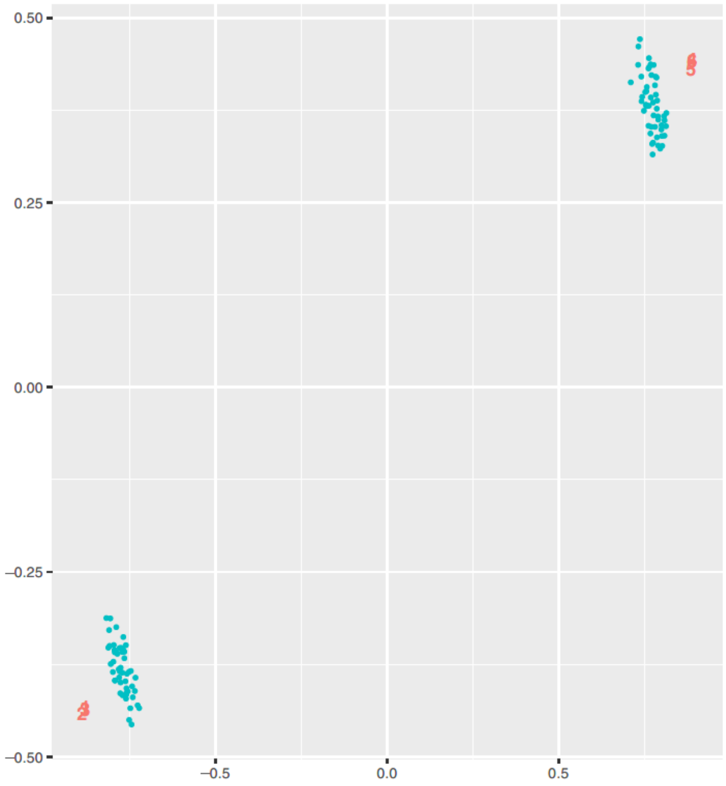
    different distances from item i: $d(\boldsymbol{a}_{j_1}, \boldsymbol{b}_i) < d(\boldsymbol{a}_{j_2}, \boldsymbol{b}_i)$

	algebra items			geometry items		
Response	I1	I2	I3	I4	I5	I6
1	1	1	1	0	0	0
2	1	1	1	0	0	0
3	0	0	0	1	1	1
4	0	0	0	1	1	1

Properties

- Theoretical advantages
  - drop some of the homogeneity assumptions

(a) Two response patterns



Respondents 51-100 give correct responses to Items 4-6 only

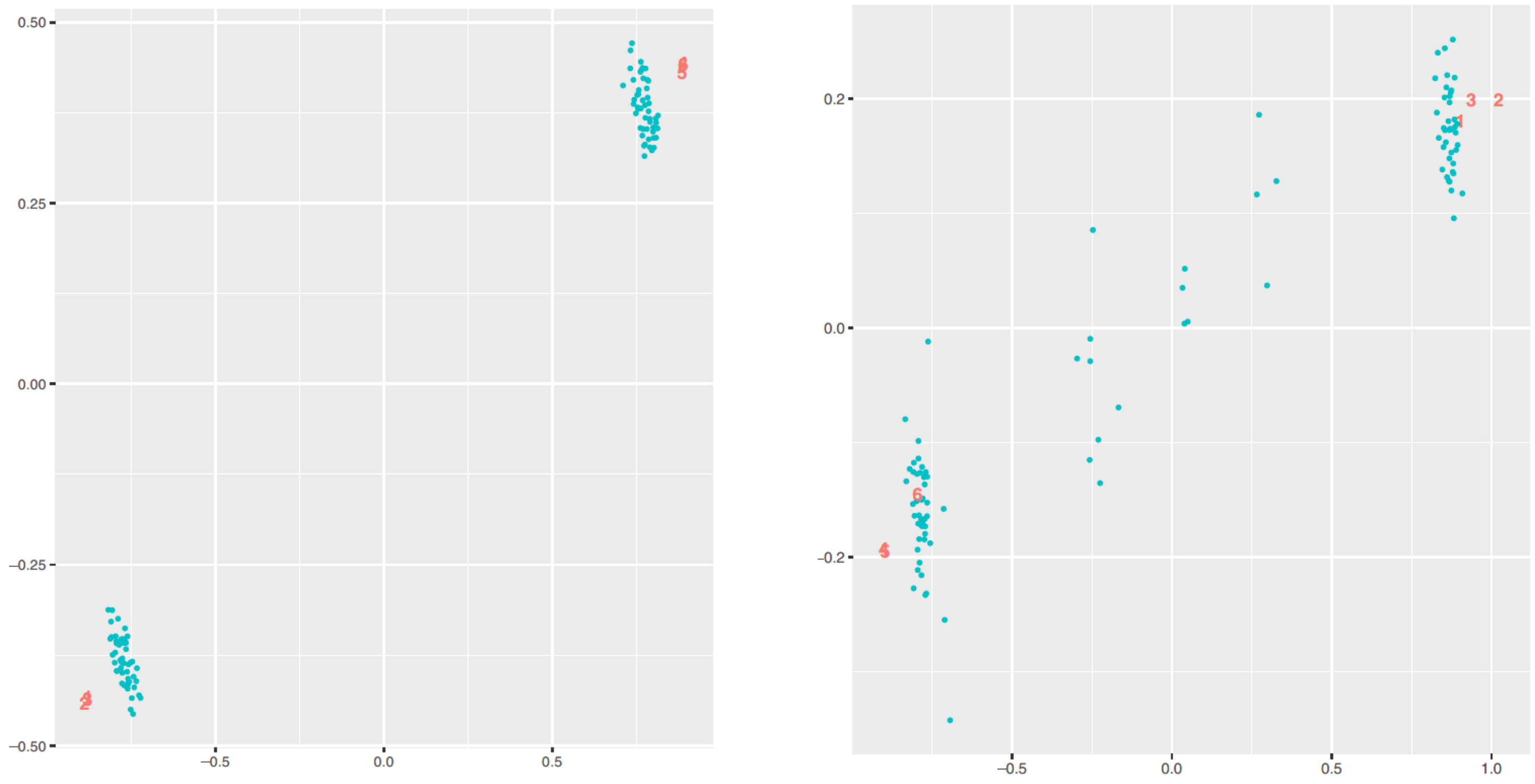
Respondents 1–50 give correct responses to Items 1–3 only

Properties

- Theoretical advantages
  - drop some of the homogeneity assumptions

(a) Two response patterns

(b) Two response patterns, with randomness



80 have two patterns
+
20 give random responses

Related Models

- Other models with relaxed assumptions
  - polytomous item models
  - testlet and bifactor models
  - finite mixture models
  - …

1. require knowledge of the interaction structure
2. heterogeneity between latent classes, but assume homogeneity within latent classes

Related Models

- Other models with interactions among respondents and items
  - two-parameter IRT model

$$\text{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j, \beta_i, \lambda_i)) = \lambda_i \alpha_j + \beta_i$$

➡ cannot visualize interactions

Related Models

- Other models with interactions among respondents and items
  - two-parameter IRT model

$$\text{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j,\, \beta_i,\, \lambda_i)) \;=\; \lambda_i\, \alpha_j + \beta_i$$

➡ cannot visualize interactions

  - interaction IRT model

$$\text{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j,\, \beta_i,\, \epsilon_{j,i})) \;=\; \alpha_j + \beta_i + \epsilon_{j,i}$$

latent space model can be viewed as a special case: $\epsilon_{j,i} \;=\; -\gamma\, d(\boldsymbol{a}_j,\, \boldsymbol{b}_i)$

➡ the restrictions 1. facilitate estimation
2. make sense in practice (capture transitivity)

Related Models

- Other models with interactions among respondents and items
  - bilinear mixed effects models and related models

    ➡ the multiplicative effects version

  - differential item functioning

    an interaction term is formed with a known categorical attribute of respondents (e.g., gender) and an item indicator

    ➡ require pre-knowledge

Bayesian Inference

- Markov Chain Monte Carlo (MCMC)

$$\text{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j, \beta_i, \boldsymbol{a}_j, \boldsymbol{b}_i)) = \alpha_j + \beta_i + g(\boldsymbol{a}_j, \boldsymbol{b}_i)$$

$$g(\boldsymbol{a}_j, \boldsymbol{b}_i) = -\gamma \; d(\boldsymbol{a}_j, \boldsymbol{b}_i)$$

  - (Posterior)~ (Prior) (Likelihood)

Bayesian Inference

- Markov Chain Monte Carlo (MCMC)

$$\operatorname{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j, \beta_i, \boldsymbol{a}_j, \boldsymbol{b}_i)) = \alpha_j + \beta_i + g(\boldsymbol{a}_j, \boldsymbol{b}_i)$$

$$g(\boldsymbol{a}_j, \boldsymbol{b}_i) = -\gamma \ d(\boldsymbol{a}_j, \boldsymbol{b}_i)$$

- (Posterior)~ (Prior) (Likelihood)

$$f(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, \boldsymbol{A}, \boldsymbol{B} \mid \boldsymbol{y}) \propto \left[\prod_{j=1}^{N} f(\alpha_j)\right] \left[\prod_{i=1}^{I} f(\beta_i)\right] f(\gamma) \left[\prod_{j=1}^{N} f(\boldsymbol{a}_j)\right] \left[\prod_{i=1}^{I} f(\boldsymbol{b}_i)\right] \times \left[\prod_{j=1}^{N}\prod_{i=1}^{I} \mathbb{P}\Big(Y_{j,i} = y_{j,i} \mid \alpha_j, \beta_i, \gamma, \boldsymbol{a}_j, \boldsymbol{b}_i\Big)\right],$$

