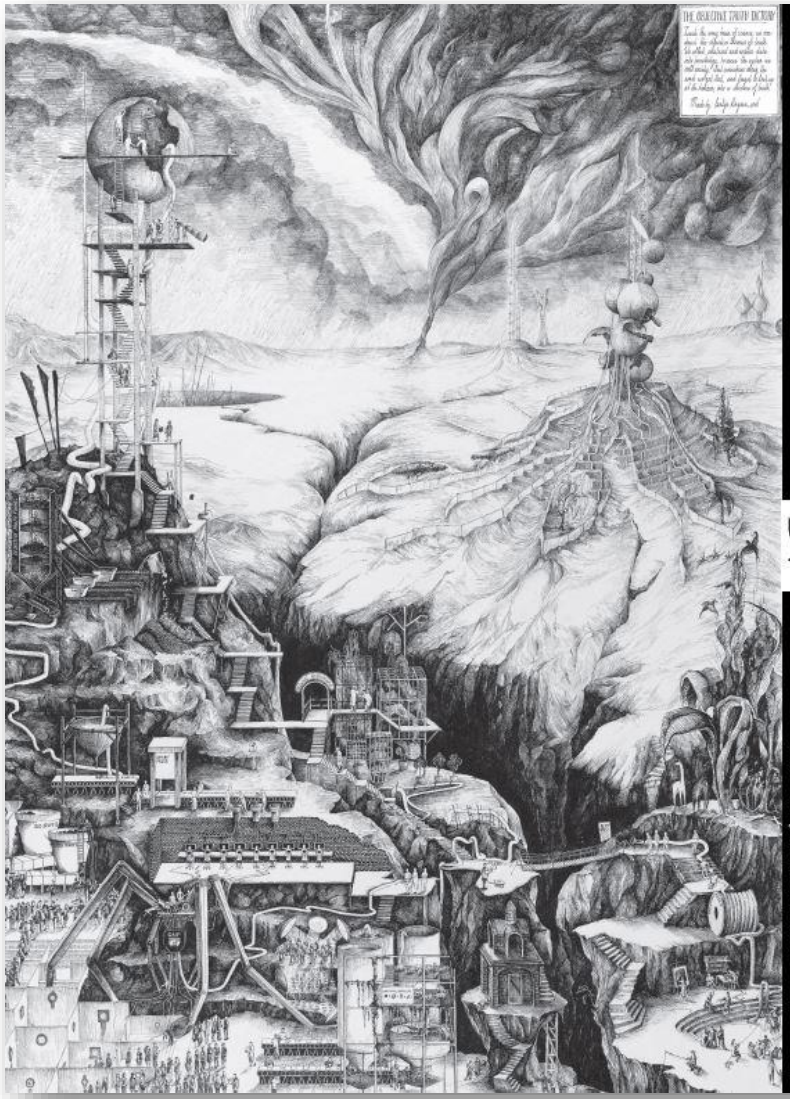


Alternative Information

Alternative Information

Duco Veen

DV

DV

Bayesian Statistics, Expert Elicitation and Information Theory in the Social Sciences

Bayesian Statistics, Expert Elicitation and Information Theory in the Social Sciences

Slides available at: ducoveen.com

Duco

- Let me briefly introduce myself

Outline

- About the project
- First elicitations
- Quality control
- Elicitation for hierarchical model

Please feel free to interrupt / as questions / discuss at any time!

About the project

- Part of a larger grant on small sample issues
  - Rens van de Schoot
- Roughly 2 parts
  - Simulation studies – Sanne Smid
  - Expert elicitation – Duco Veen

Universiteit Utrecht

About the project – Expert elicitation

- In the social sciences
  - Systematic review 2016: 2 cases of expert knowledge in 25 years of Bayes in Psychology
  - Experts with limited statistical knowledge

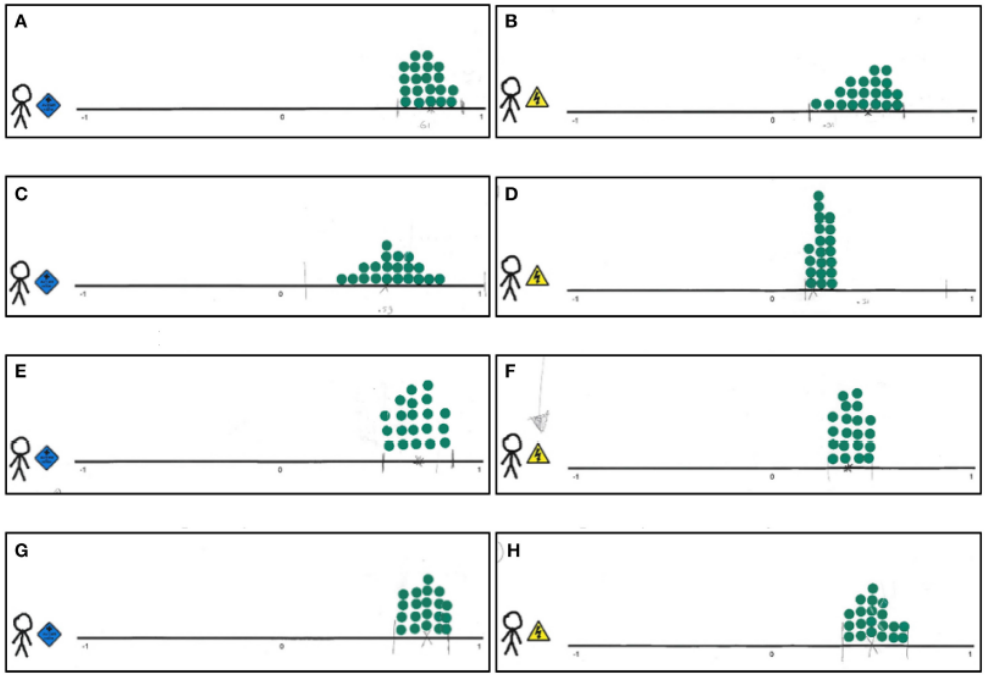
- End at Latent Growth Curve Model (Hierarchical model)

Expert elicitation – What to do?

- Direct vs. Indirect
  - quantile elicitation
  - predicting data

- Group vs. Individual

Expert elicitation – Work by collages



(Zondervan-Zwijnenburg et al., 2016)

Improving elicitation quality

- Avoid triggering of heuristics and biases
- Employ face-to-face elicitation
- Training experts and facilitators

Improving elicitation quality

- Providing Feedback
  - Intuition laypeople improved through graphical elicitation techniques (Goldstein & Rothschild, 2014)
  - Interpretation expert's beliefs
  - Explicit dialogue
- Can be incorporated trough software
  - Recommendation in O'Hagan et al. (2006)

Improving elicitation quality

- Providing Feedback
  - Intuition laypeople improved through graphical elicitation techniques (Goldstein & Rothschild, 2014)
  - Interpretation expert's beliefs
  - Explicit dialogue

- **Can be incorporated trough software**
  - **Recommendation in O'Hagan et al. (2006)**

Expert elicitation – Digitizing for feedback

- What is out there? – Systematic review
- What do we think works well with our experts?
  - Direct – indirect?

Systematic Review Software Use

- All references to O'Hagan et al. (2006) (n=840)
- Scopus search (n=1578)
  - papers with prior distributions and experts' beliefs
  - Based on search Johnson et al. (2010) sys. review elicitation methods
- All R packages on CRAN at September 15th 2016 (n=9178)

Systematic Review Software Use

- MATCH (Morris et al., 2014)
  - Based on SHELF (Oakley, 2016)
- Single use elicitation programs

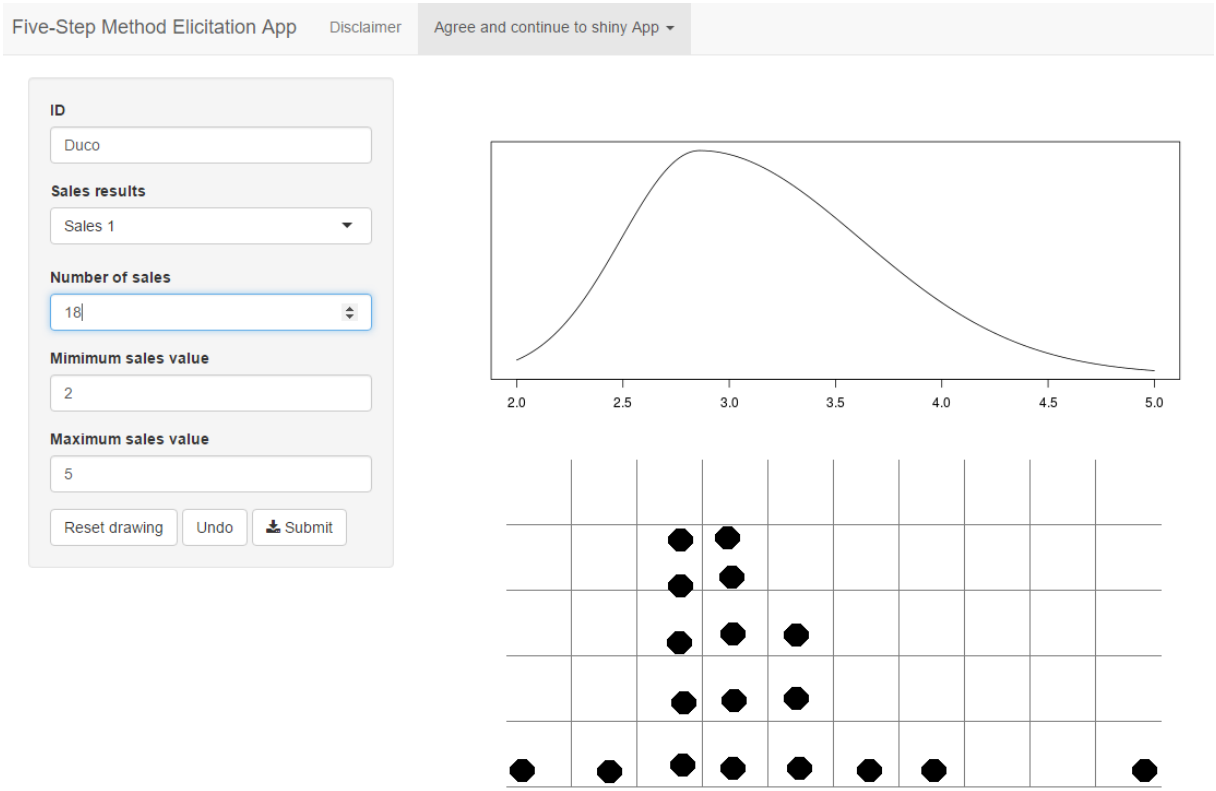
Expert elicitation – Digitizing for feedback

