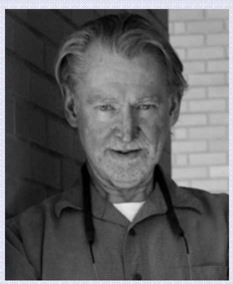
Bayesian Estimation of Multinomial Processing Tree Models with Heterogeneity in Participants and Items

Dora Matzke

Conor Dolan

William Batchelder

EJ Wagenmakers



UNIVERSITY OF AMSTERDAM

dog

paper

father

cat

pen

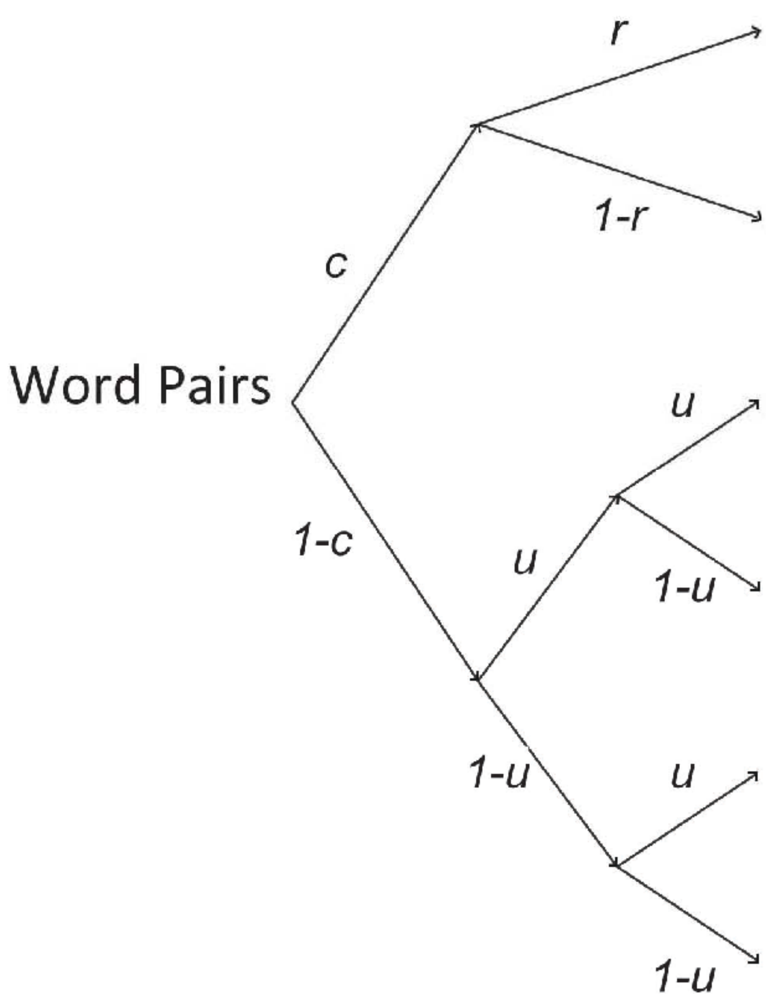
son

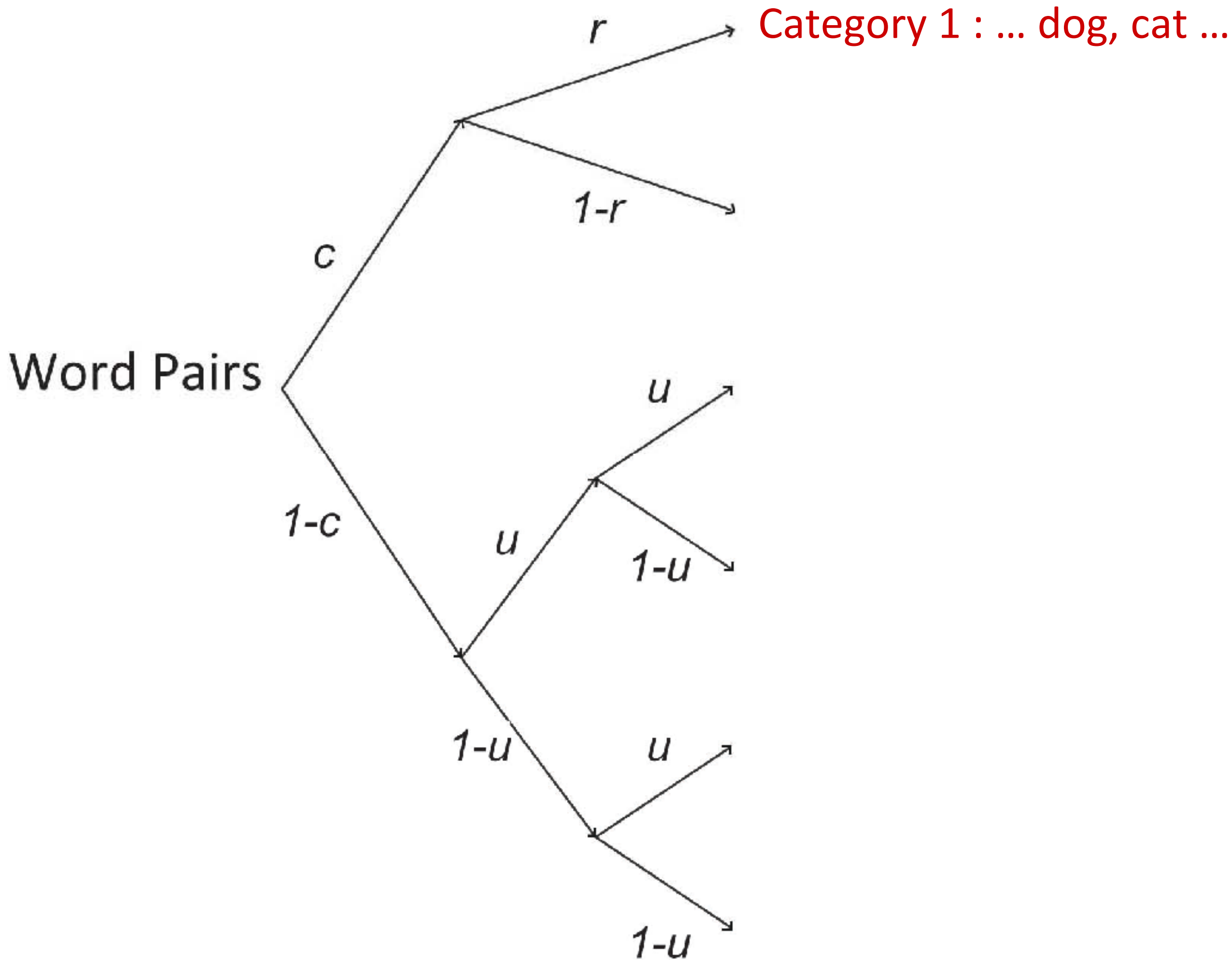
Pair-Clustering Paradigm
(Batchelder & Riefer, 1980, 1986)