Bayesian Inference

- Markov Chain Monte Carlo (MCMC)

$$\operatorname{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j,\ \beta_i,\ \boldsymbol{a}_j,\ \boldsymbol{b}_i)) \ = \ \alpha_j + \beta_i + g(\boldsymbol{a}_j,\ \boldsymbol{b}_i)$$

$$g(\boldsymbol{a}_j,\ \boldsymbol{b}_i) \ = \ -\gamma \ d(\boldsymbol{a}_j, \boldsymbol{b}_i)$$

  - (Posterior)~ (Prior) (Likelihood)

priors:

$$\begin{aligned}
\alpha_j \mid \sigma^2 &\overset{\text{ind}}{\sim} \mathrm{N}\left(0,\ \sigma^2\right), \quad \sigma^2 > 0, \quad j = 1, \ldots, N \\
\beta_i \mid \tau_\beta^2 &\overset{\text{ind}}{\sim} \mathrm{N}\left(0,\ \tau_\beta^2\right), \quad \tau_\beta^2 > 0, \quad i = 1, \ldots, I \\
\log \gamma \mid \mu_\gamma,\ \tau_\gamma^2 &\sim \mathrm{N}\left(\mu_\gamma,\ \tau_\gamma^2\right), \quad \mu_\gamma \in \mathbb{R}, \quad \tau_\gamma^2 > 0 \\
\sigma^2 \mid a_\sigma,\ b_\sigma &\sim \text{Inv-Gamma}\left(a_\sigma,\ b_\sigma\right), \quad a_\sigma > 0, \quad b_\sigma > 0 \\
\boldsymbol{a}_j &\overset{\text{iid}}{\sim} \mathrm{MVN}_p\left(\boldsymbol{0},\ \boldsymbol{I}_p\right), \quad j = 1, \ldots, N \\
\boldsymbol{b}_i &\overset{\text{iid}}{\sim} \mathrm{MVN}_p\left(\boldsymbol{0},\ \boldsymbol{I}_p\right), \quad i = 1, \ldots, I,
\end{aligned}$$

$$\tau_\beta^2 = 4,\ a_\sigma = 1,\ b_\sigma = 1,\ \mu_\gamma = 0.5,\ \tau_\gamma^2 = 1$$

Bayesian Inference

- Identifiability

a) ℓ_1-distance: $d(\boldsymbol{a},\, \boldsymbol{b}) = ||\boldsymbol{a} - \boldsymbol{b}||_1 = \sum_{i=1}^{p} |a_i - b_i|$

b) ℓ_2-distance: $d(\boldsymbol{a},\, \boldsymbol{b}) = ||\boldsymbol{a} - \boldsymbol{b}||_2 = \sqrt{\sum_{i=1}^{p}(a_i - b_i)^2}$

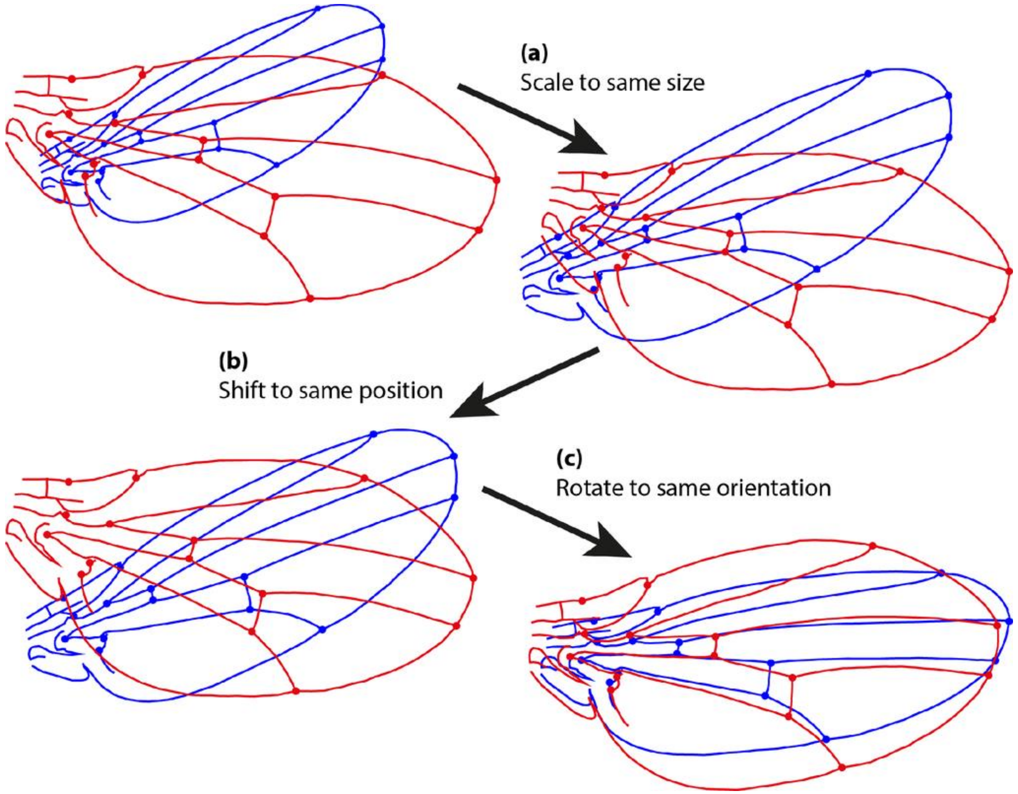
c) ℓ_∞-distance: $d(\boldsymbol{a},\, \boldsymbol{b}) = ||\boldsymbol{a} - \boldsymbol{b}||_\infty = \max_{1 \le i \le p} |a_i - b_i|$

undirected: distances are inherently symmetric

→ invariant to translations, reflections, and rotations

Bayesian Inference

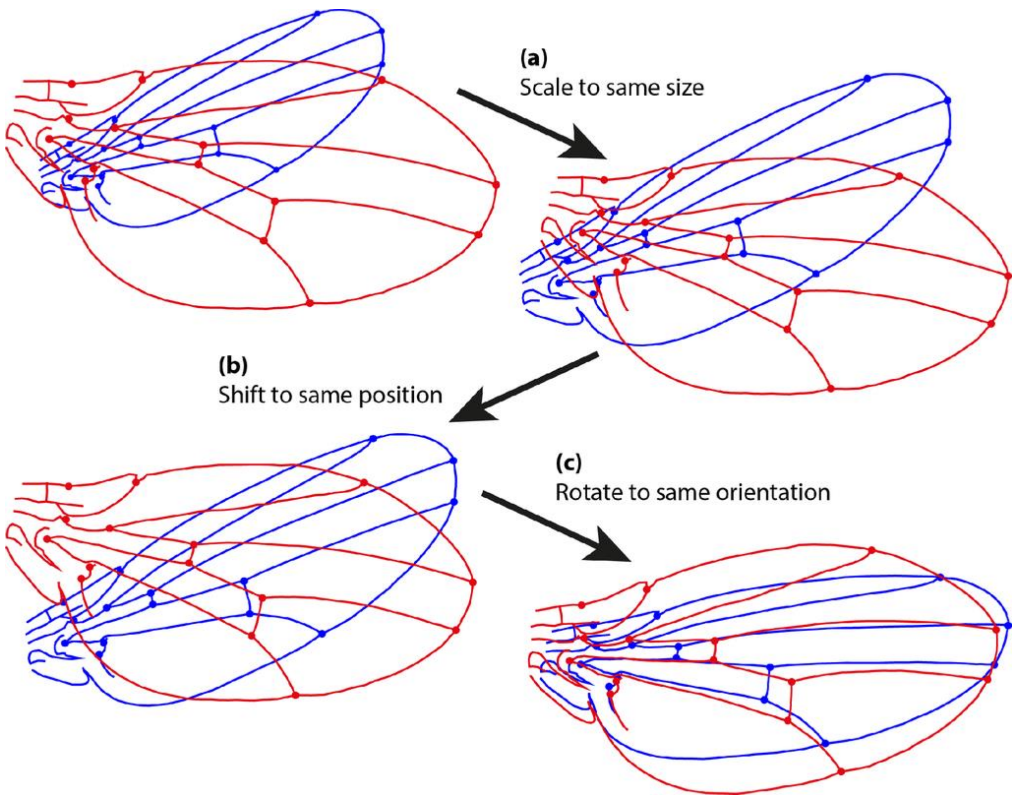
- Identifiability
  - post-processing the MCMC output with Procrustes matching



three transformation steps of an ordinary Procrustes fit
for two configurations (from WiKi)

Bayesian Inference

- Identifiability
  - post-processing the MCMC output with Procrustes matching



three transformation steps of an ordinary Procrustes fit for two configurations (from WiKi)

observed matrix A
target matrix B

find a transformation T, to produce greatest similarity between AT and B

→ relative distances between positions

Bayesian Inference

- Model selection

$$\text{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j, \beta_i, \boldsymbol{a}_j, \boldsymbol{b}_i)) = \alpha_j + \beta_i + g(\boldsymbol{a}_j, \boldsymbol{b}_i)$$

$$g(\boldsymbol{a}_j, \boldsymbol{b}_i) = -\gamma \, d(\boldsymbol{a}_j, \boldsymbol{b}_i)$$

$\gamma = 0$ or $\gamma > 0$?