- Experts had difficulty with concept of hyperparameters with uncertainty
- Cut elicitation into smaller steps
- Combine direct and indirect

Expert elicitation – Five-step method

1) Elicit location parameter using trial roulette – direct elicitation
2) Provide feedback
3) Elicit scale and shape parameters
4) Provide feedback
5) Use elicited distribution

Five-step method – Steps 1 & 2



Five-step method – Steps 3 & 4

ID

Paper Example 2

Sales results

Sales 1

Total

1840

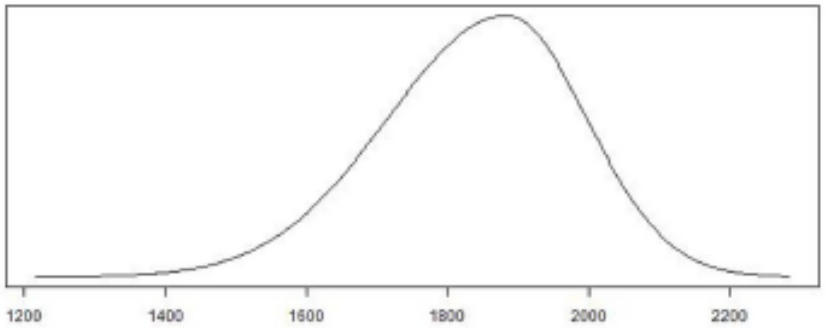
Reasonable lowerbound

1400

Reasonable upperbound

2100

Submit



Universiteit Utrecht

Five-step method – Steps 5

- Use elicited distributions
- But what about quality control?

Universiteit Utrecht

Questions so far?

Please feel free to interrupt / as questions / discuss at any time!

Expert elicitation – Quality control

- How and with what?
- Some critical notes

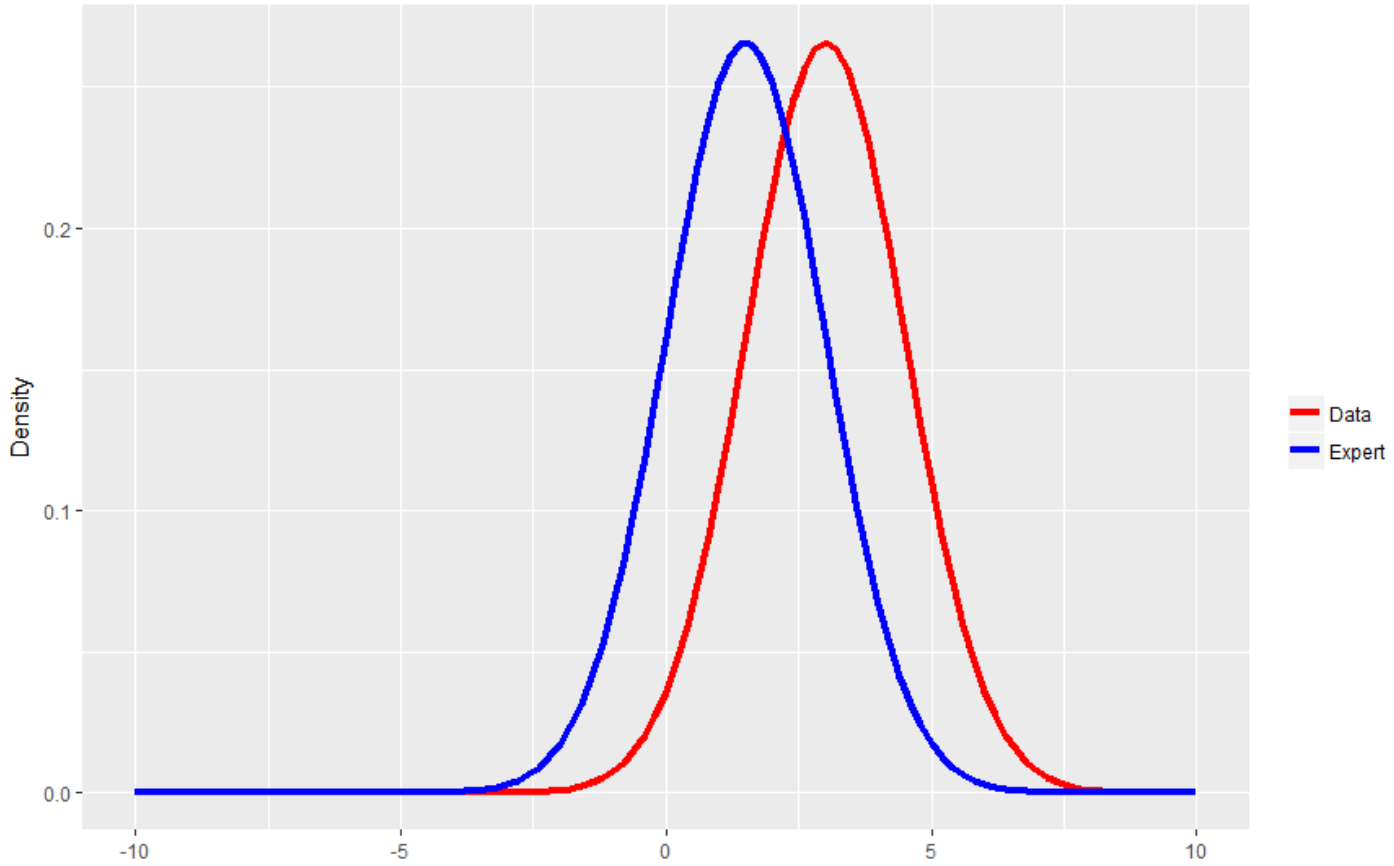
Universiteit Utrecht

Expert elicitation – Quality control

- Classical method
  - Calibration questions?
- Direct comparison expert priors and data
  - Prior predictive distributions – save bet to be uncertain
  - Prior-data conflict measure

Quality Control

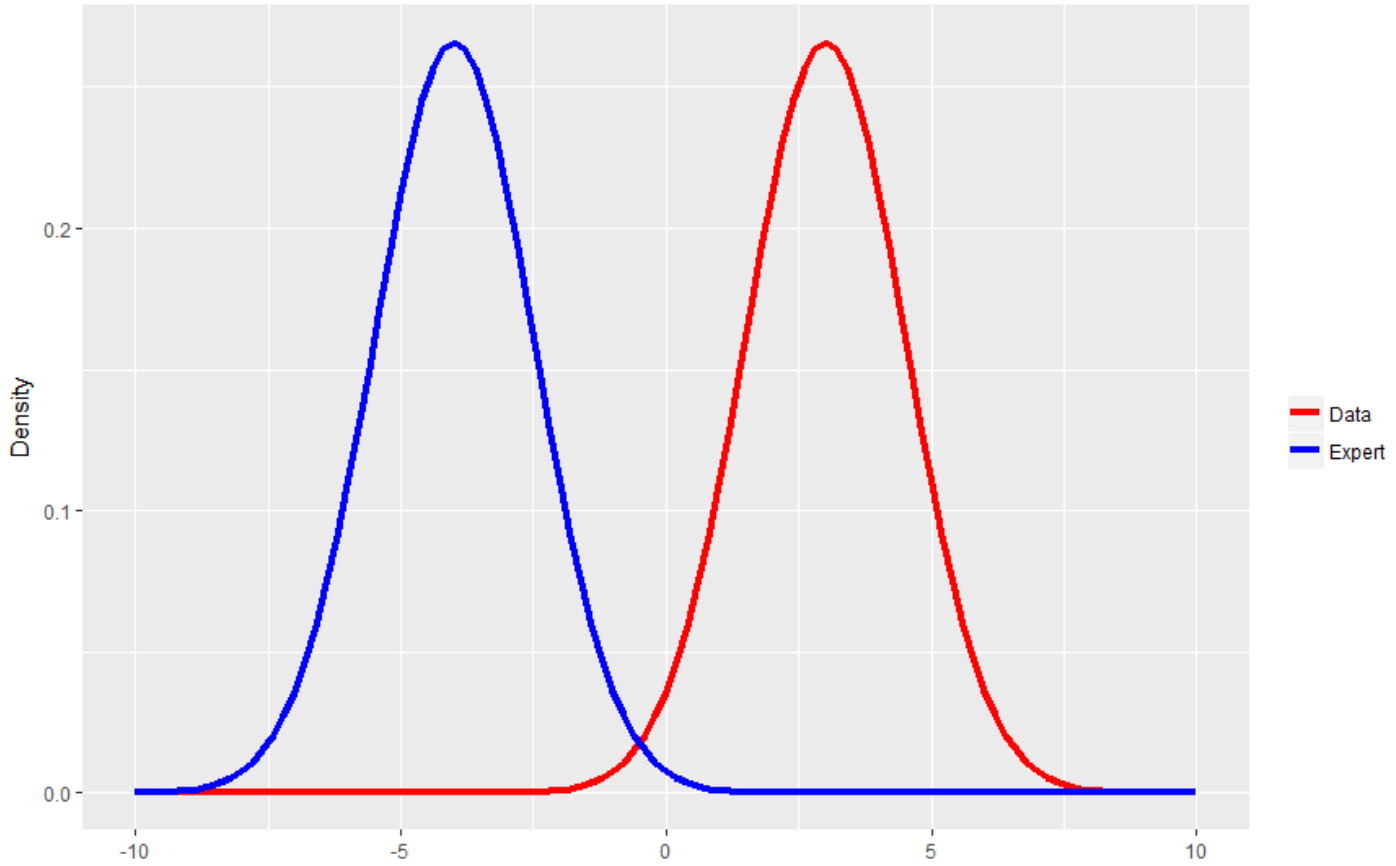
Prior-data Agreement



Universiteit Utrecht

profitwise

Prior-data Disagreement



Universiteit Utrecht

profitwise

Data Agreement Criterion[1]

- Ratio of two Kullback-Leibler divergences[2]