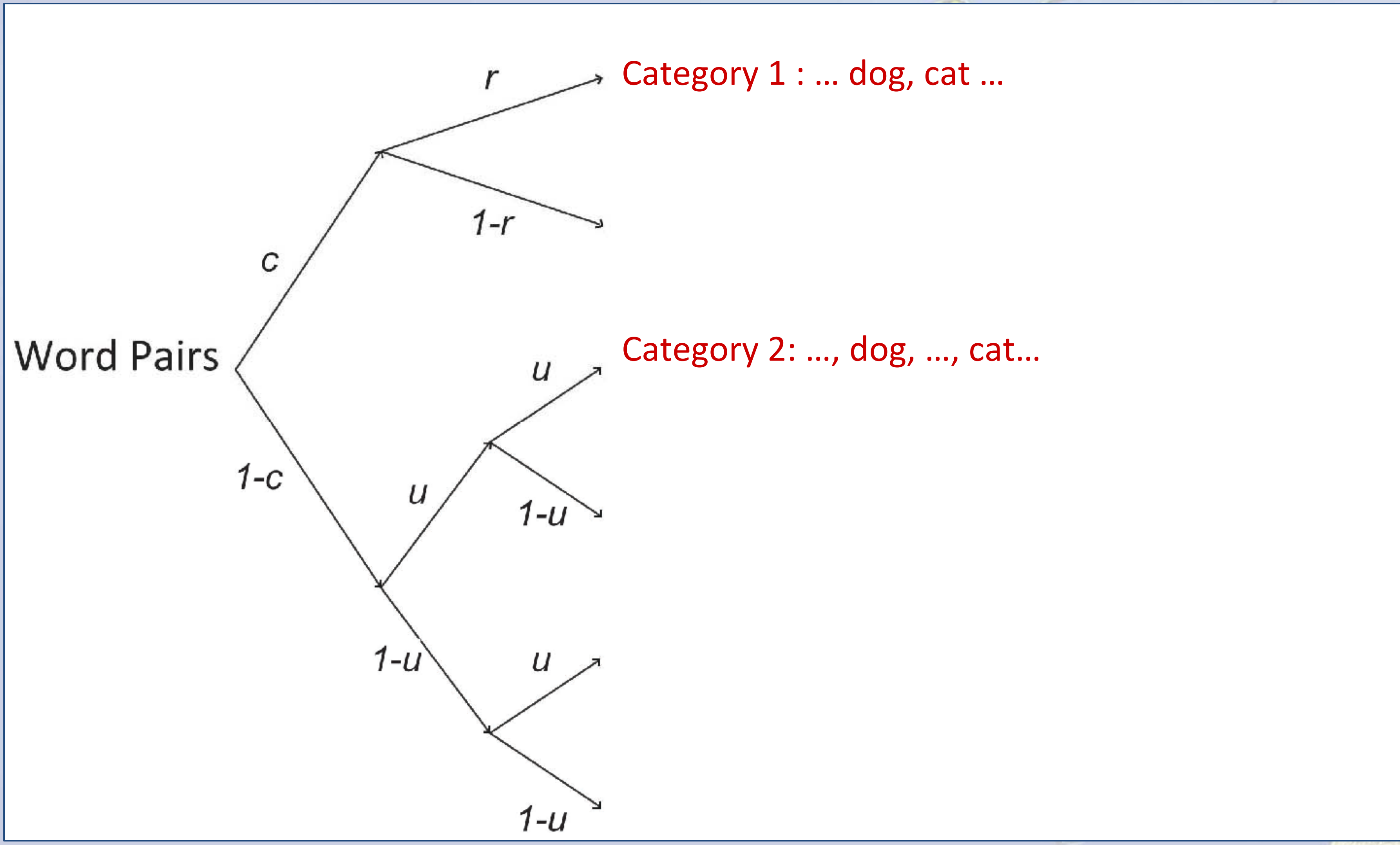
dog – cat

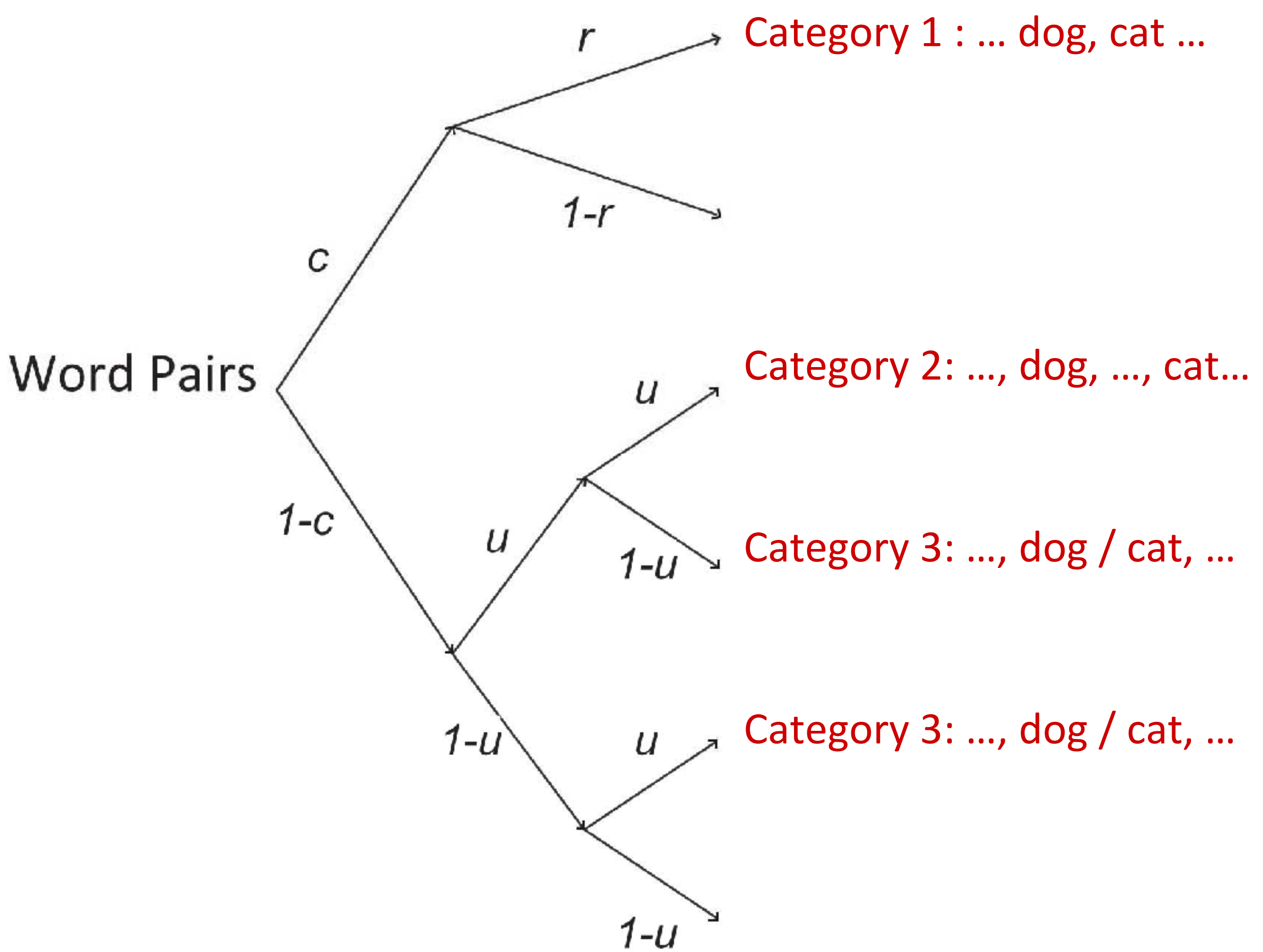
paper – pen

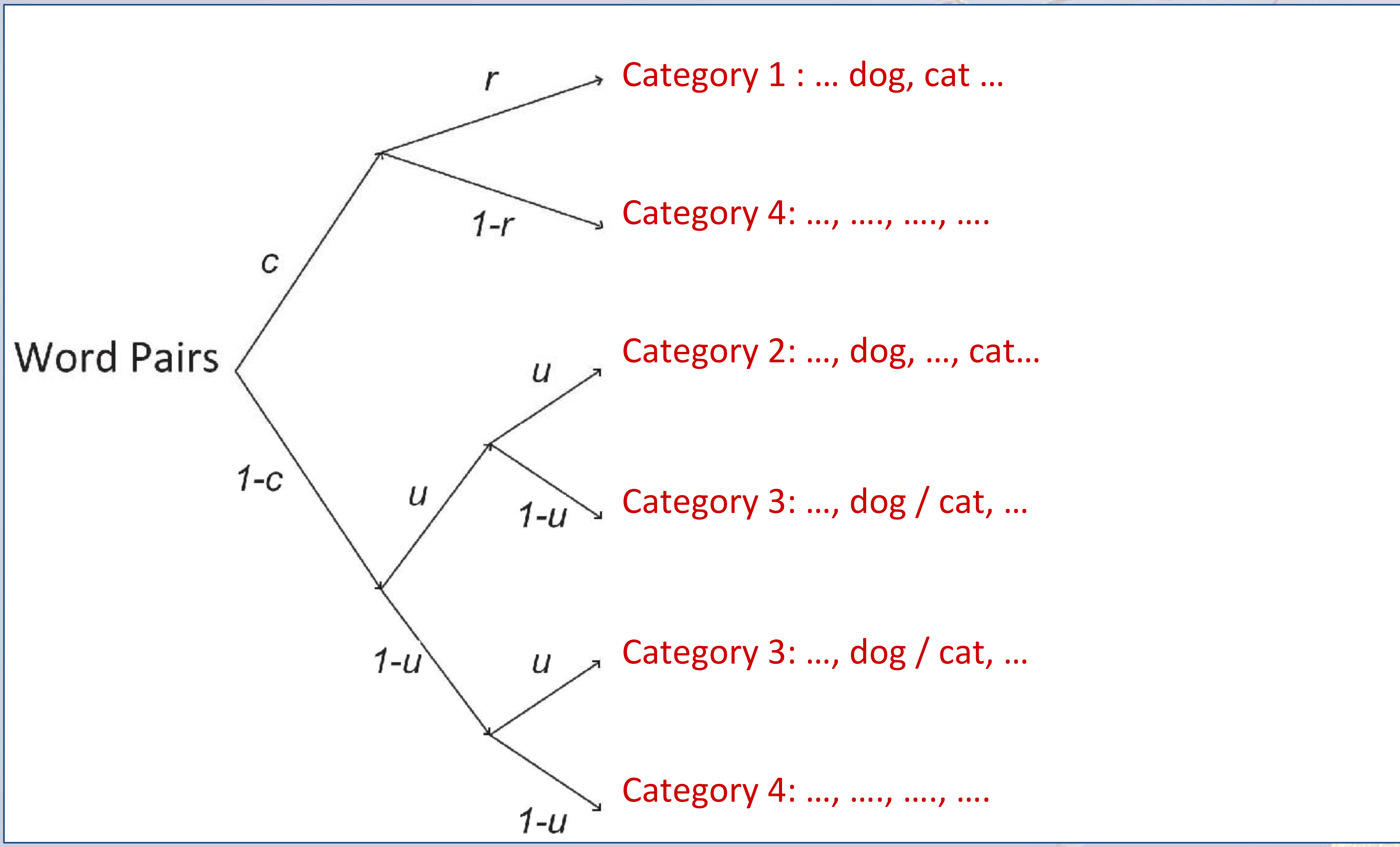
father – son

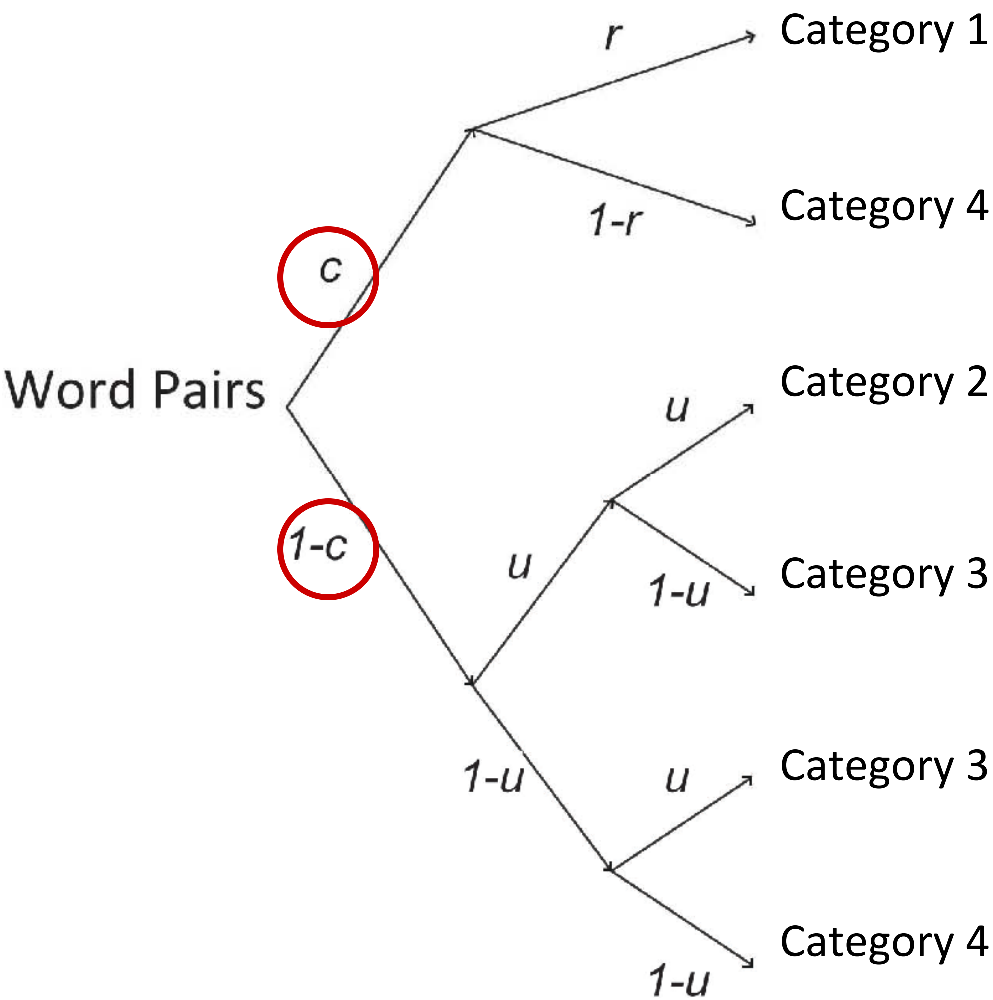

Word Pairs
c
1-c
r
1-r
u
1-u
u
1-u
u
1-u


Word Pairs
c
1-c
r
1-r
Category 1 : … dog, cat …
u
1-u
u
1-u
1-u
u
1-u

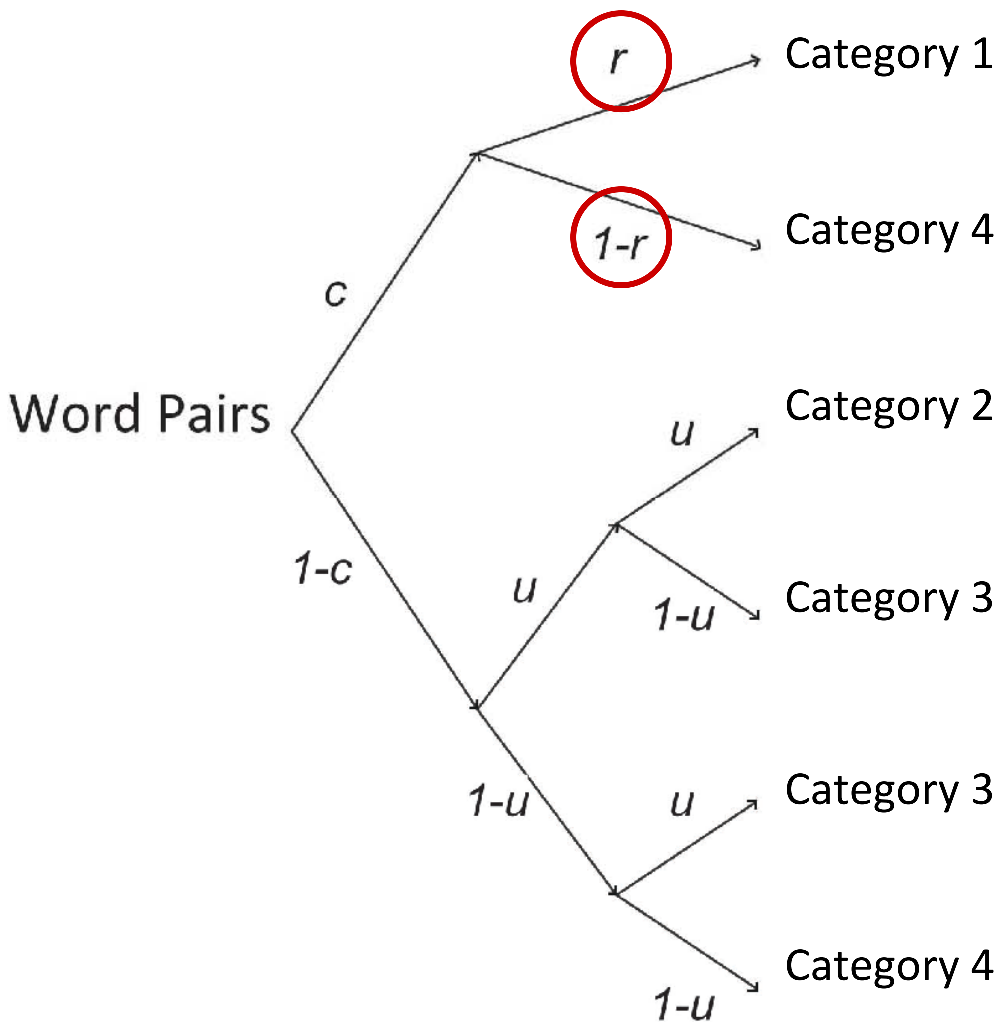

Word Pairs
c
1-c
r
1-r
u
1-u
u
1-u
u
1-u
Category 1 : … dog, cat …
Category 2: …, dog, …, cat…


Word Pairs
c
1-c
r
1-r
u
1-u
u
1-u
u
1-u
Category 1 : … dog, cat …
Category 2: …, dog, …, cat…
Category 3: …, dog / cat, …
Category 3: …, dog / cat, …


Word Pairs
c
1-c
r
1-r
u
1-u
u
1-u
u
1-u
Category 1 : … dog, cat …
Category 4: …, …., …., ….
Category 2: …, dog, …, cat…
Category 3: …, dog / cat, …