Bayesian Inference

- Model selection

$$\text{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j, \beta_i, \boldsymbol{a}_j, \boldsymbol{b}_i)) = \alpha_j + \beta_i + g(\boldsymbol{a}_j, \boldsymbol{b}_i)$$

$$g(\boldsymbol{a}_j, \boldsymbol{b}_i) = -\gamma\ d(\boldsymbol{a}_j, \boldsymbol{b}_i)$$

$\gamma = 0$ or $\gamma > 0$?

  - spike-and-slab prior

    γ is likely to be sampled from N(0,0.01) or N(0,1)?

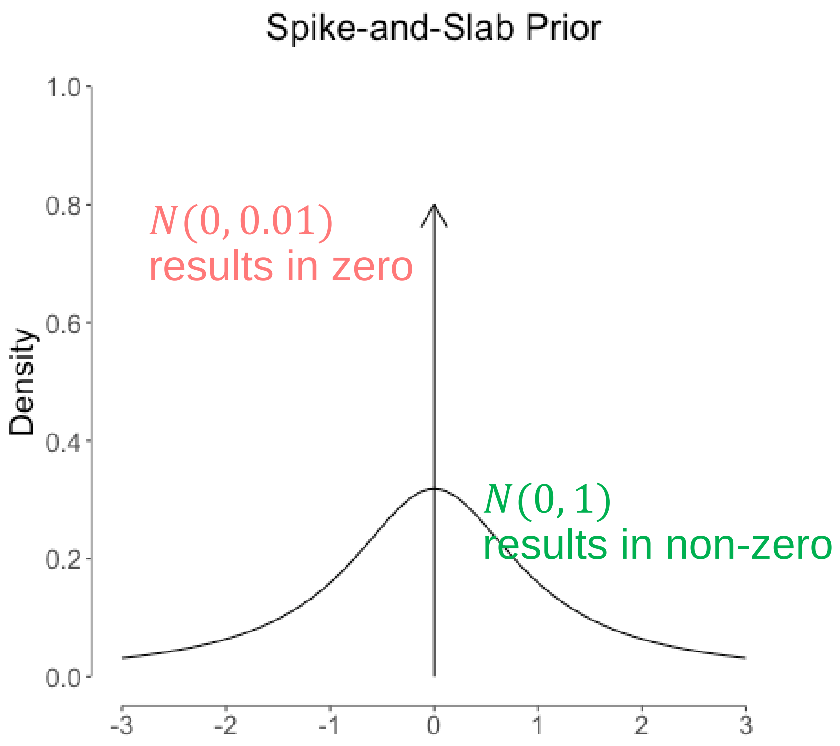


Fig made by Fabian Dablander

Bayesian Inference

- Model selection

$$\text{logit}(\mathbb{P}(Y_{j,i} = 1 \mid \alpha_j,\ \beta_i,\ \boldsymbol{a}_j,\ \boldsymbol{b}_i)) \ = \ \alpha_j + \beta_i + g(\boldsymbol{a}_j,\ \boldsymbol{b}_i)$$

$$g(\boldsymbol{a}_j,\ \boldsymbol{b}_i) \ = -\gamma \ d(\boldsymbol{a}_j, \boldsymbol{b}_i)$$

$\gamma = 0$ or $\gamma > 0$?

- spike-and-slab prior ($\delta \in \{0,1\}$)

$$\log \gamma \ \sim \ (1-\delta)\ N_{spike}(\mu_{\gamma_0},\ \tau_{\gamma_0}^2) + \delta\ N_{slab}(\mu_{\gamma_1},\ \tau_{\gamma_1}^2)$$

$N_{spike}(-3,1)$: mean 0.08, SD 0.01

$N_{slab}(0.5,1)$: mean 2.72, SD 3.56

⇒ $\gamma = 0$, the posterior probability of $\delta = 1 < 0.5$

$\gamma > 0$, otherwise

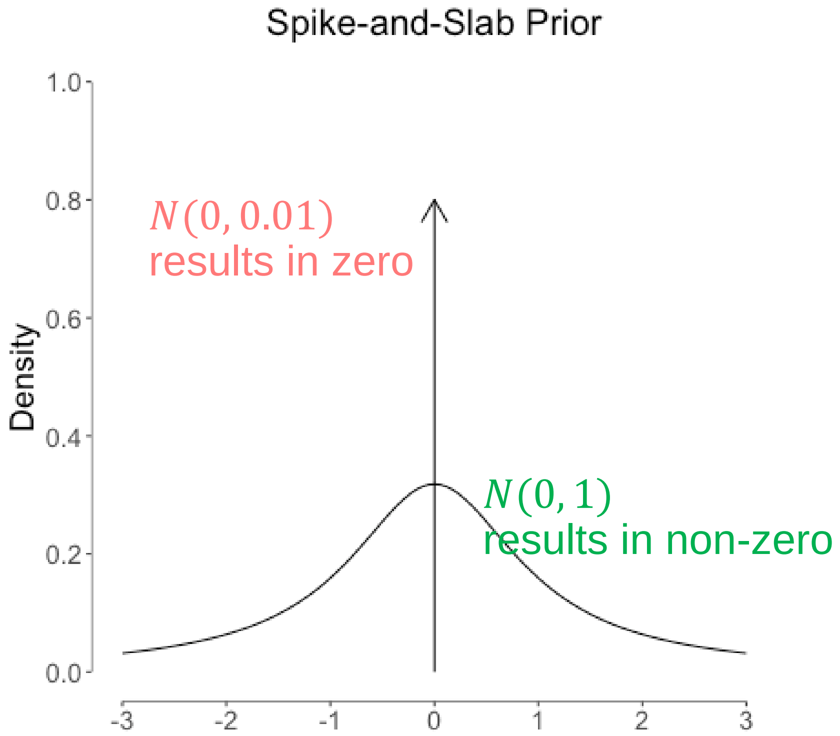


Fig made by Fabian Dablander

Simulation Study

- The accuracy of model selection approach

1. Rasch:

   14 items & 200 respondents

   100 datasets with $\gamma = 0$

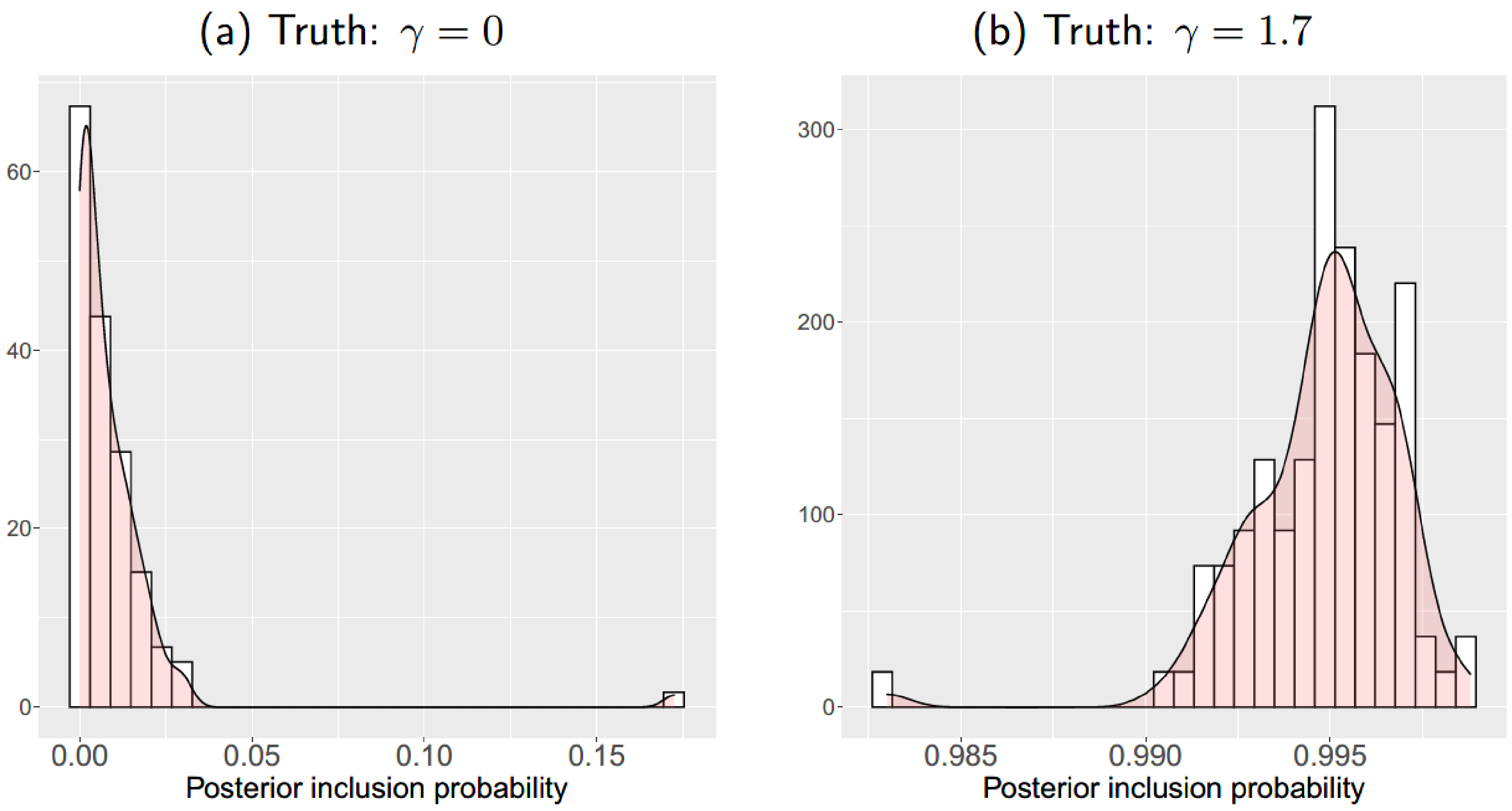
2. latent space model:

   first 100 respondents $\rightarrow$ item 1-7
   last 100 respondents $\rightarrow$ item 8-14