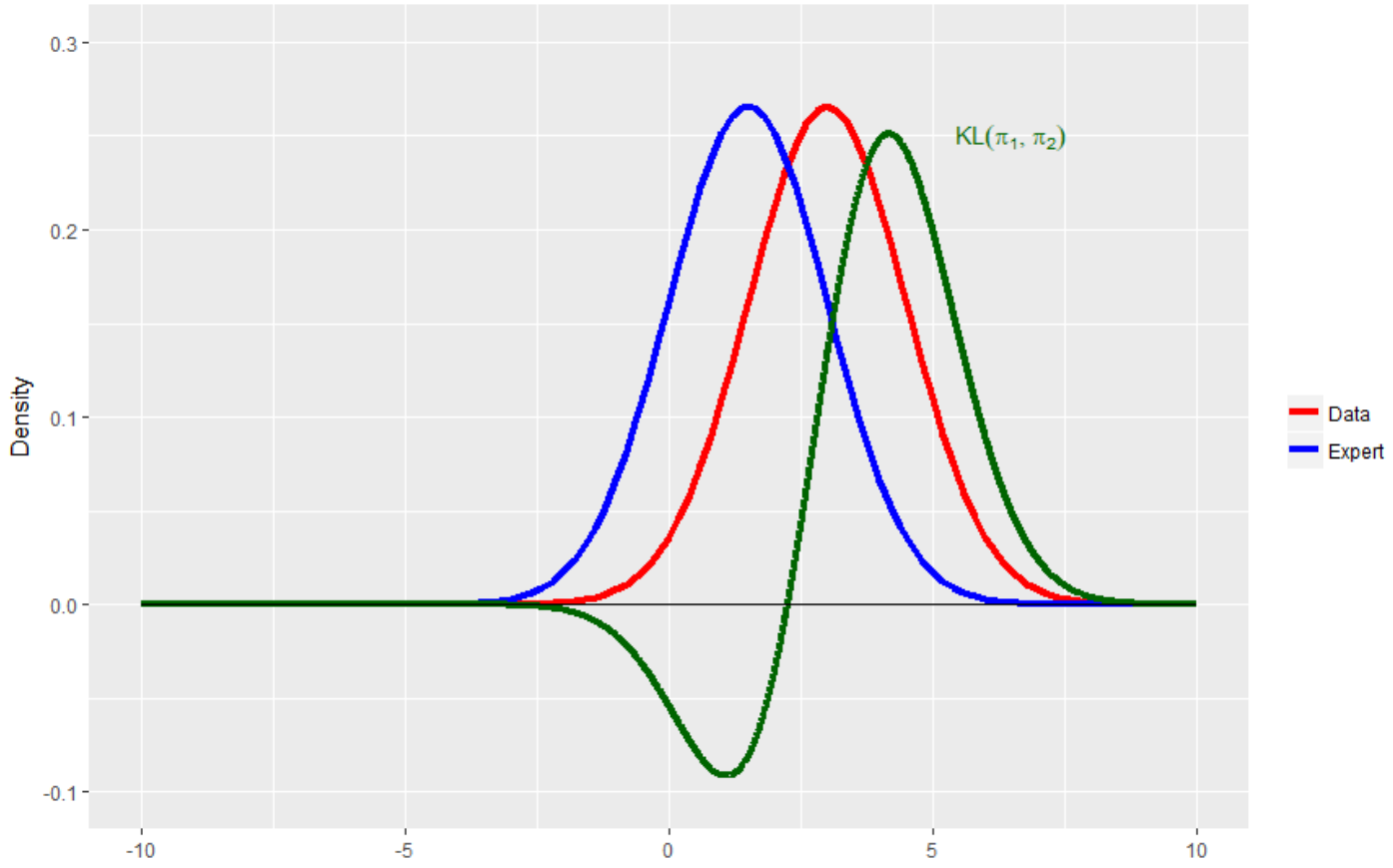
$$KL(\pi_1||\pi_2) = \int_{\Theta} \pi_1(\theta) \log \frac{\pi_1(\theta)}{\pi_2(\theta)} d\theta$$

Data Agreement Criterion[1]

- Ratio of two Kullback-Leibler divergences[2]

$$KL(\pi_1||\pi_2) = \int_{\Theta} \pi_1(\theta) \log \frac{\pi_1(\theta)}{\pi_2(\theta)} d\theta$$

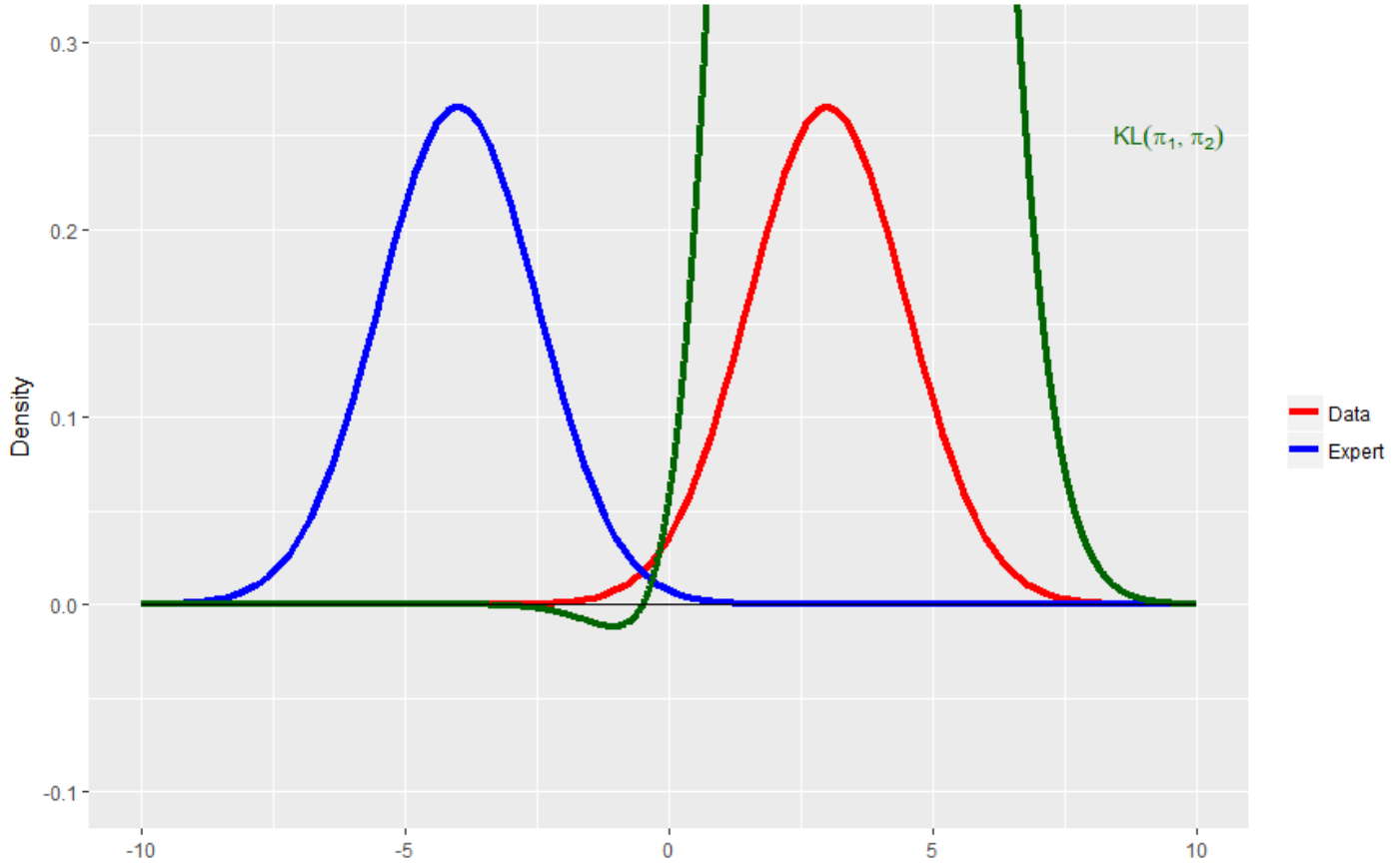
Kullback-Leibler Divergence



Universiteit Utrecht

profitwise

Kullback-Leibler Divergence



Data Agreement Criterion

$$\mathrm{DAC} = \frac{KL[\pi^{\mathrm{J}}(\theta|\mathbf{y}) \,||\pi(\theta)]}{KL[\pi^{\mathrm{J}}(\theta|\mathbf{y}) \,||\pi^{\mathrm{J}}(\theta)]}$$

Data Agreement Criterion

$$\mathrm{DAC} = \frac{KL[\pi^{\mathrm{J}}(\theta|\mathbf{y}) \,||\pi(\theta)]}{KL[\pi^{\mathrm{J}}(\theta|\mathbf{y}) \,||\pi^{\mathrm{J}}(\theta)]}$$

Data Agreement Criterion

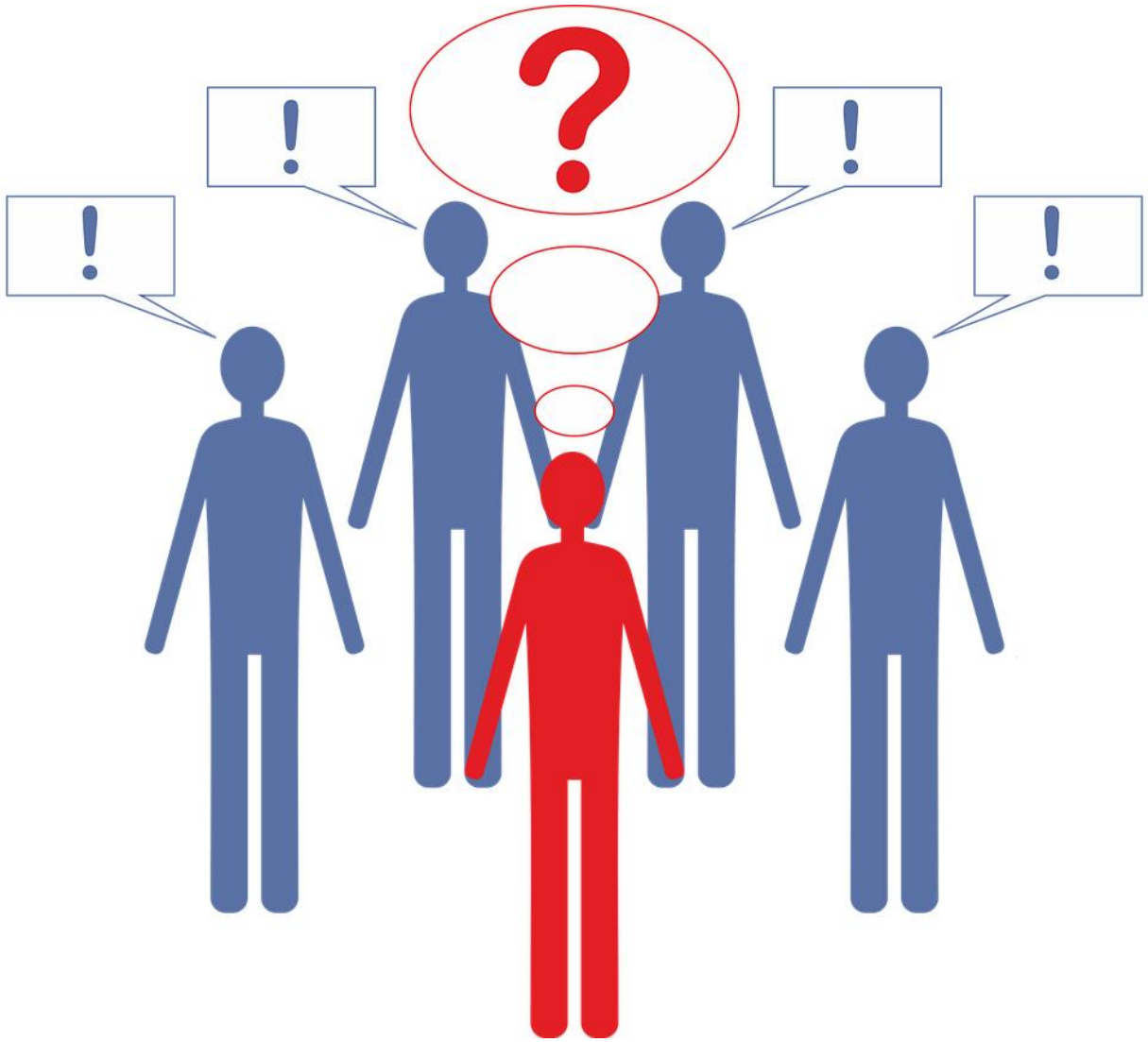
$$\mathrm{DAC} = \frac{KL[\pi^{\mathrm{J}}(\theta|\mathbf{y}) \,||\pi(\theta)]}{KL[\pi^{\mathrm{J}}(\theta|\mathbf{y}) \,||\pi^{\mathrm{J}}(\theta)]}$$