Category 3: …, dog / cat, …
Category 4: …, …., …., ….

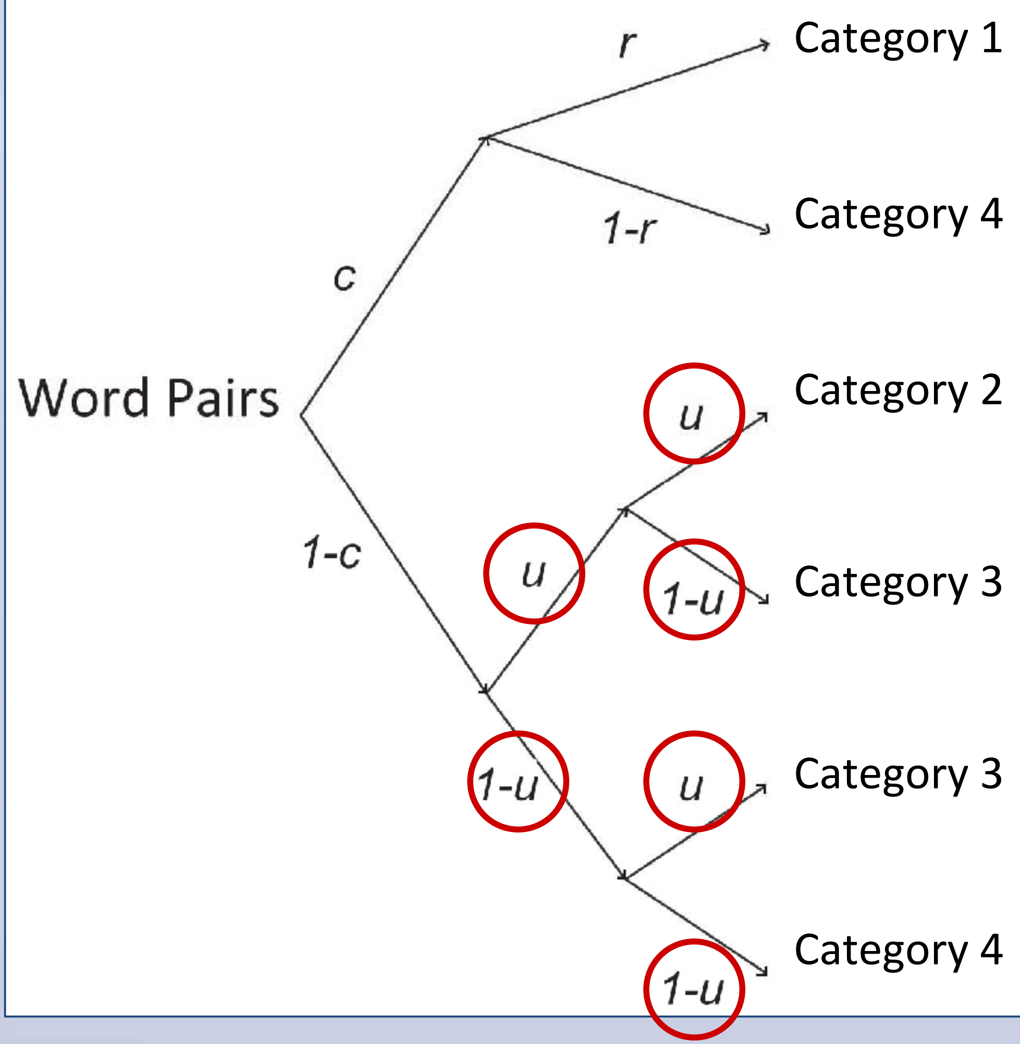


c : cluster-storage



c : cluster-storage

r: cluster-retrieval



c : cluster-storage

r: cluster-retrieval

u: storage-retrieval

Parameter heterogeneity

- Traditional aggregate analysis biased if parameters are heterogeneous
- Incorporate participant heterogeneity in MPT models
  - Latent-trait approach (Klauer, 2010)
- Crossed-random effects extension
  - Participant and item heterogeneity

Outline

- Latent-trait approach (Klauer, 2010)
- Crossed-random effects extension
- Fitting experimental free recall data

Outline

- Latent-trait approach (Klauer, 2010)
- Crossed-random effects extension
- Fitting experimental free recall data

Latent-trait approach (Klauer, 2010)

- Parameter heterogeneity as a result of individual differences between participants
- Data aggregated only over word pairs
- (c_i, r_i, u_i) for $i=1,\ldots,I$.

Latent-trait approach (Klauer, 2010)

- Hierarchical Bayesian modeling (e.g., Gelman & Hill, 2007)
- Probit transformed individual parameters are multivariate normal distributed
  - Between subject variability
  - Parameter correlations

Outline

- Latent-trait approach (Klauer, 2010)
- Crossed-random effects extension
- Fitting experimental free recall data

Crossed-random effects approach

- Parameter heterogeneity as a result of differences between participants and items
- Raw participant-by-item data
- $\Theta_{ik} = (c_{ik}, r_{ik}, u_{ik})$ for i=1,…,I, and k=1,…,K.