   100 datasets with $\gamma = 1.7$

– compute the proportion of times $\delta = 1$ in the MCMC posterior sample

Simulation Study



Histogram of the estimated posterior probability of the event $\delta = 1$, called "posterior inclusion probability." **a** Data are generated from the Rasch model with $\gamma = 0$. **b** Data are generated from the latent space model with $\gamma = 1.7$

Example 1: Attitudes to Abortion

- Data and estimation

1. The woman decides on her own that she does not wish to have the child
2. The couple agree that they do not wish to have the child
3. The woman is not married and does not wish to marry the man
4. The couple cannot afford any more children.
5. There is a strong chance of a defect in the baby
6. The woman's health is seriously endangered by the pregnancy
7. The woman became pregnant as a result of rape

0.42, 0.52, 0.47, 0.53, 0.86, 0.94, 0.93

– 20,000 iterations (10,000 burn-in)

Example 1: Attitudes to Abortion

- Data and estimation

1. The woman decides on her own that she does not wish to have the child
2. The couple agree that they do not wish to have the child
3. The woman is not married and does not wish to marry the man
4. The couple cannot afford any more children.
5. There is a strong chance of a defect in the baby
6. The woman's health is seriously endangered by the pregnancy
7. The woman became pregnant as a result of rape

0.42, 0.52, 0.47, 0.53, 0.86, 0.94, 0.93

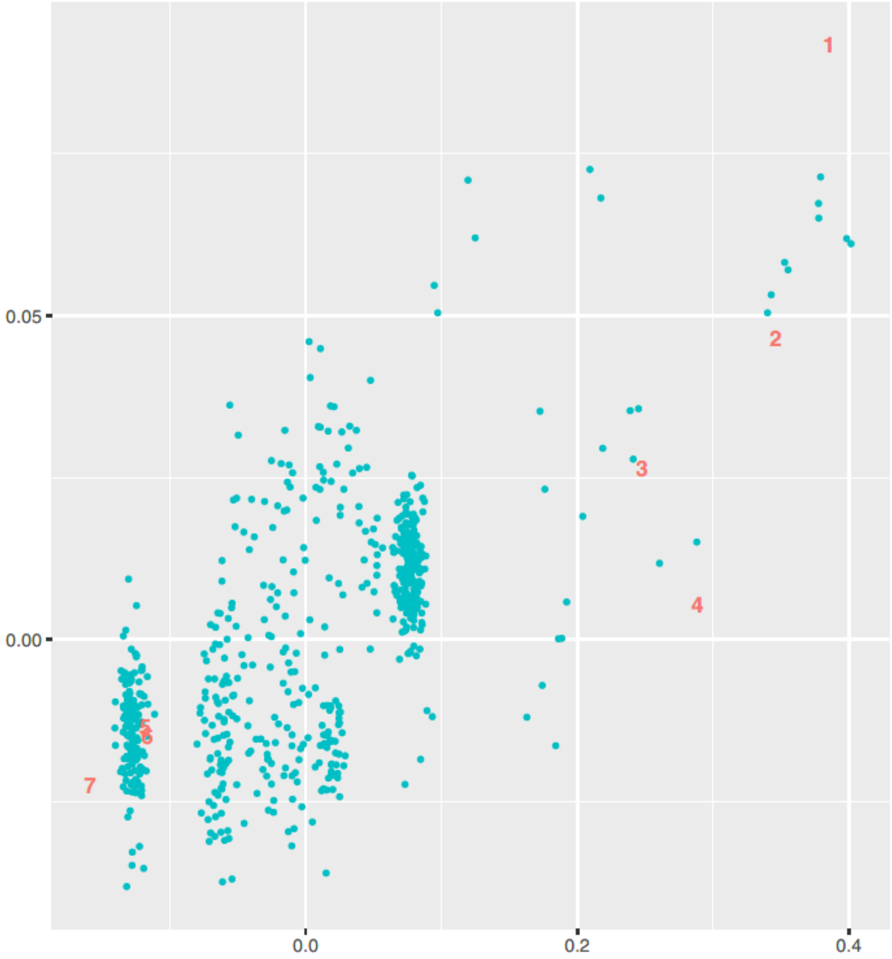
– 20,000 iterations (10,000 burn-in)

– Convergence: the scale reduction factor < 1.06

– Model selection: probability of $\delta = 1$ was 0.99

→ move forward with the latent space model

Example 1: Attitudes to Abortion



$\gamma = 1.25\ [0.92, 1.54]$

FIGURE 3.
Estimated latent space for the attitudes to abortion data. Red numbers represent items, and blue dots represent respondents

Example 1: Attitudes to Abortion

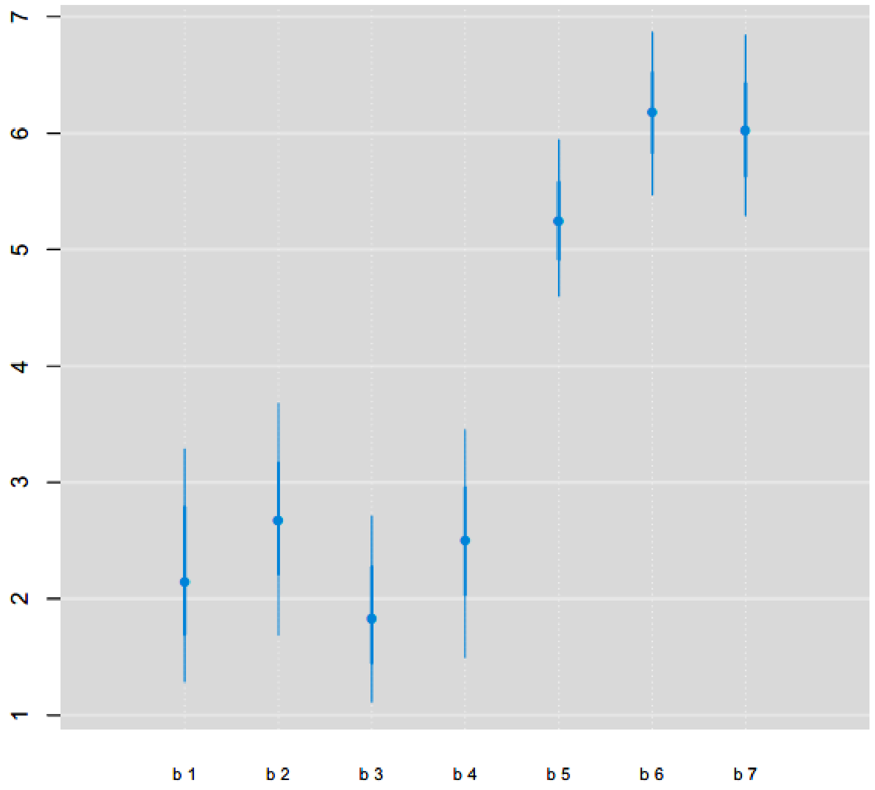
the region of $X > 0.3$ and $Y > 0.025$

ID	I1	I2	I3	I4	I5	I6	I7
27	1	0	0	1	0	0	0
92	1	1	1	1	1	0	0
132	1	1	0	1	0	0	0
191	1	1	1	1	0	1	0
273	1	1	0	0	0	0	0
330	1	0	0	1	0	0	0
653	1	1	0	1	0	0	0
662	1	1	1	0	0	0	0
675	1	1	1	0	0	0	0

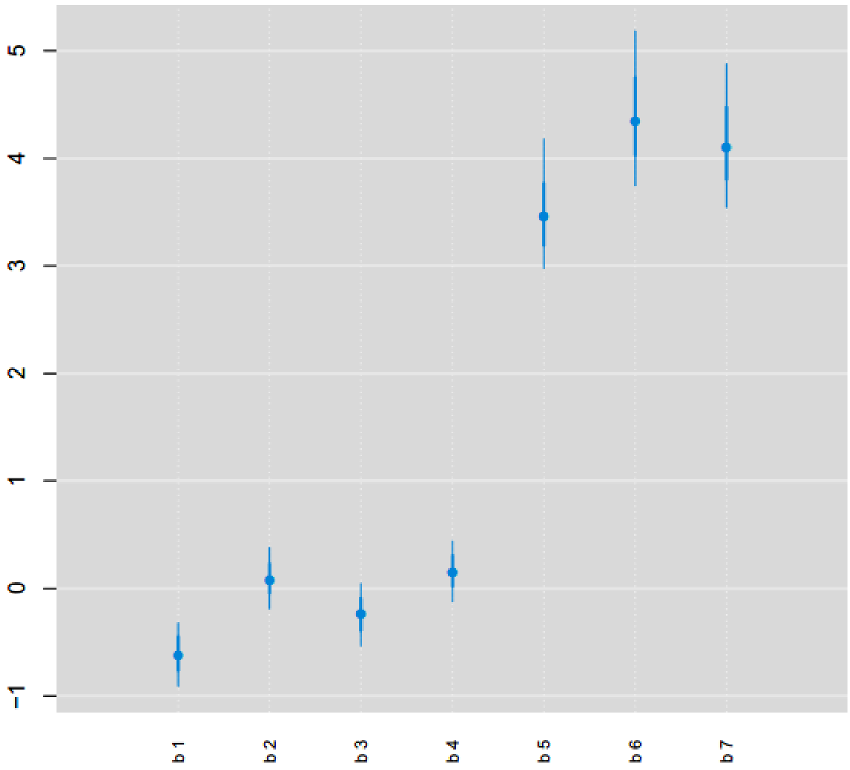
These people tend to give positive responses to I1–I4, but negative responses to I5–I7