Multiple Experts

$$\mathrm{DAC}_d = \frac{KL[\pi^{\mathrm{J}}(\theta|\mathbf{y}) \,||\pi_d(\theta)]}{KL[\pi^{\mathrm{J}}(\theta|\mathbf{y}) \,||\pi^{\mathrm{J}}(\theta)]}$$

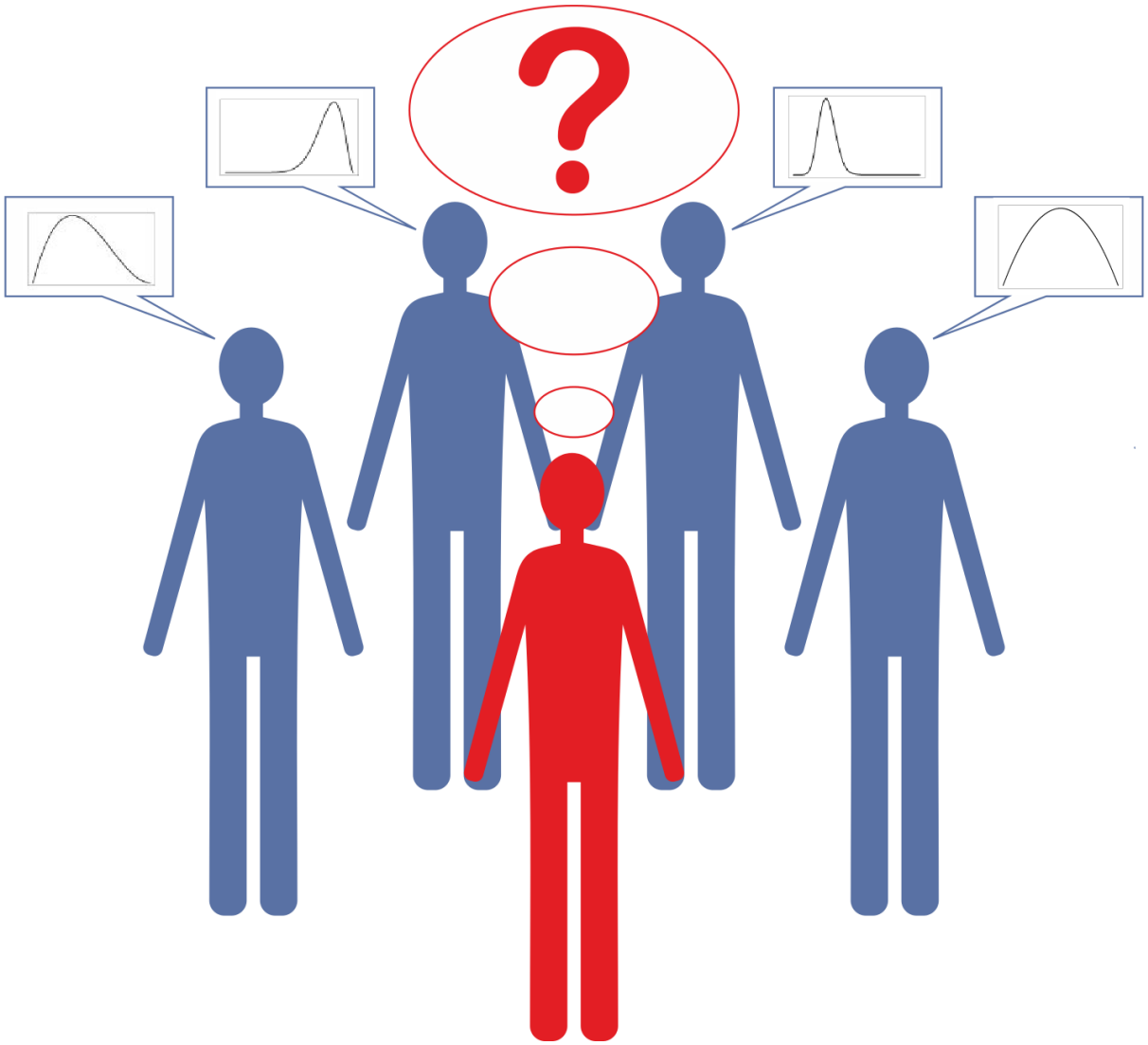
Universiteit Utrecht

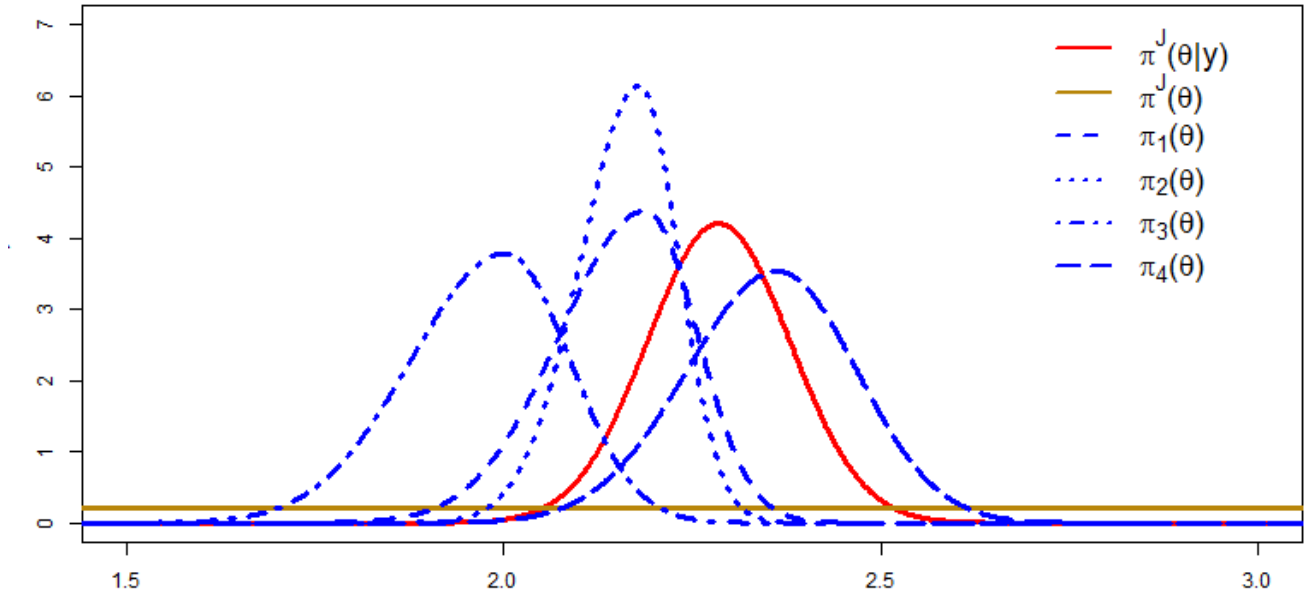
profitwise

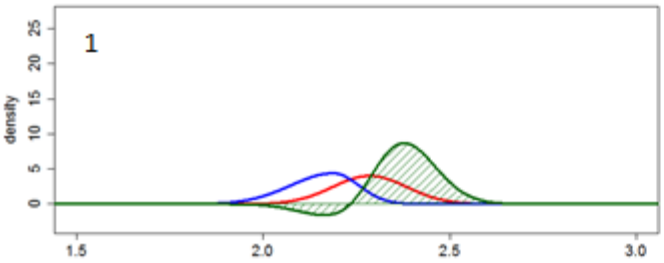


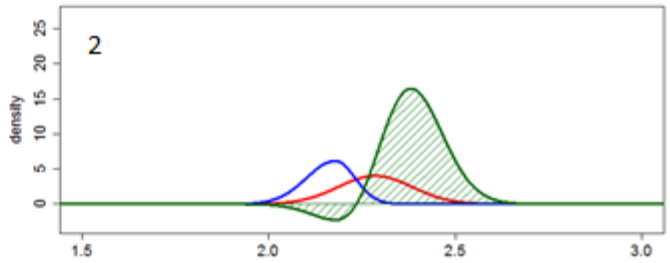
Universiteit Utrecht

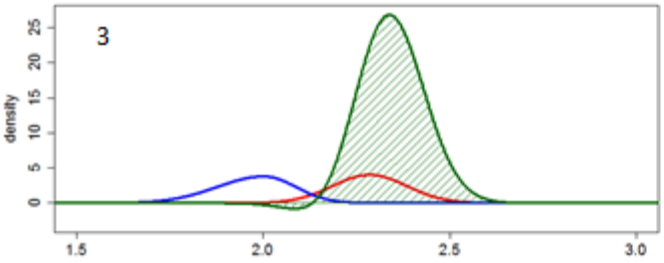
profitwise

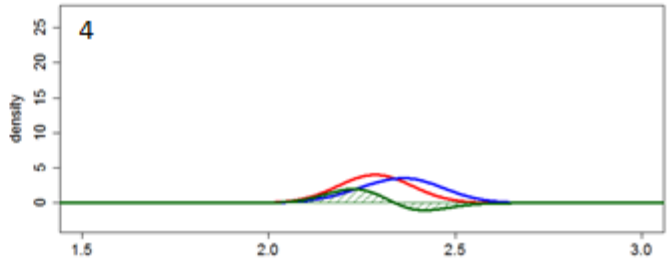


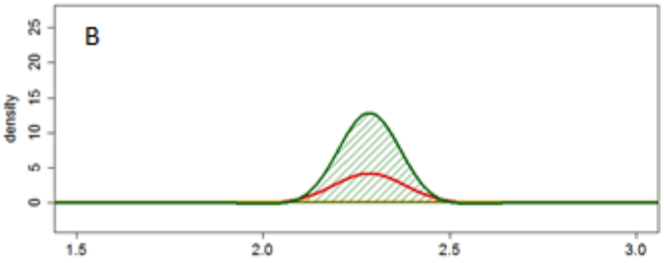

0
1
2
3
4
5
6
7
1.5
2.0
2.5
3.0
$\pi^J(\theta|y)$
$\pi^J(\theta)$
$\pi_1(\theta)$
$\pi_2(\theta)$
$\pi_3(\theta)$
$\pi_4(\theta)$