Crossed-random effects approach

- Reduce the number of parameters
- Participant effects + item effects (e.g, Rouder et al., 2007, 2008)

Data

for i=1,…,I, and k=1,…,K

- $Data_{ik}$ ~ Multinomial($Pr(C_1)_{ik}$,…,$Pr(C_4)_{ik}$)

Data

for i=1,…,I, and k=1,…,K

- $Data_{ik} \sim Multinomial(Pr(C_1)_{ik},\ldots,Pr(C_4)_{ik})$

- $Pr(C_1)_{ik} = c_{ik} \times r_{ik}$

  $Pr(C_2)_{ik} = (1-c_{ik}) \times u_{ik}^2$

  $Pr(C_3)_{ik} = (1-c_{ik}) \times 2 \times u_{ik} \times (1-u_{ik})$

  $Pr(C_4)_{ik} = c_{ik} \times (1-r_{ik}) + (1-c_{ik}) \times (1-u_{ik})^2$

Probit transformation

for i=1,…,I, and k=1,…,K

- $c_{ij} = \phi(c_{ij}^{prb})$
  $r_{ij} = \phi(r_{ij}^{prb})$
  $u_{ij} = \phi(u_{ij}^{prb})$

Additivity

for i=1,…,I, and k=1,…,K

- $c_{ik}^{prb} = \mu_c + \delta_{ci}^{part} + \delta_{ck}^{item}$

  $r_{ik}^{prb} = \mu_r + \delta_{ri}^{part} + \delta_{rk}^{item}$

  $u_{ik}^{prb} = \mu_u + \delta_{ui}^{part} + \delta_{uk}^{item}$

Prior distributions

$\mu + \delta^{part} + \delta^{item}$

- $\mu_c \sim \text{Normal}(0,1)$
  $\mu_r \sim \text{Normal}(0,1)$
  $\mu_u \sim \text{Normal}(0,1)$

Prior distributions

$$\mu + \delta_i^{part} + \delta_k^{item}$$

for i=1,…,I

- $(\delta_{ci}^{part}, \delta_{ri}^{part}, \delta_{ui}^{part}) \sim \text{MNormal}((0,0,0), S)$

Prior distributions

$\mu + \delta_i^{part} + \delta_k^{item}$

for i=1,…,I

- $(\delta_{ci}^{part}, \delta_{ri}^{part}, \delta_{ui}^{part}) \sim MNormal((0,0,0), S)$
- $S \sim$ Scaled - Inverse-Wishart(I, df = 4, $\boldsymbol{\xi}^{part}$)
  - $\xi_c^{part} \sim$ Uniform(0,100)
  - $\xi_r^{part} \sim$ Uniform(0,100)
  - $\xi_u^{part} \sim$ Uniform(0,100)

Prior distributions

$\mu + \delta_i^{part} + \delta_k^{item}$

for k=1,…,K

- $\delta_{ck}^{part} \sim \text{Normal}(0, \sigma_c^2)$
- $\delta_{rk}^{part} \sim \text{Normal}(0, \sigma_r^2)$
- $\delta_{uk}^{part} \sim \text{Normal}(0, \sigma_u^2)$

Prior distributions

$\mu + \delta_i^{part} + \delta_k^{item}$

- $\sigma_c^2 \sim \text{Inverse} - \text{Gamma}\ (1,1, \xi_c^{item})$
  $\sigma_r^2 \sim \text{Inverse} - \text{Gamma}\ (1,1, \xi_r^{item})$
  $\sigma_u^2 \sim \text{Inverse} - \text{Gamma}\ (1,1, \xi_u^{item})$
  - $\xi_c^{item} \sim \text{Uniform}(0,100)$
  - $\xi_r^{item} \sim \text{Uniform}(0,100)$
  - $\xi_u^{item} \sim \text{Uniform}(0,100)$

Implementation

- MCMC sampling
- Implemented in WinBUGS
  - Easy-to-use

Outline

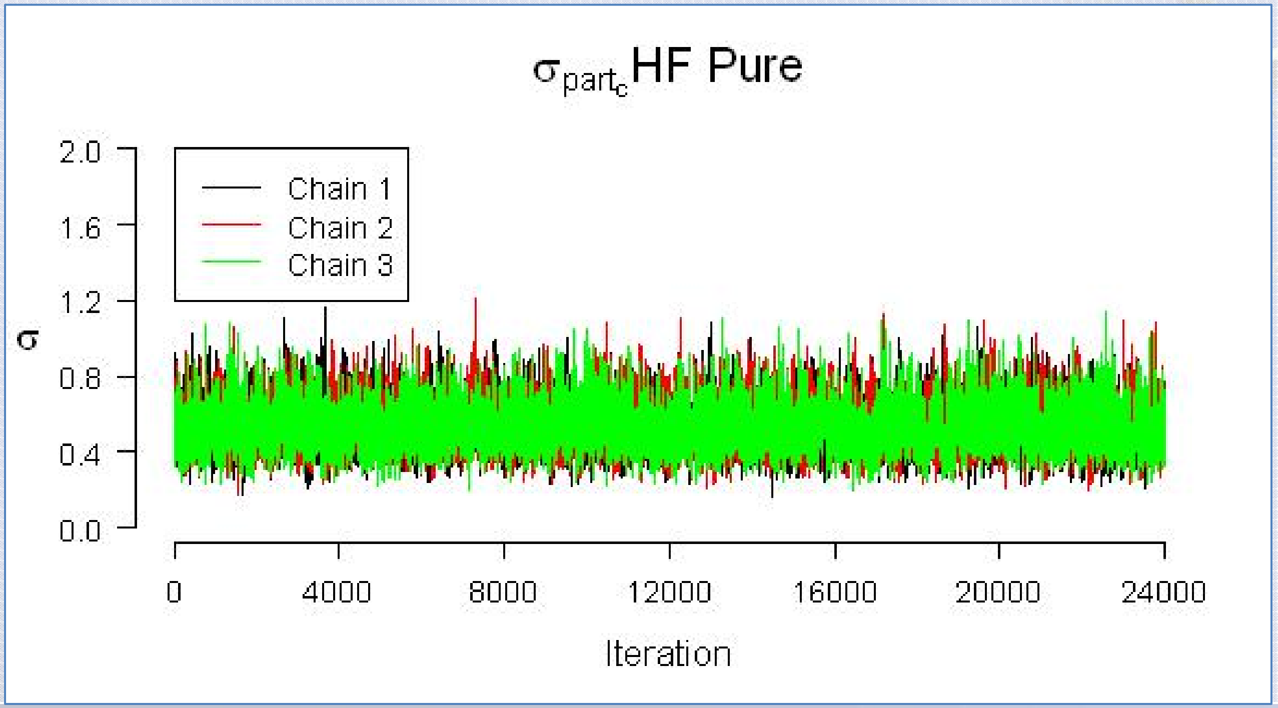
- Latent-trait approach (Klauer, 2010)
- Crossed-random effects extension
- Fitting experimental free recall data