Example 1: Attitudes to Abortion

(a) LS β_i

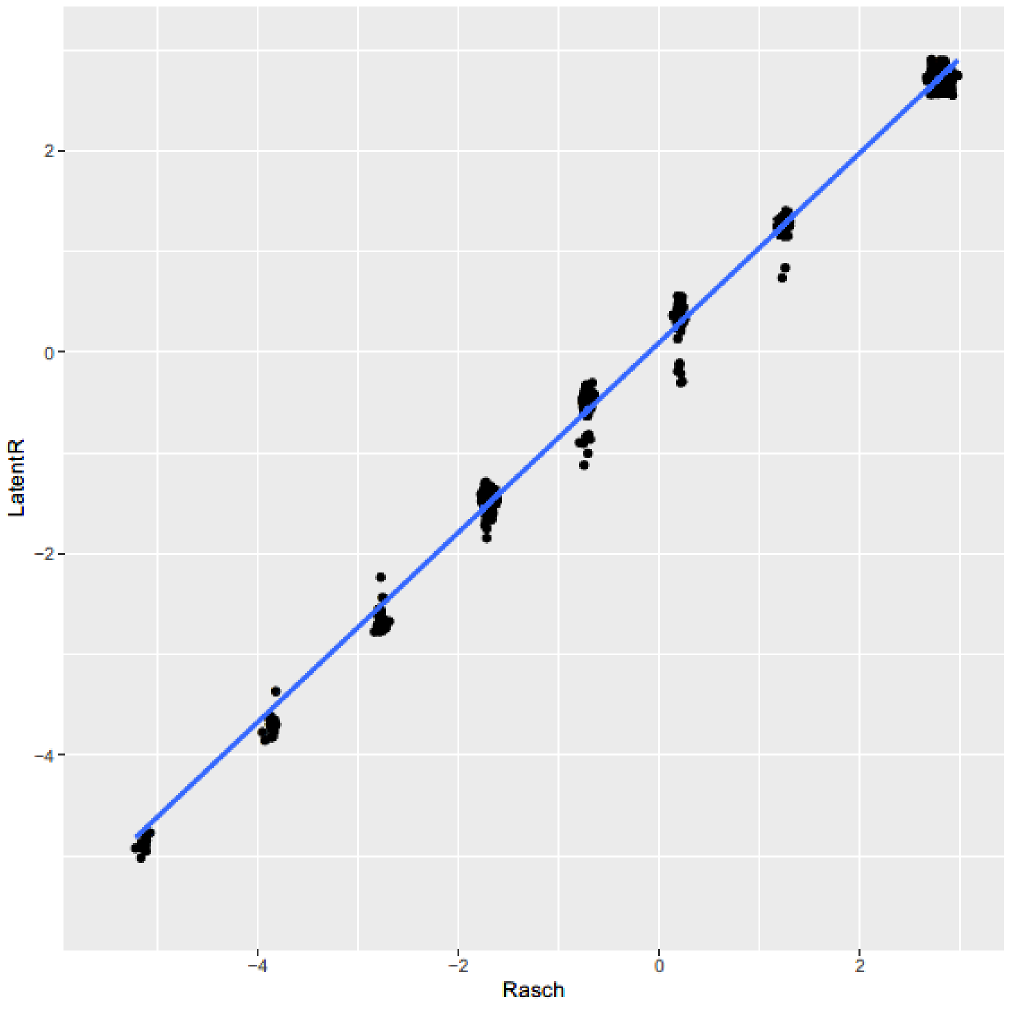


(b) Rasch β_i

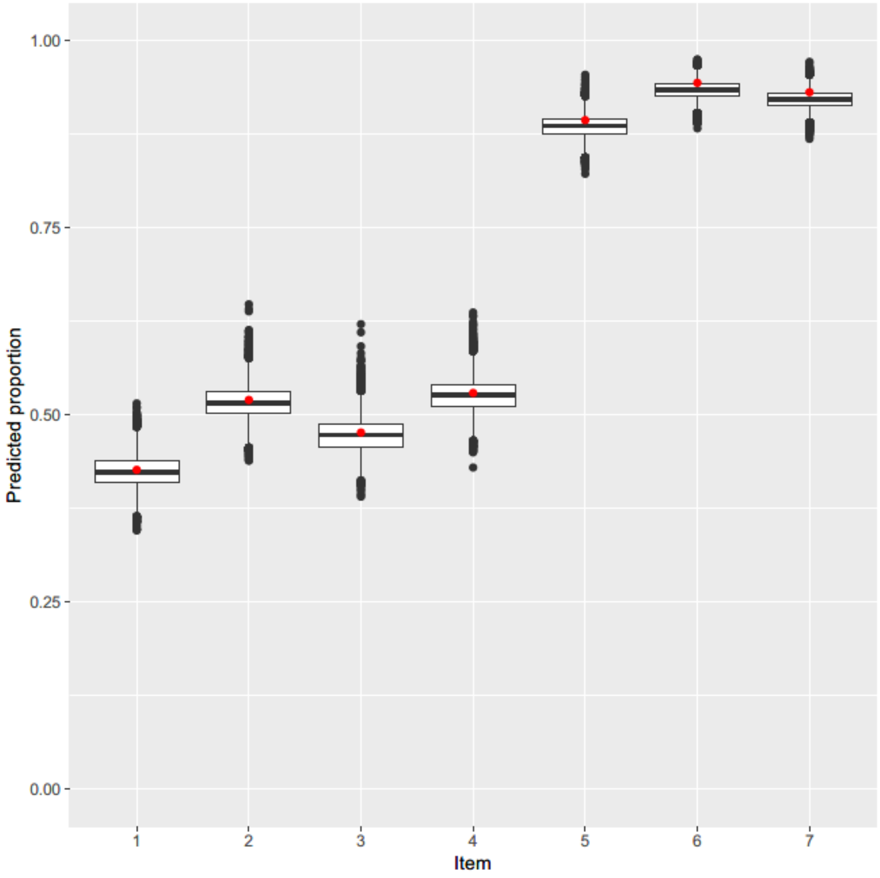


Example 1: Attitudes to Abortion

α_j



Example 1: Attitudes to Abortion



over 10,000 replicated data

Predicted proportions of the positive responses for the seven items for the attitudes to abortion data. The red dot in each box indicates the proportion of positive responses calculated from the raw data

Example 2: Deductive Reasoning

- Data and estimation
  - the Competence Profile Test of Deductive Reasoning—Verbal assessment (DRV)
  - 24 binary items (0 = correct, 1 = incorrect)
  - 418 school students (162 female)

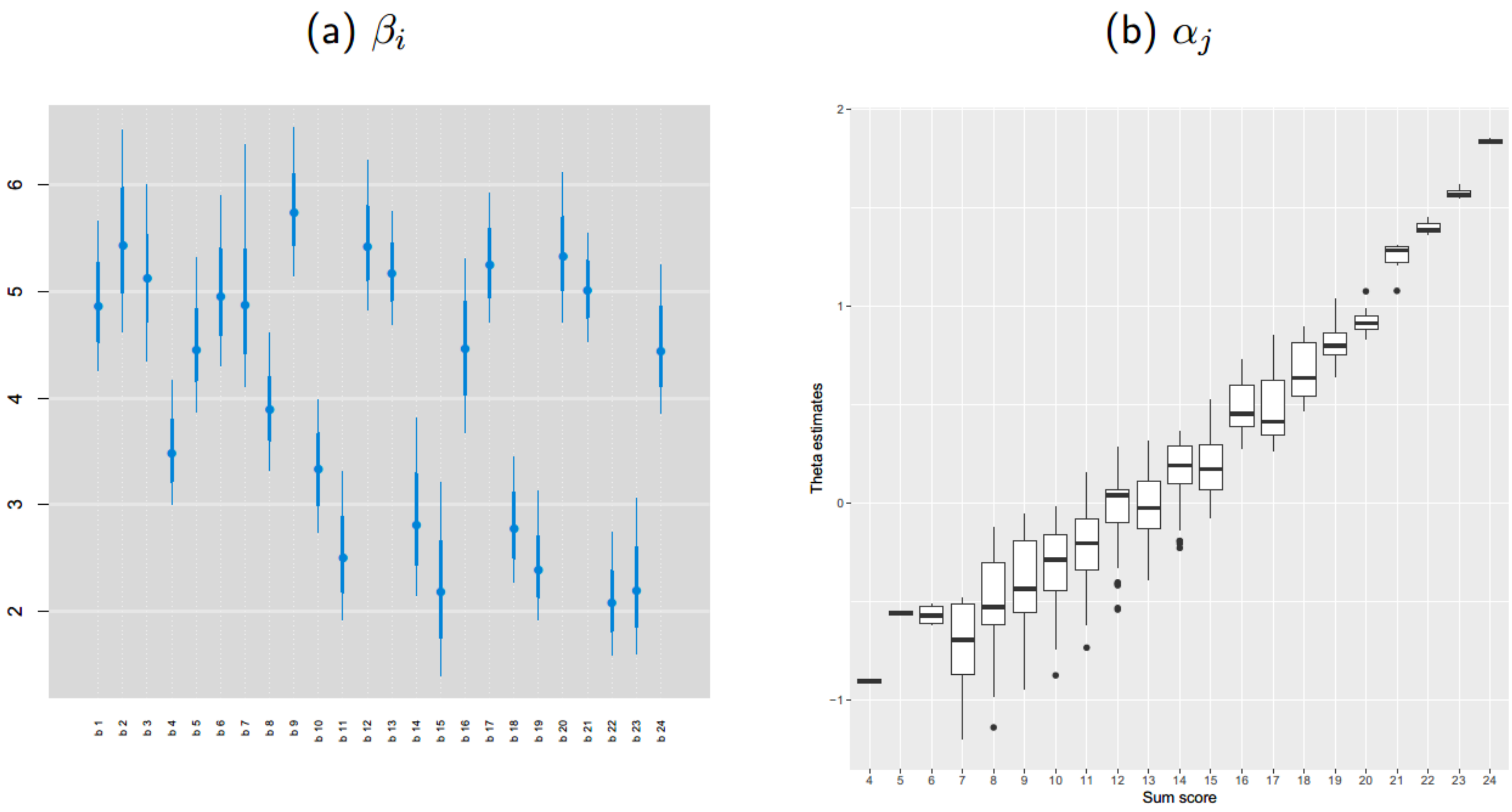
Precedent of antecedent	Content of conditional	Type of inference
No Negation (UN)	Concrete (CO)	Modus Ponens (MP)
Negation (N)	Abstract (AB)	Modus Tollens (MT)
	Counterfactual (CF)	Negation of Antecedent (NA)
		Affirmation of Consequence (AC)