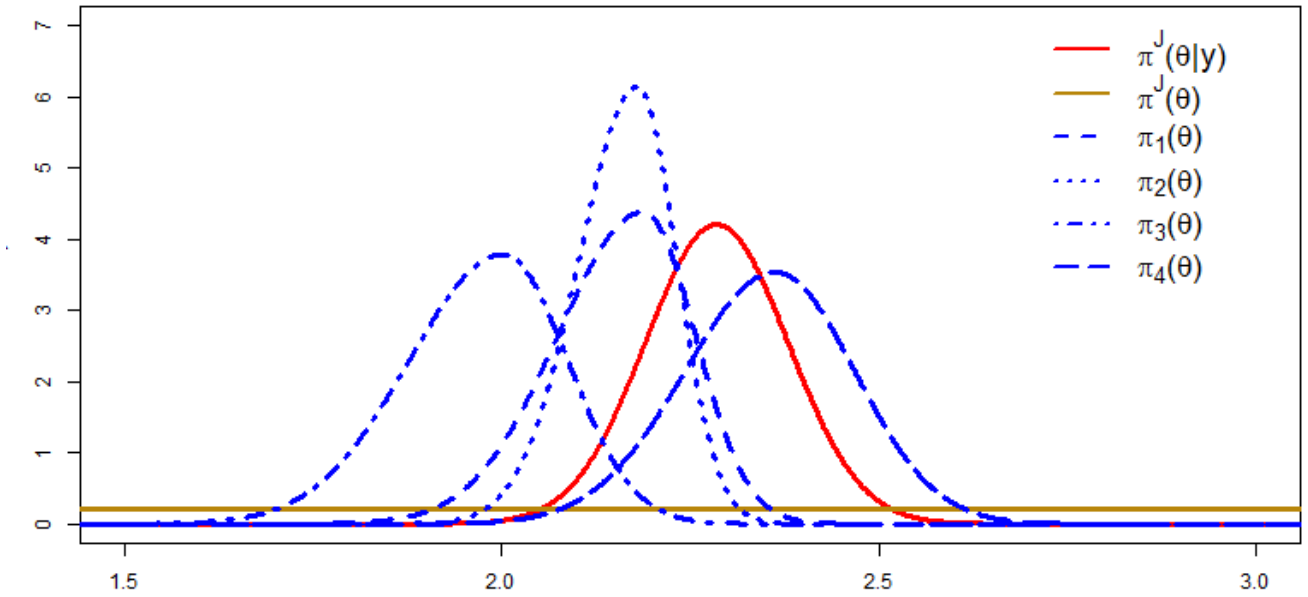






Universiteit Utrecht

profitwise



	KL divergence	DAC_d	Ranking
Expert 1	**1.43**	**0.56**	**2**
Expert 2	**2.86**	**1.12**	**3**
Expert 3	**5.76**	**2.26**	**4**
Expert 4	**0.19**	**0.07**	**1**
Benchmark	**2.55**	**-**	**-**

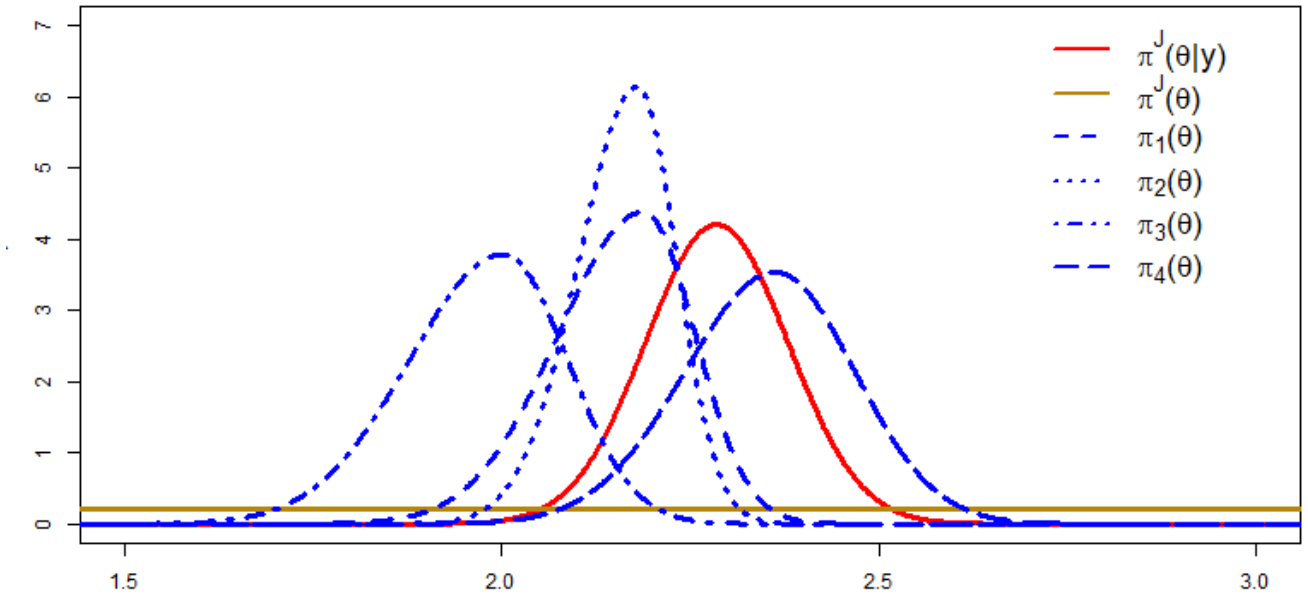
Alternative suggested by reviewer

- Compare to Bayes Factor
- Using Marginal Likelihood instead of KL divergence
- Differentiate experts by assessing the probability of the data averaged across their specified prior beliefs.

$$\mathrm{BF}_{JD} = \frac{m^J(\mathbf{y})}{m_d(\mathbf{y})}$$

Comparing BF and DAC

$$DAC_{2,d}^{J} = \frac{m^{J}(\boldsymbol{y})}{m_{d}(\boldsymbol{y})} \exp\{KL[\pi^{J}(.|\boldsymbol{y})||\pi_{d}(.|\boldsymbol{y})]\}$$



	DAC_d	Ranking	BF_{JD}	Ranking
Expert 1	**0.56**	**2**	**0.21**	**3**
Expert 2	**1.12**	**3**	**0.17**	**2**
Expert 3	**2.26**	**4**	**5.31**	**4**
Expert 4	**0.07**	**1**	**0.07**	**1**