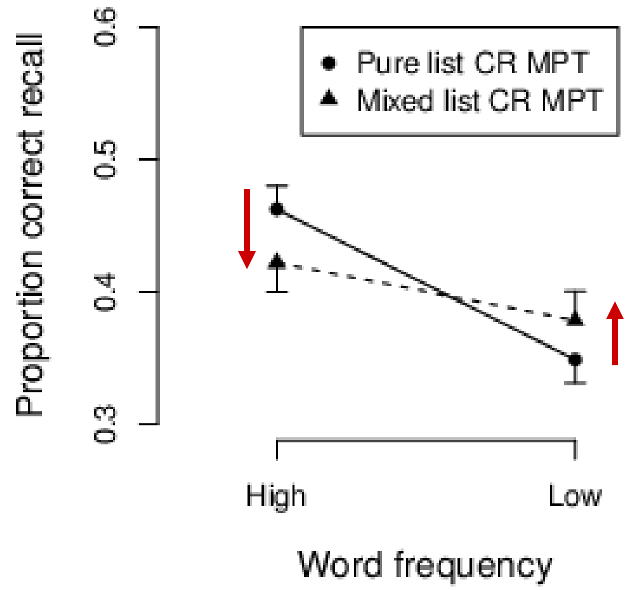
Fitting Real Data

- Pure list vs. mixed list word frequency paradox
  - HF Pure > LF Pure
  - HF Mixed >= LF Mixed
- Orthographically related word pairs
- 65 participants and 4 list conditions
  - HF Pure
  - LF Pure
  - HF Mixed
  - LF Mixed

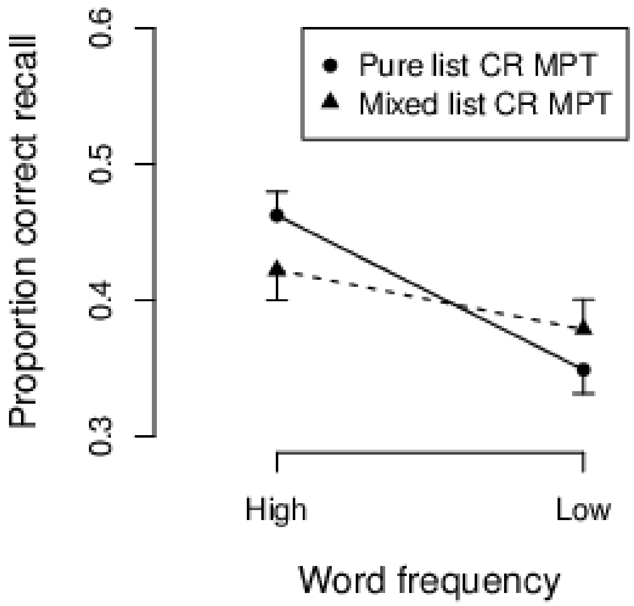
Fitting Real Data

- Convergence: OK
- Posterior predictive model checks: OK

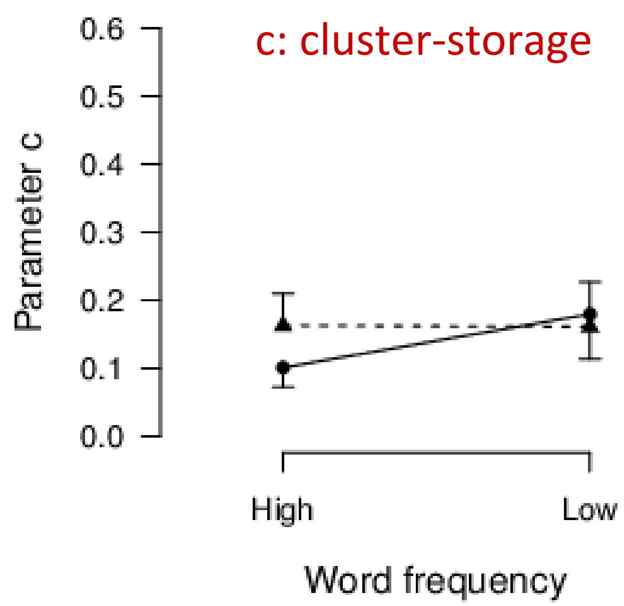




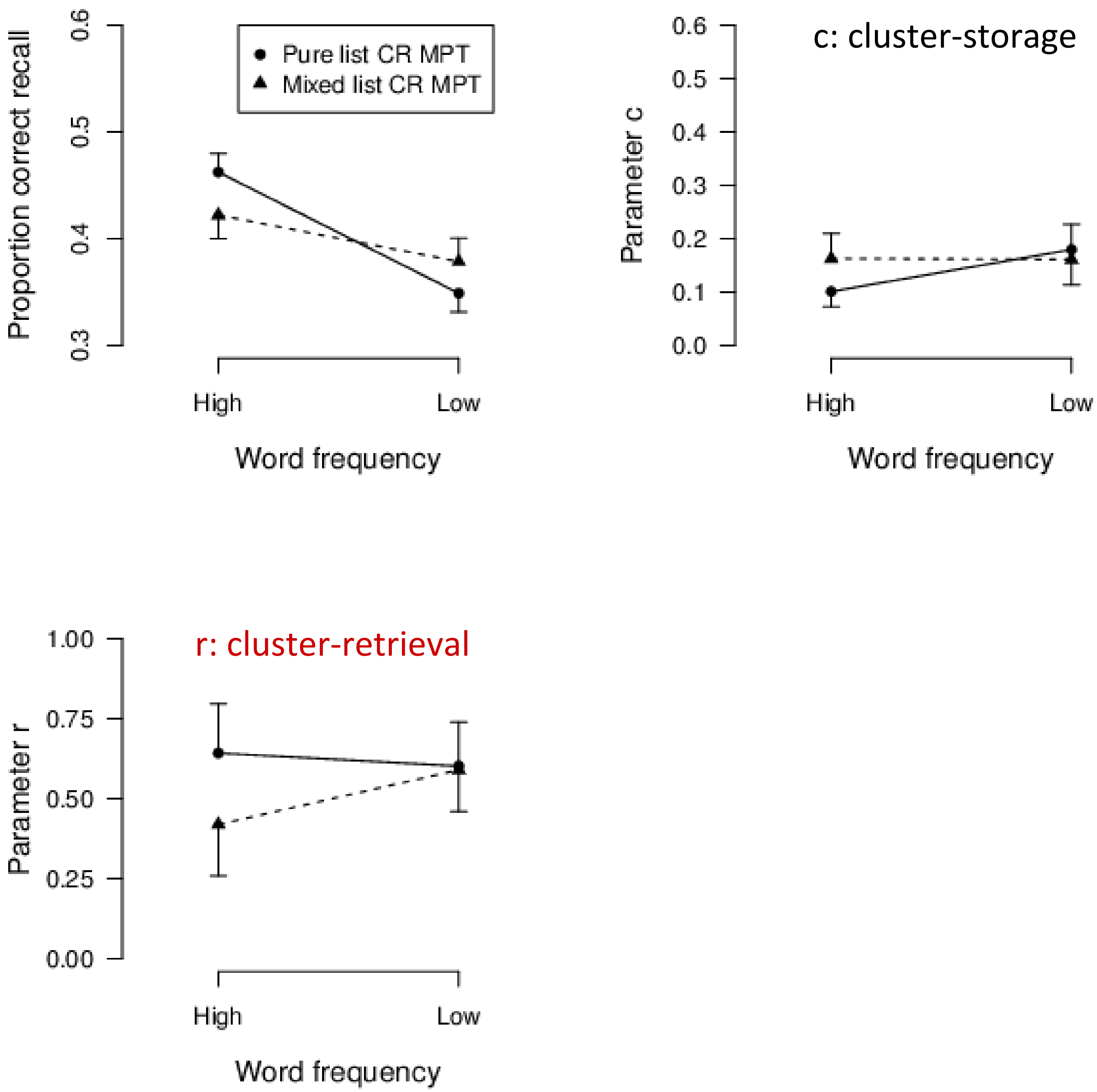
Pure list CR MPT
Mixed list CR MPT
Proportion correct recall
0.3
0.4
0.5
0.6
High
Low
Word frequency