Example 2: Deductive Reasoning

- Data and estimation
  - the Competence Profile Test of Deductive Reasoning—Verbal assessment (DRV)
  - 24 binary items (0 = correct, 1 = incorrect)
  - 418 school students (162 female)
  - 20,000 iterations (10,000 burn-in)
  - Convergence: the scale reduction factor < 1.1
  - Model selection: probability of $\delta = 1$ was 0.99

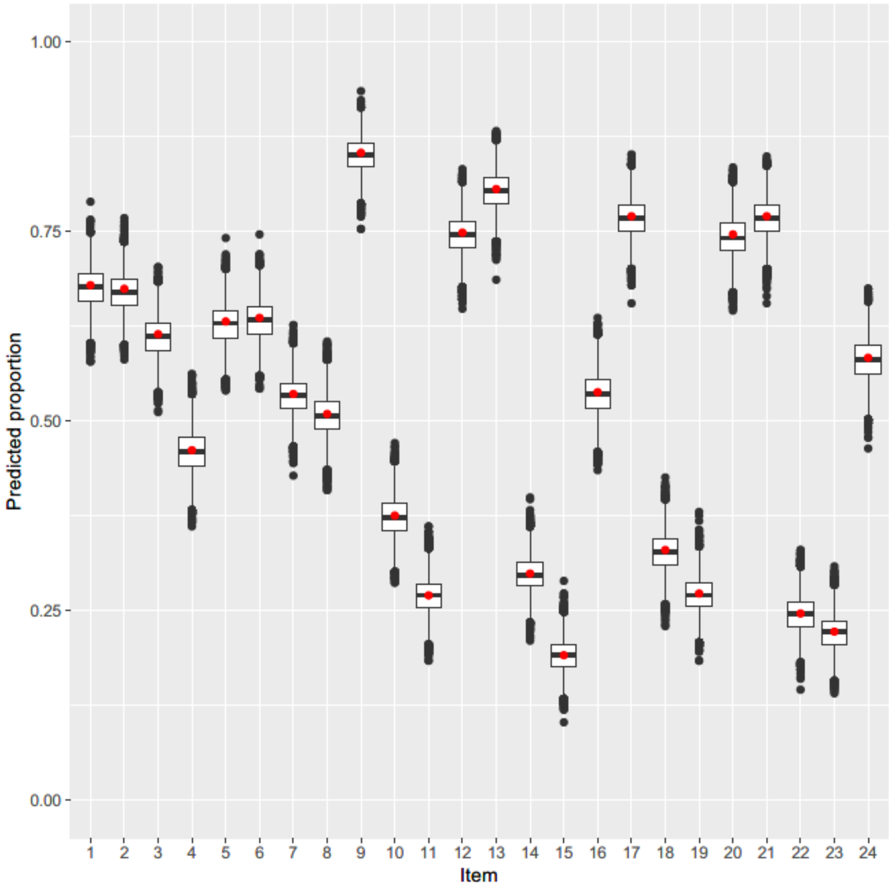
➡ move forward with the latent space model

Example 2: Deductive Reasoning



a 95% posterior credible intervals for the β_i estimates (b1 and b24 on the X-axis represent Items 1 to 24), and **b** the distribution of the α_j estimates per total test score for the DRV data. The estimates are from the latent space model

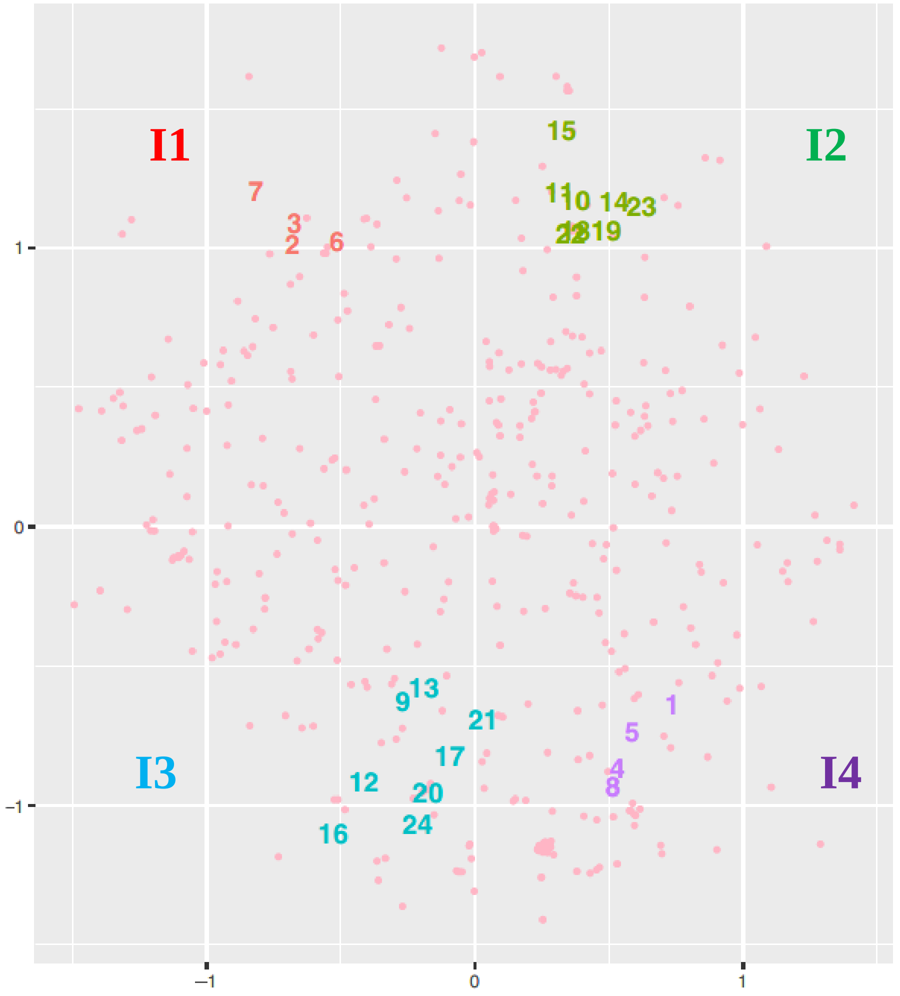
Example 2: Deductive Reasoning



Box plots of the predicted proportions of the correct responses for the 24 DRV test items from 10,000 replicated data. The red dot in each box indicates the proportion of the correct responses for the corresponding item from the raw data