Universiteit Utrecht

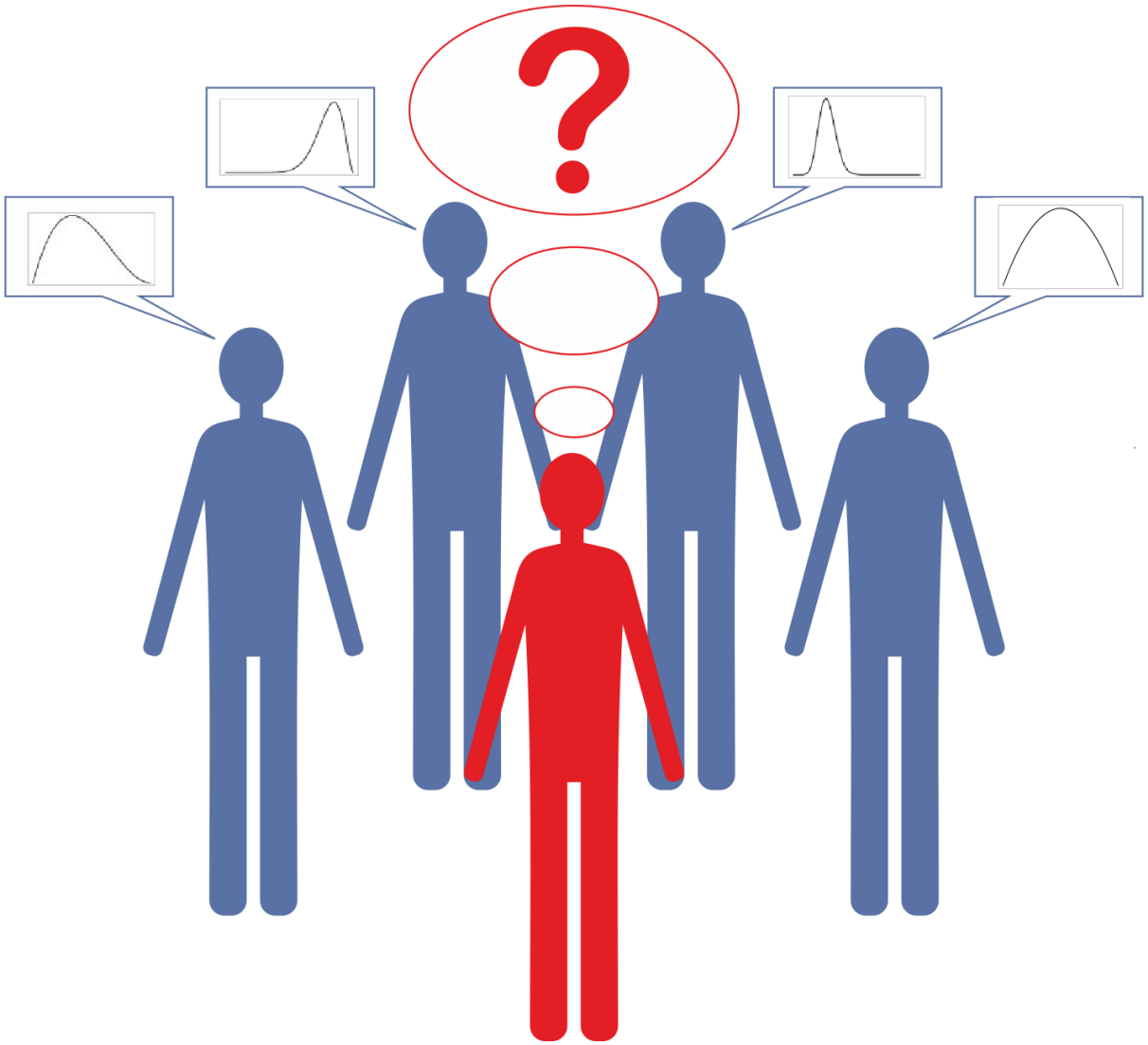
profitwise

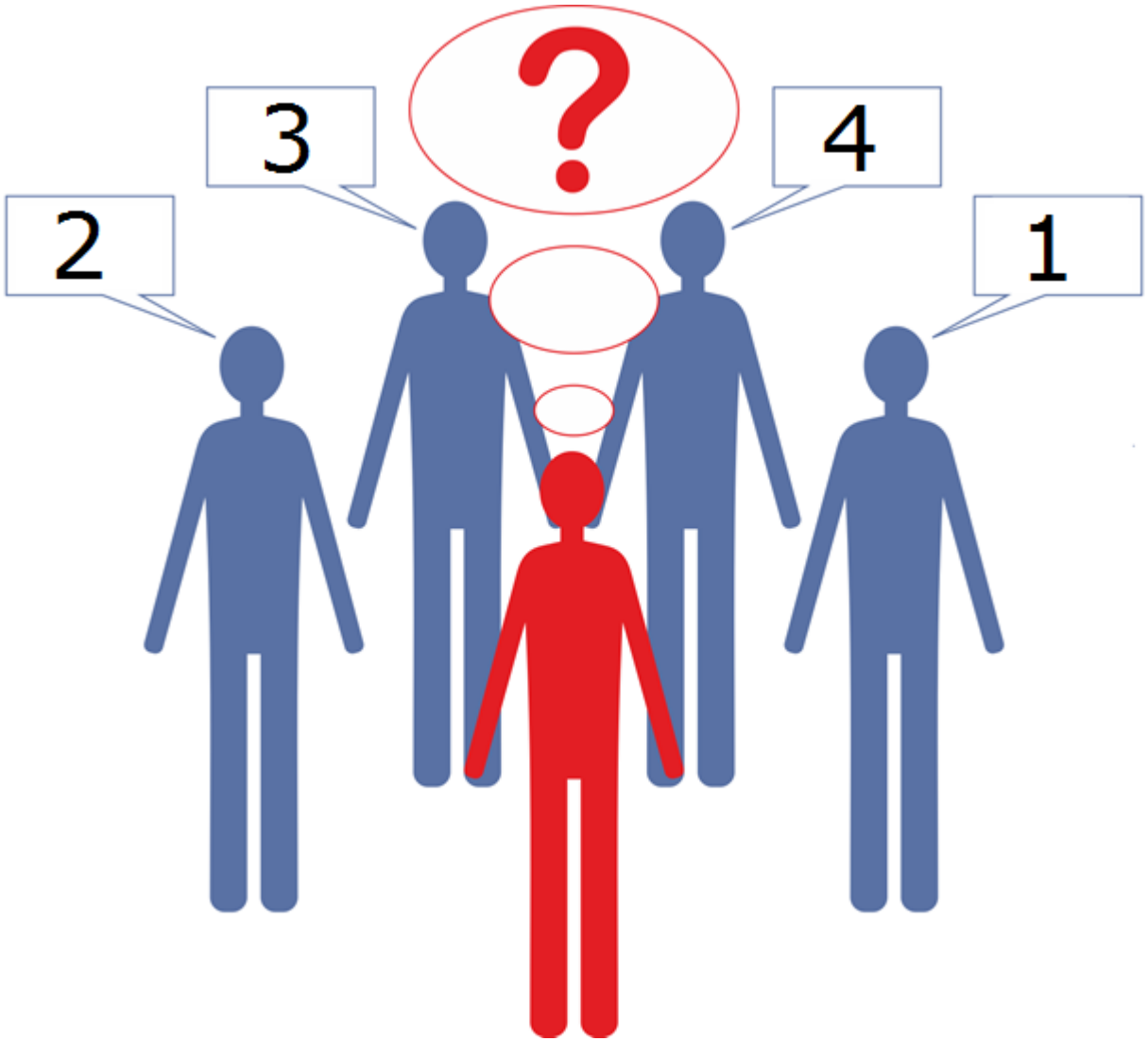
The limiting case

- Expert specifying Dirac delta function
- If location of delta function coincides with region of parameter space for which likelihood > 0
- Both KL divergence and Marginal Likelihood are infinity
  - They mean the opposite though

Universiteit Utrecht

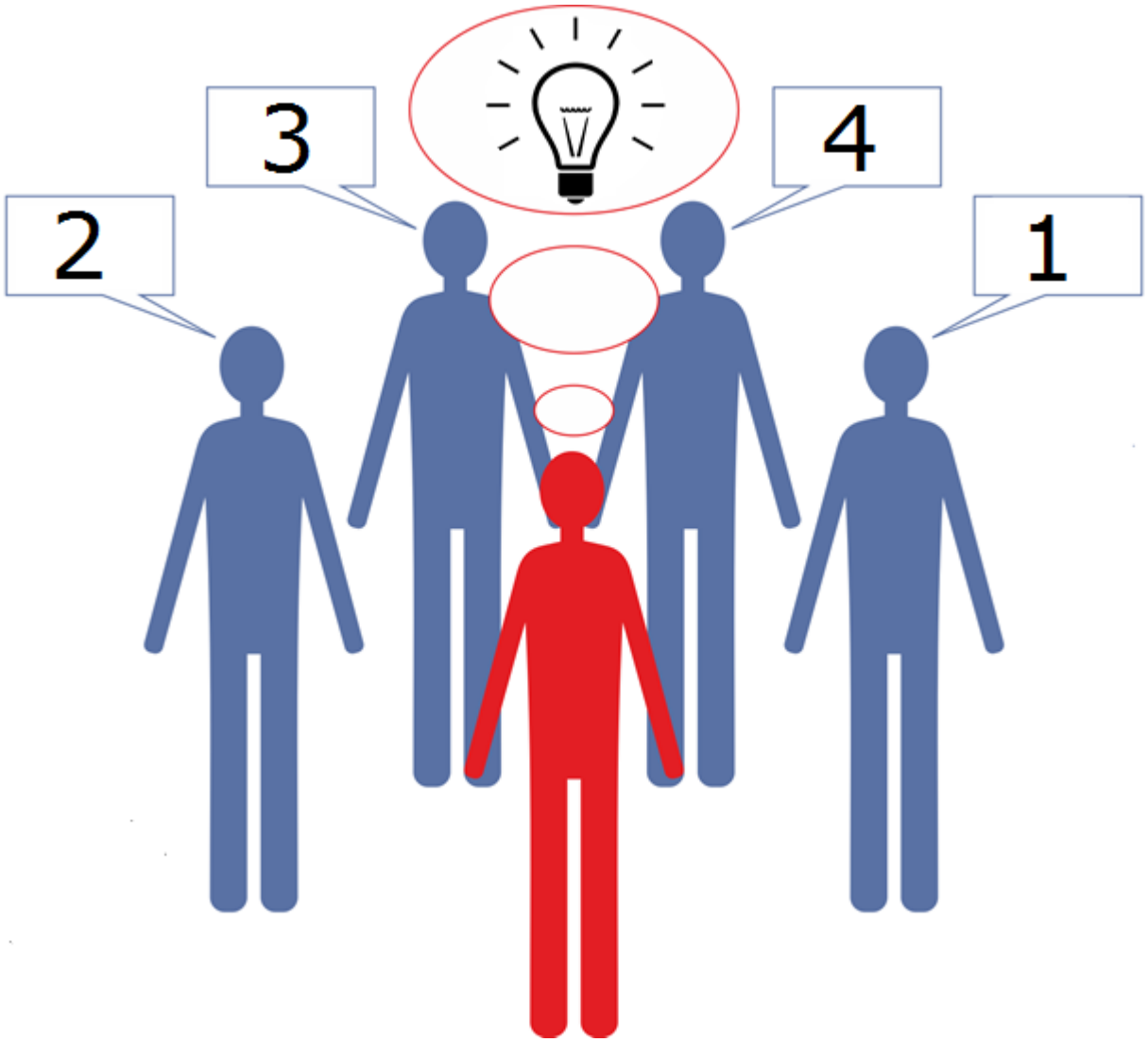
profitwise





Universiteit Utrecht

profitwise



Universiteit Utrecht

profitwise

Prior-data (dis)agreement – Critical notes

- Based on one sample data
  - Is the data correct / representative
- Would like to use feedback / learning and reach convergence over time
  - Not allowed in our case by ethical committee
  - Great for future research
- Are benchmarks meaningful?

Elicitation for hierarchical model

- Extending the Five-step method from before
- Context of impact of pediatric burn injuries

Universiteit Utrecht

Impact of pediatric burn injuries

Impact of pediatric burn injuries

Posttraumatic stress symptoms

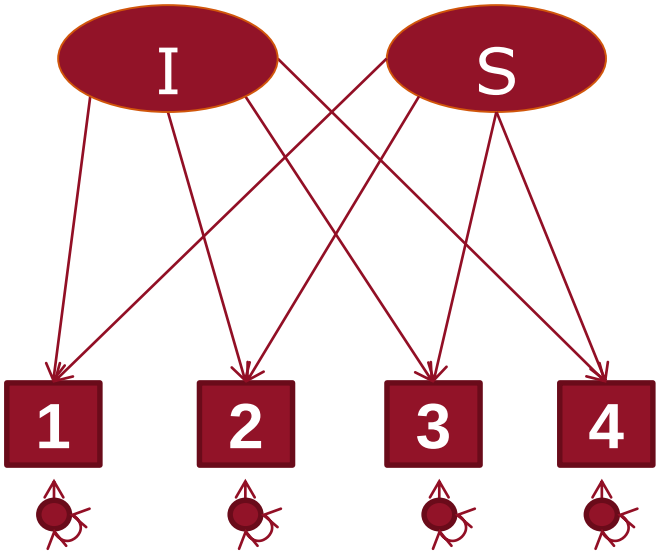
Elicitation for hierarchical model

- 8–18-year old from Netherlands and Belgium
- Minimal 24 stay
- Minimal percentage of body burned of 1%
- Self-reported posttraumatic stress symptoms