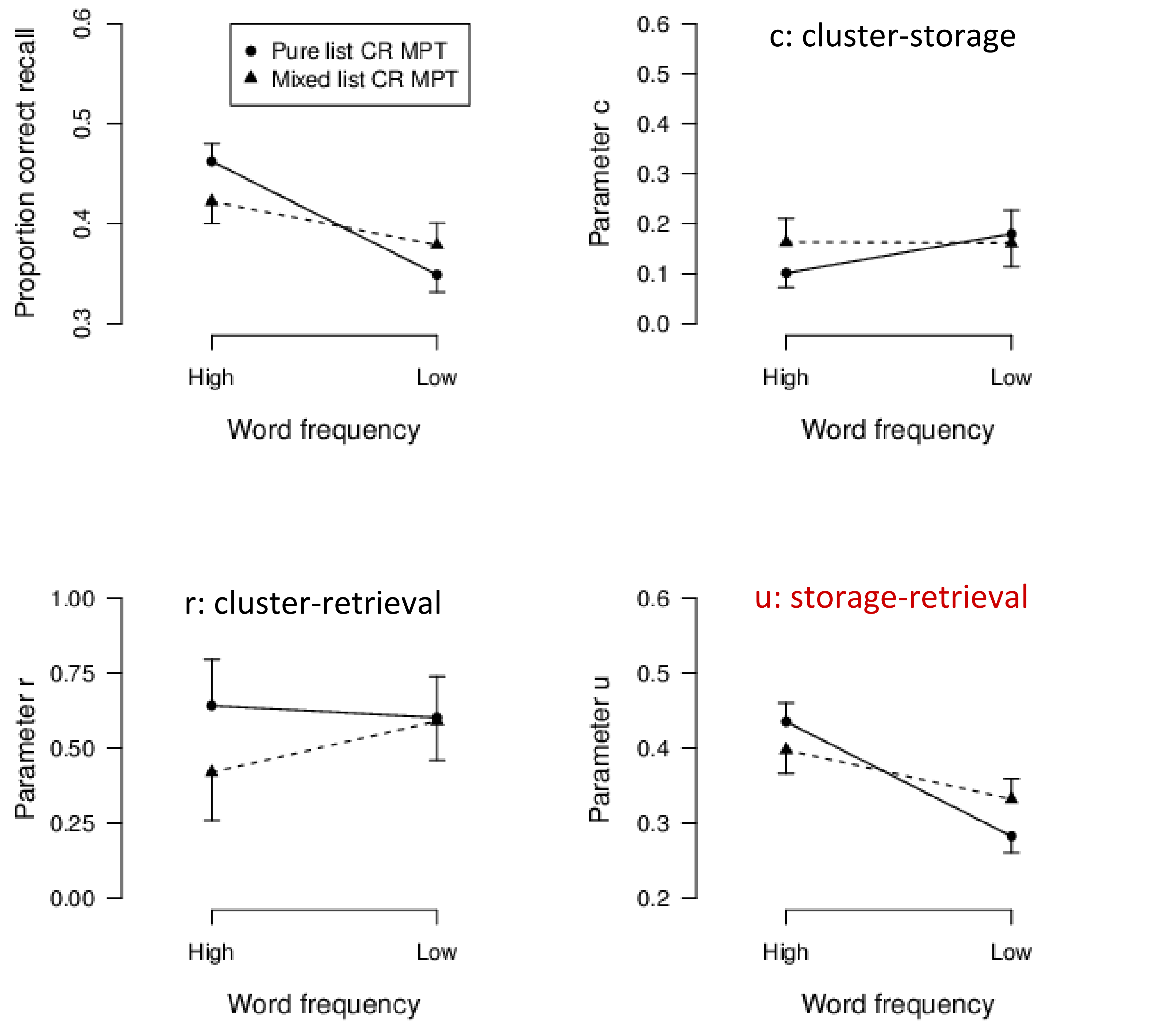
Pure list CR MPT
Mixed list CR MPT
Proportion correct recall
0.6
0.5
0.4
0.3
High
Low
Word frequency



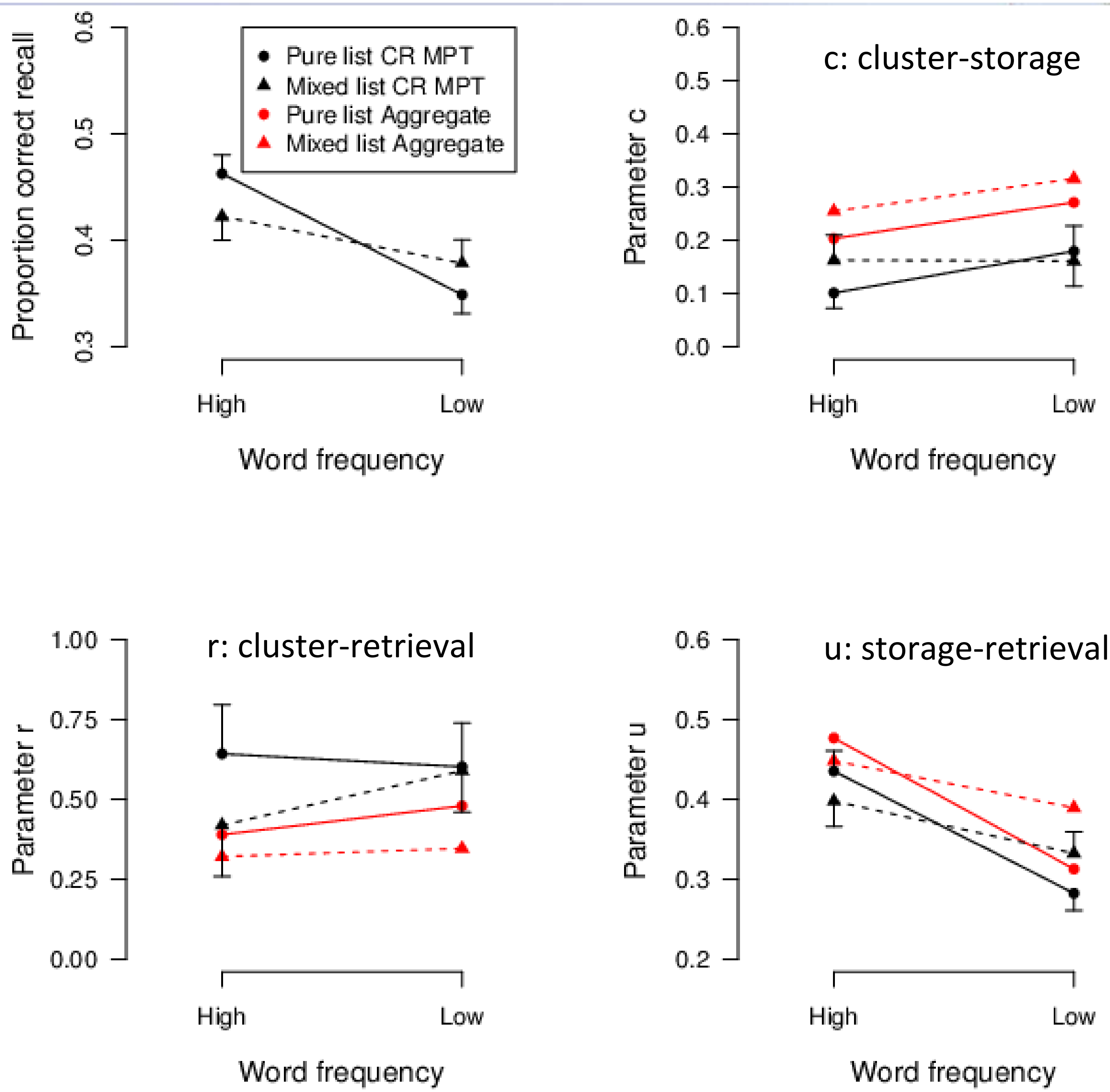
c: cluster-storage
Parameter c
0.6
0.5
0.4
0.3
0.2
0.1
0.0
High
Low
Word frequency