Example 2: Deductive Reasoning

(a) DRV latent space



$\gamma = 2.23\ [2.08, 2.35]$

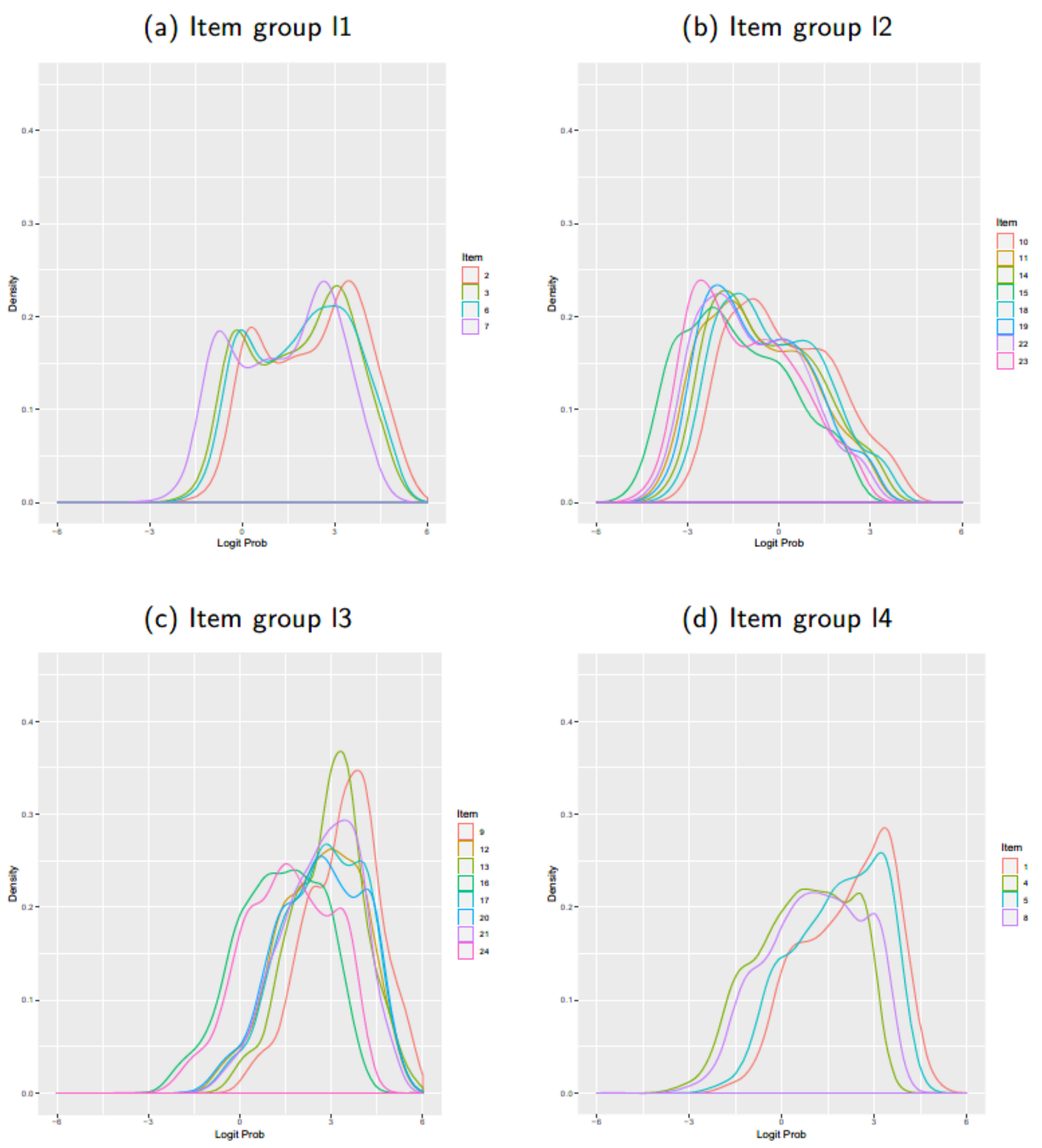
Example 2: Deductive Reasoning

Members of the four item groups identified in the DRV data latent space

Item group	Group details
I1	UN_CO_NA (2); UN_CO_AC (3); N_CO_NA (6); N_CO_AC (7)
I2	UN_AB_NA (10); UN_AB_AC (11); N_AB_NC (14); N_AB_AC (15); UN_CF_NA (18); UN_CF_AC (19); N_CF_NA (22); N_CF_MT (23)
I3	UN_AB_MP (9); UN_AB_MT (12); N_AB_MP (13);N_AB_MT (16); UN_CF_MP (17); UN_CF_MT (20); N_CF_MP (21); N_CF_MT (24)
I4	UN_CO_MP (1); UN_CO_MT (4); N_CO_MP (5); N_CO_MT (8)

Numbers in parenthesis indicate item numbers. The acronyms in the item labels indicate the following design factors and their levels: (1) UN vs. N: no negation (UN) and Negation (N) for the presentation of the antecedent factor. (2) CO vs. AB vs. AC: Concrete (CO), Abstract (AB), and Counterfactual (CF) for the content of conditional factor. (3) MP vs. MT vs. NA vs. AC: Modus Ponens (MP), Modus Tollens (MT), Negation of Antecedent (NA), and Affirmation of Consequent (AC) for the type of inference factor

Example 2: Deductive Reasoning

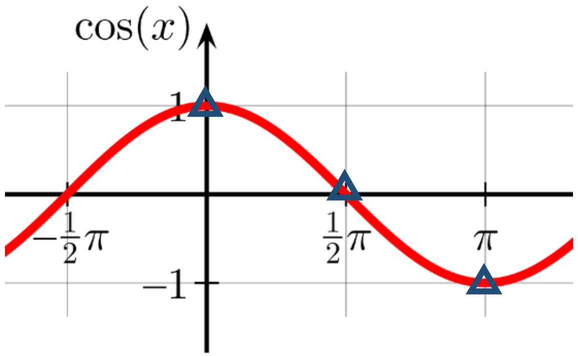


Example 2: Deductive Reasoning

- Cosine similarity between item groups

$$\cos(\theta) = \frac{\boldsymbol{a}^\top \boldsymbol{b}}{||\boldsymbol{a}||_2 \; ||\boldsymbol{b}||_2}$$