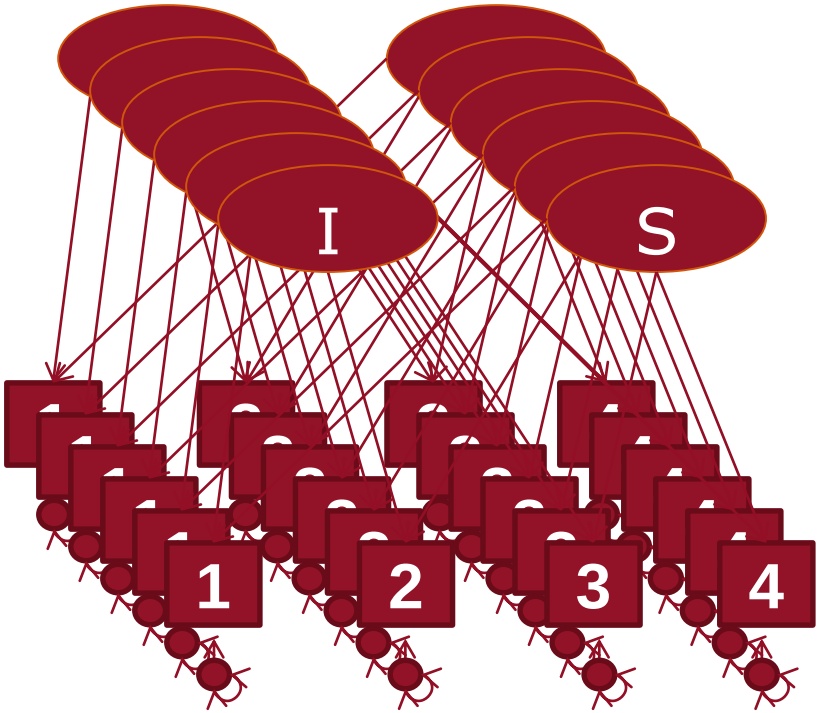
Elicitation for hierarchical model

- 7 nurses specialized at working with burn-injuries
- 7 psychologists working with the children
- From all 3 Dutch burn-institutes
- Audio recordings of elicitations for qualitative information

Single model



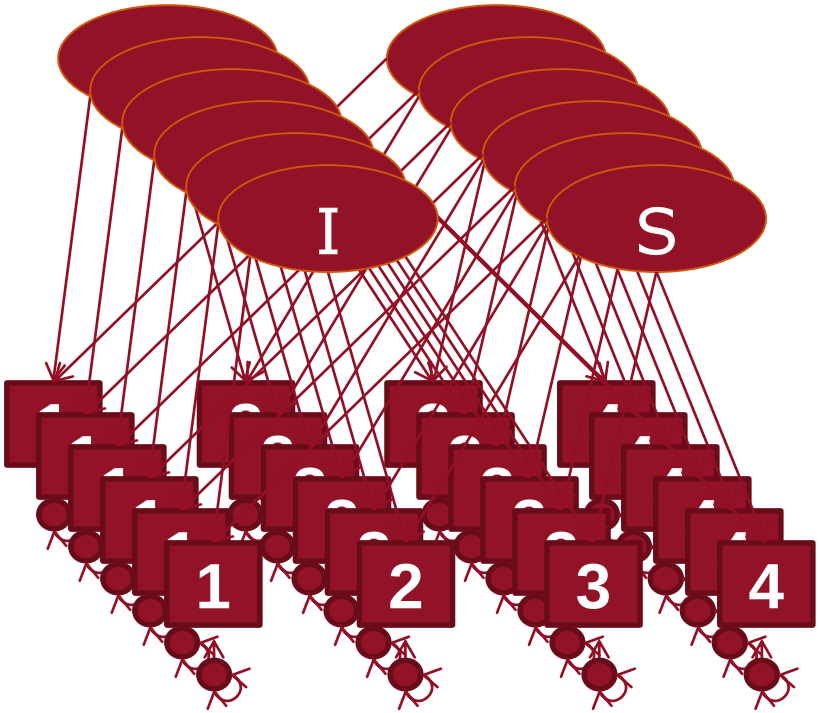
Hierarchical model



$I \sim N(\mu_I, \sigma_I^2)$

$S \sim N(\mu_S, \sigma_S^2)$

Hierarchical model



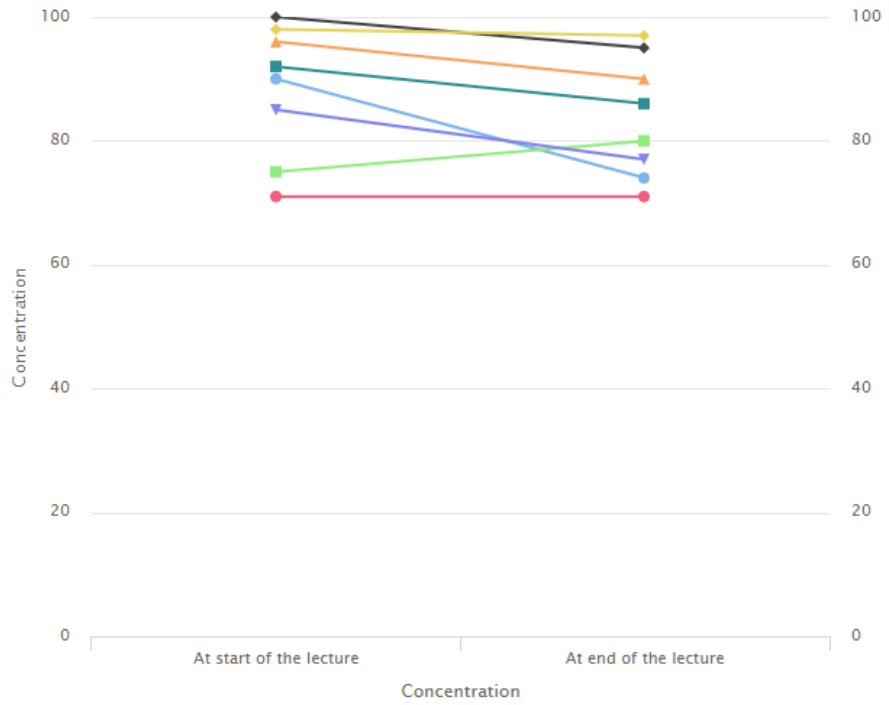
$I \sim N(\mu_I, \sigma_I^2)$

$S \sim N(\mu_S, \sigma_S^2)$

Add Aditional Line

Show average trajectory

Submit



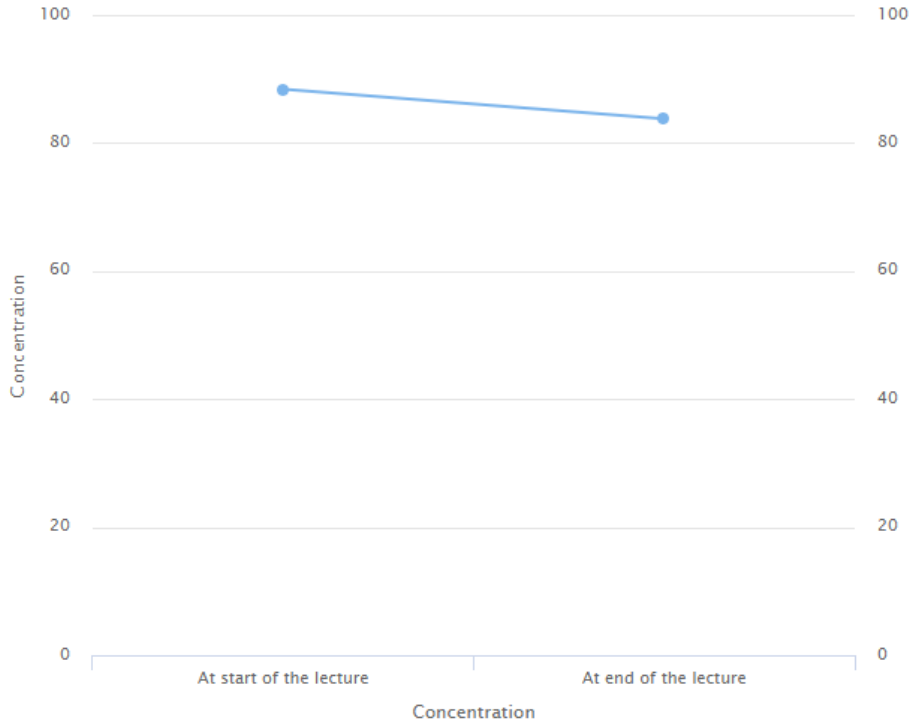
	data
Average concentration at start of the lecture	88.38
Average change in concentration from start to end of the lecture	-4.62

Universiteit Utrecht

Add Aditional Line

☑ Show average trajectory

Submit



	data
Average concentration at start of the lecture	88.38
Average change in concentration from start to end of the lecture	-4.62

Universiteit Utrecht

Reasonable lowerbound average concentration at start of lecture

95

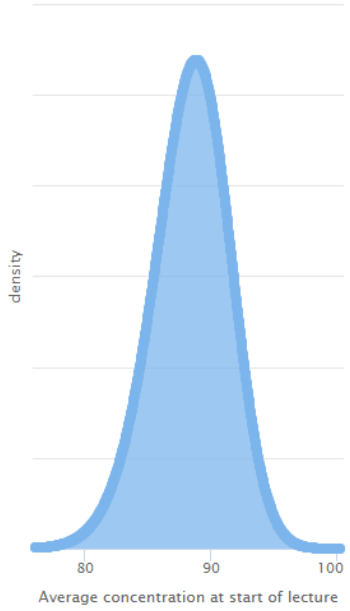
Average average concentration at start of lecture

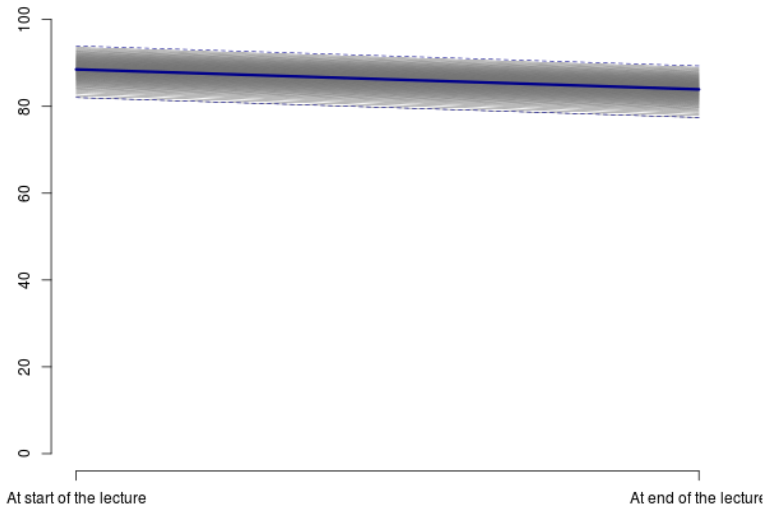
88,4

Reasonable upperbound average concentration at start of lecture

80

Fit distribution | Show implications





2.5%	25%	50%	75%	97.5%
82	86.5	88.6	90.5	93.9

95% CI	50% CI
[82, 93.9]	[86.5, 90.5]