Proportion correct recall
0.3
0.4
0.5
0.6
Pure list CR MPT
Mixed list CR MPT
High
Low
Word frequency
c: cluster-storage
Parameter c
0.0
0.1
0.2
0.3
0.4
0.5
0.6
High
Low
Word frequency
r: cluster-retrieval
Parameter r
0.00
0.25
0.50
0.75
1.00
High
Low
Word frequency



Pure list CR MPT
Mixed list CR MPT
Proportion correct recall
0.3
0.4
0.5
0.6
High
Low
Word frequency
c: cluster-storage
Parameter c
0.0
0.1
0.2
0.3
0.4
0.5
0.6
High
Low
Word frequency
r: cluster-retrieval
Parameter r
0.00
0.25
0.50
0.75
1.00
High
Low
Word frequency
u: storage-retrieval
Parameter u
0.2
0.3
0.4
0.5
0.6
High
Low
Word frequency



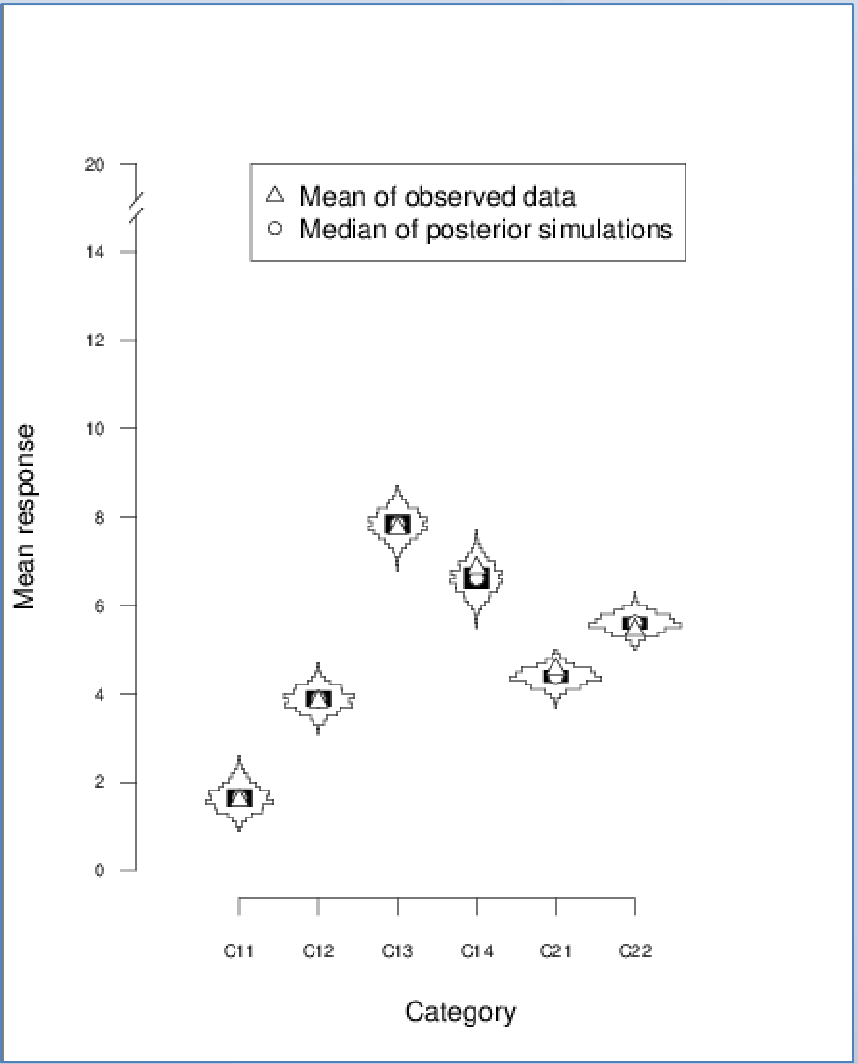
Proportion correct recall
0.3
0.4
0.5
0.6
Pure list CR MPT
Mixed list CR MPT
Pure list Aggregate
Mixed list Aggregate
High
Low
Word frequency
c: cluster-storage
Parameter c
0.0
0.1
0.2
0.3
0.4
0.5
0.6
High
Low
Word frequency
r: cluster-retrieval
Parameter r
0.00
0.25
0.50
0.75
1.00
High
Low
Word frequency
u: storage-retrieval
Parameter u
0.2
0.3
0.4
0.5
0.6
High
Low
Word frequency

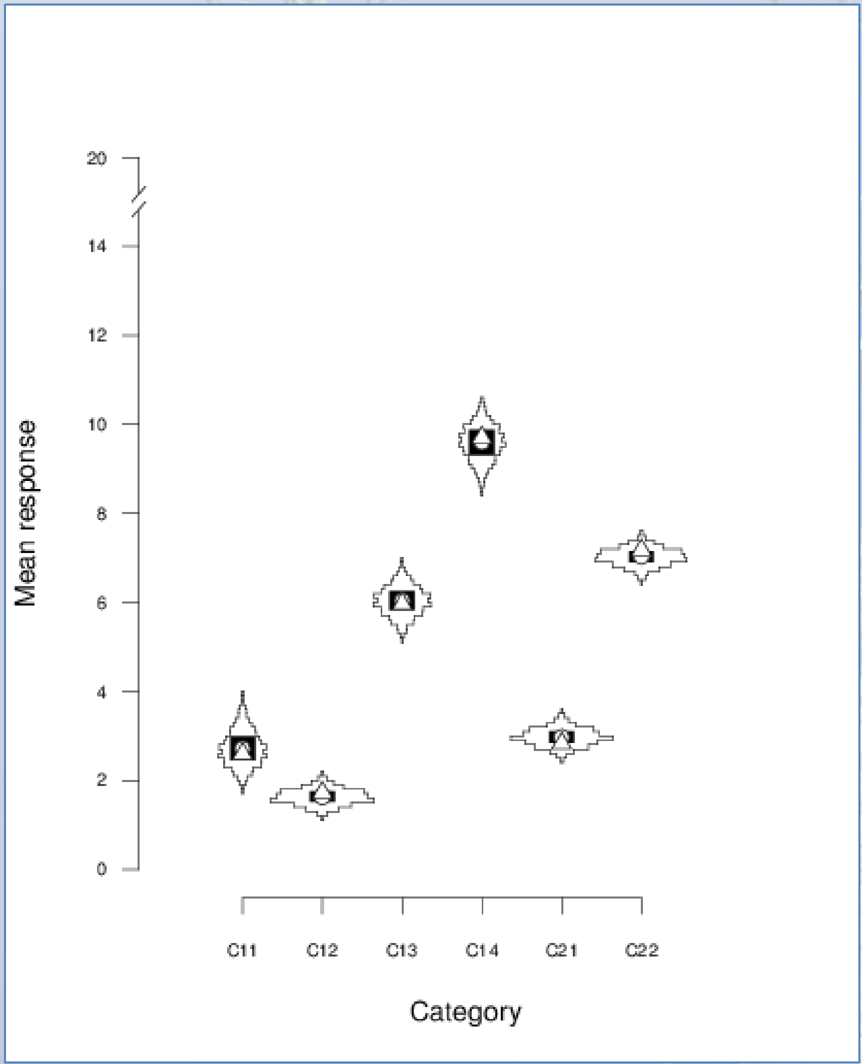
Summary

- WinBUGS implementations of a crossed–random effects MPT model
  - Conclusions about underlying cognitive processes
  - No need to aggregate
  - Flexible

Thank you for your attention!

Posterior Predictive Checks: Pure list





Posterior Predictive Checks: Mixed list