the angle between two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$

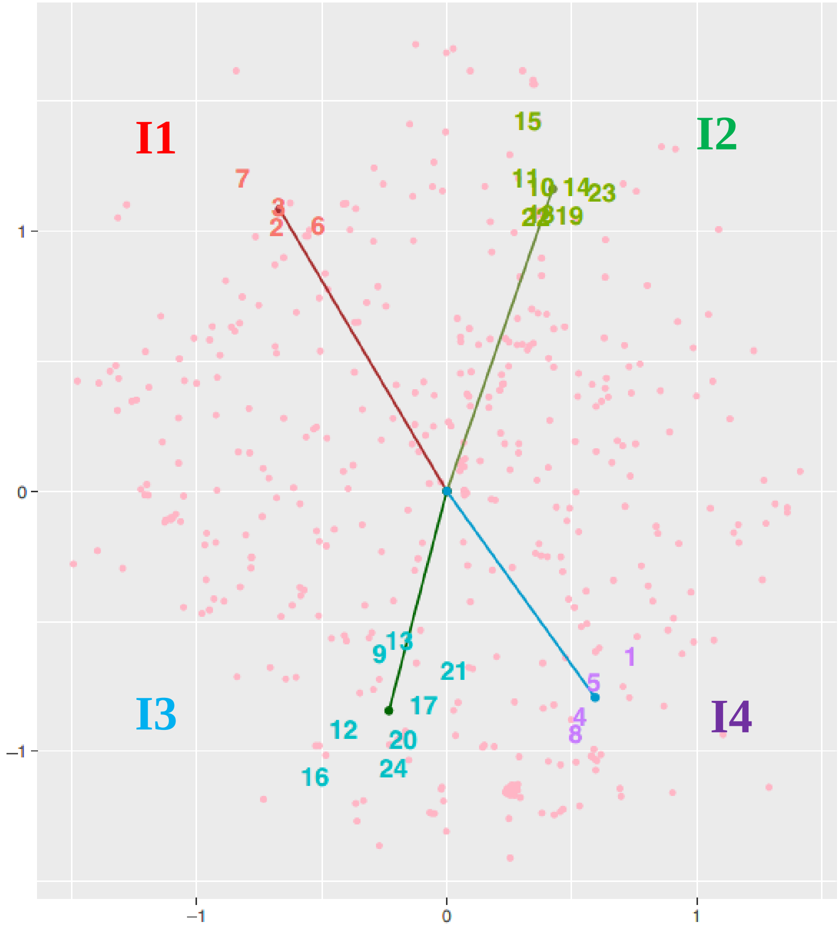


Example 2: Deductive Reasoning

Cosine similarity measures between (centers of) the four item groups

	I1	I2	I3	I4
I1	–			
I2	0.618	–		
I3	−0.680	−0.996	–	
I4	−0.996	−0.546	0.613	–

(b) with item group vectors



Example 2: Deductive Reasoning

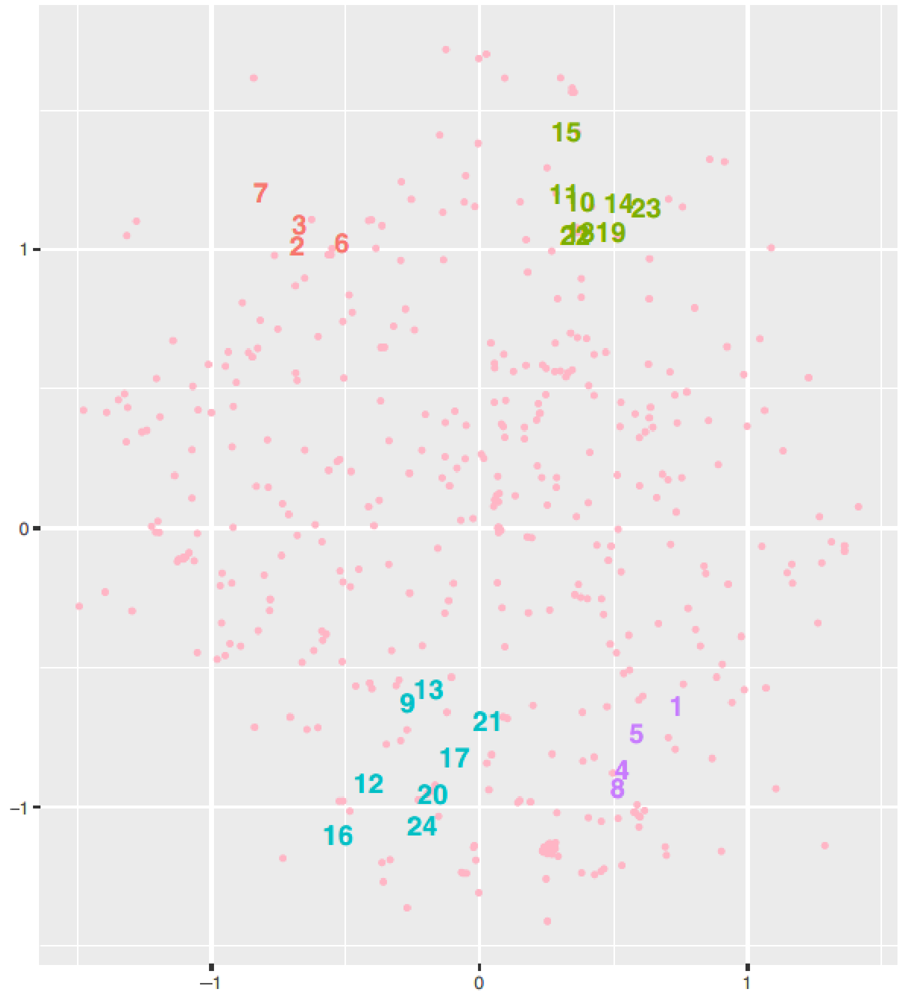
− Respondent structure

	logical fallacy inference items (NA/AC)	simpler inference items (MP/MT)	concrete conditionals (CO)	abstract or counterfactual conditionals (AB/CF)
Children near I1	√	X	X	
Children near I2	√	X		X
Children near I3	X	√		X
Children near I4	X	√	X	

Example 2: Deductive Reasoning

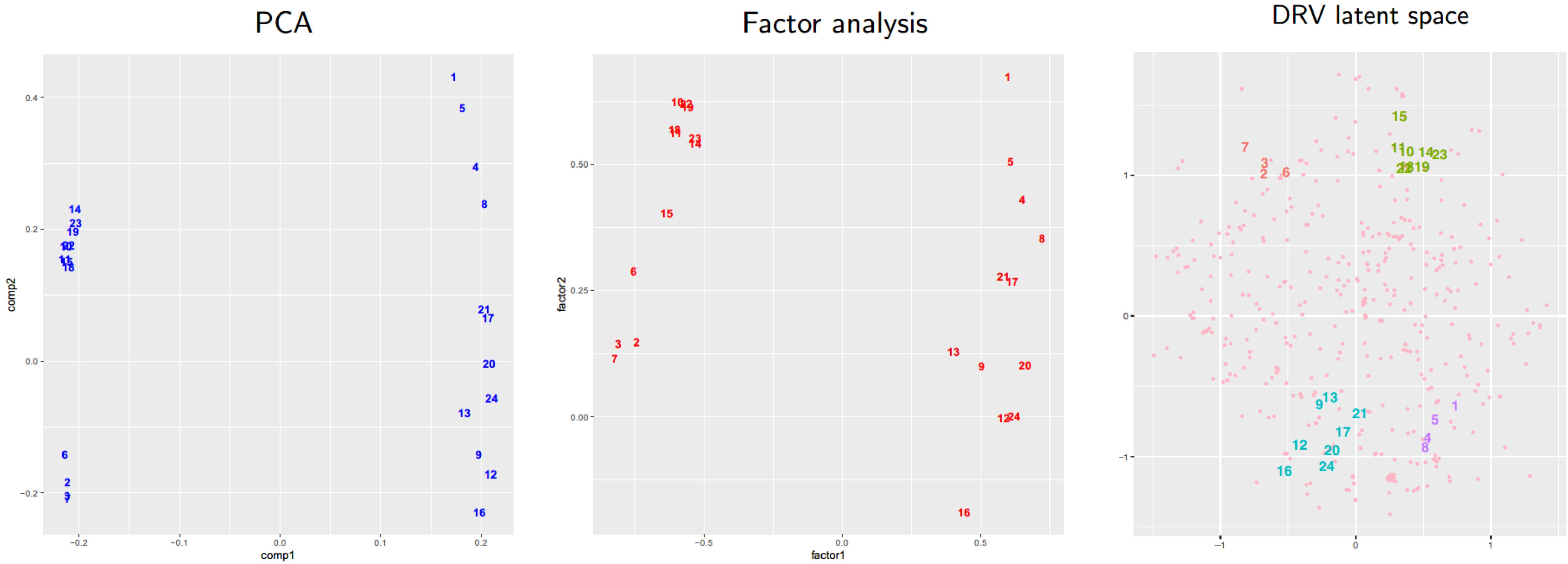
− Respondent structure

(a) DRV latent space

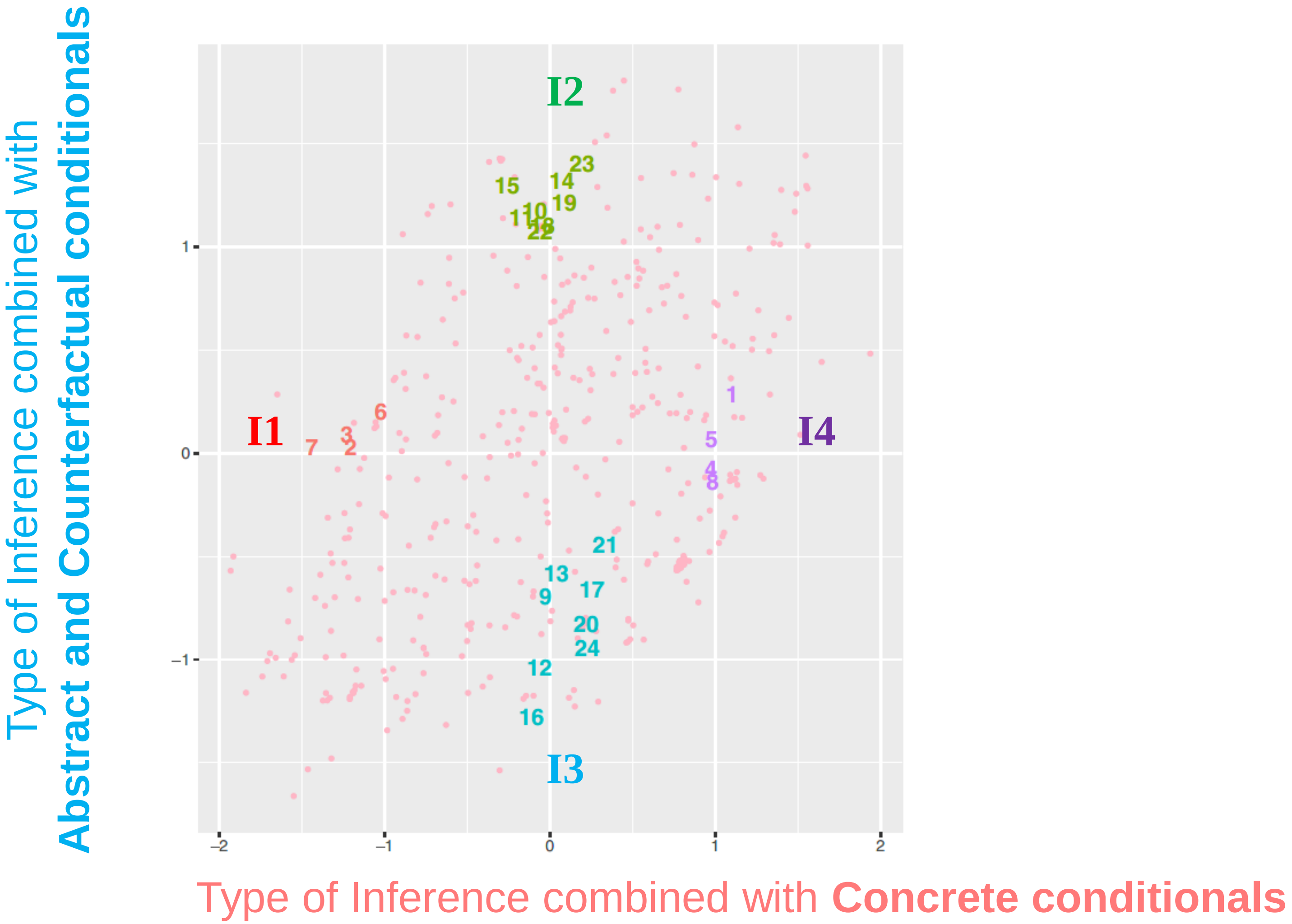


	logical fallacy inference item (NA/AC)	abstract or counterfactual onditionals (AB/CF)
Children near I1	√	
Children near I2	√	X
Children near I3	X	X
Children near I4	X	

Example 2: Deductive Reasoning



Example 2: Deductive Reasoning



Rotated latent space for the DRV data with oblim rotation. Dots represent respondents and numbers represent items. Four item groups are distinguished with four different colors. I1: Items 2, 3, 6, 7; I2: Items 10, 11, 14, 15, 18, 19, 22, 23; I3: Items 9, 12, 13, 16, 17, 20, 21, 24; I4: Items 1, 4, 5,8 (Color figure online)

Discussion

- interactions among respondents and items are present and non-negligible

- whether test items are differentiated or grouped together as blueprinted by test developers

  (e.g., the Presentation of Antecedent barely contributed to item differentiation)

- detect unintended or undesirable forms of test-taking behavior

  (e.g., respondents that are located close to the last test items)

- provide feedback on the test performance

  (e.g., identify items that individual test takers may be struggling with)

THANKS FOR ATTENTION!

REPORTER

YINGSHI HUANG