Reasonable lowerbound average change in concentration

0

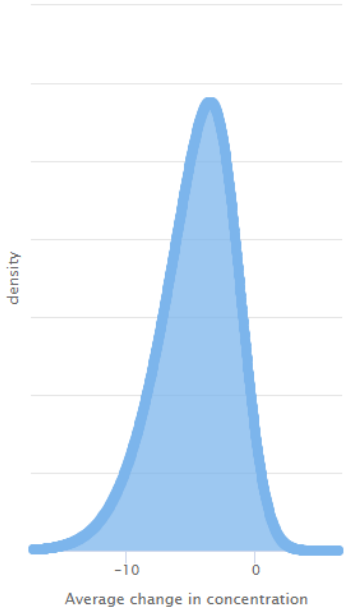
Average change in concentration

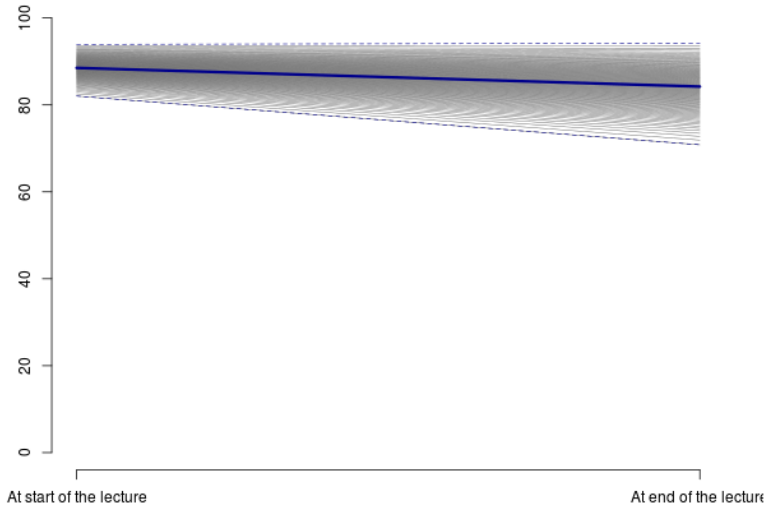
-4,6

Reasonable upperbound average change in concentration

-15

Fit distribution | Show implications



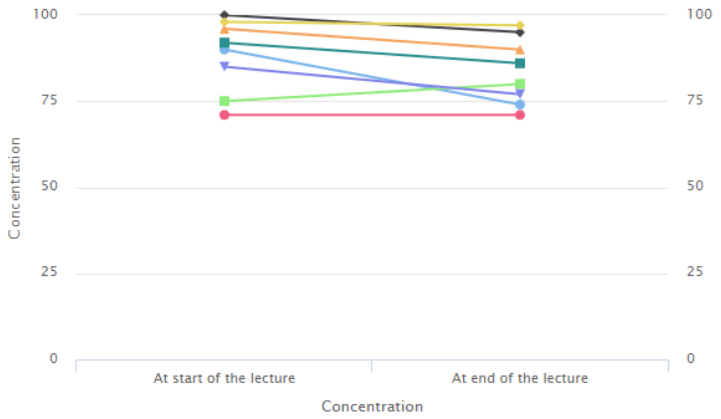


concentration

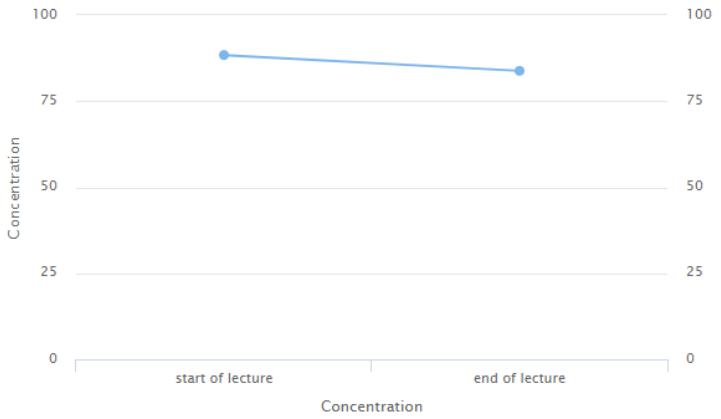
2.5%	25%	50%	75%	97.5%
-11.2	-6.4	-4.3	-2.5	0.3

95% CI	50% CI
[-11.2, 0.3]	[-6.4, -2.5]

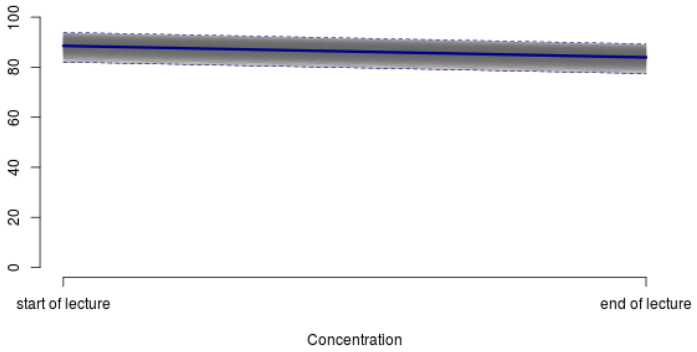
In this tab we provide a final summary of how we interpret your elicited beliefs and you can either agree to this or we go back to the relevant section of the procedure to adapt your input and our interpretation of your beliefs.

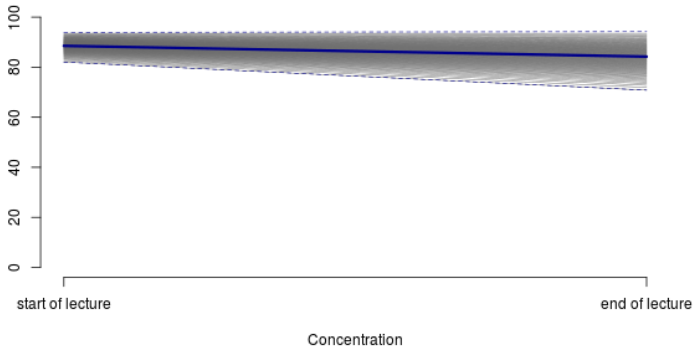


These are the concentration levels for your imagined individual children



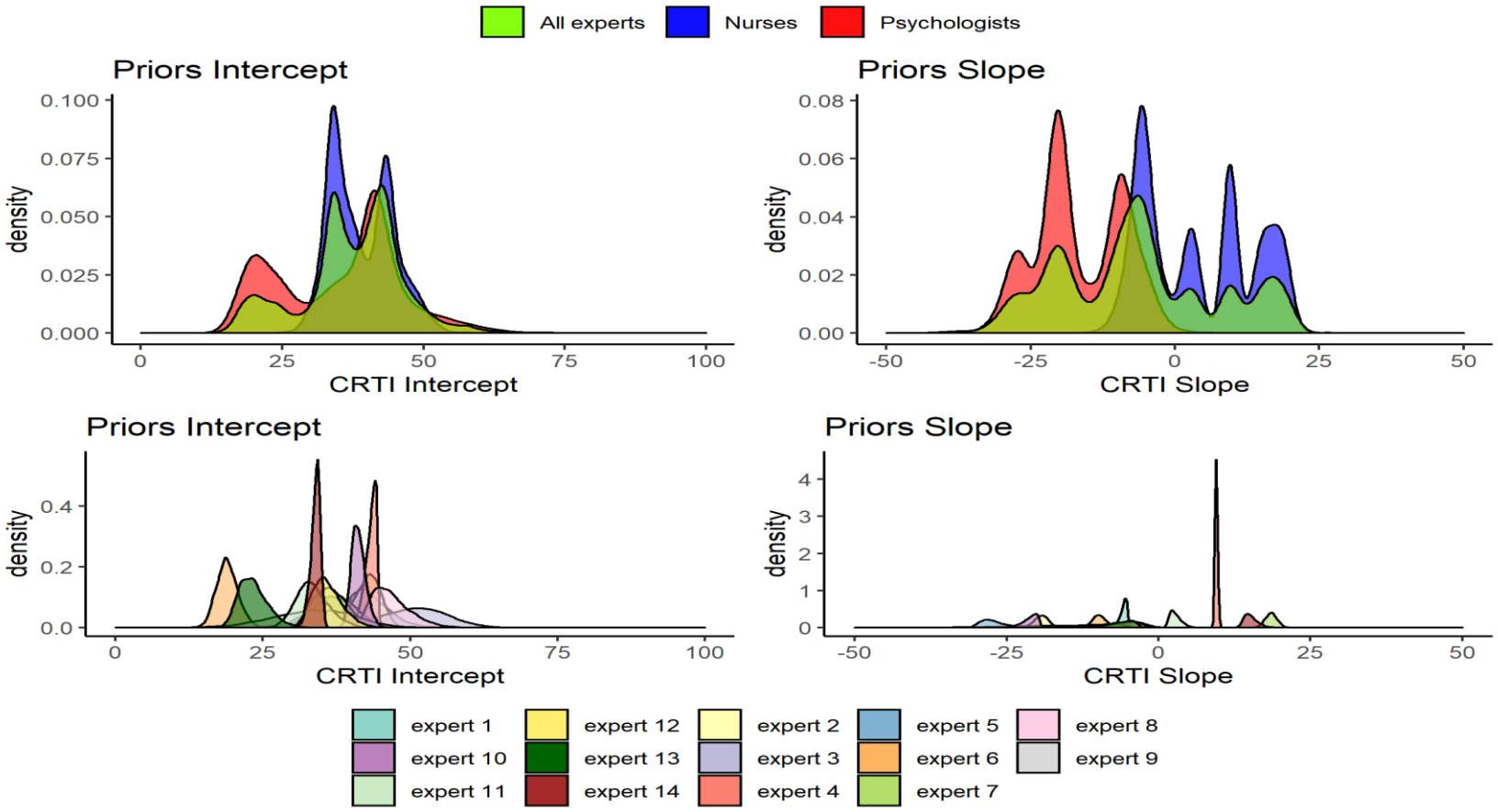
This is our interpretation of your beliefs regarding the average concentration levels at the start and the end of the lecture.



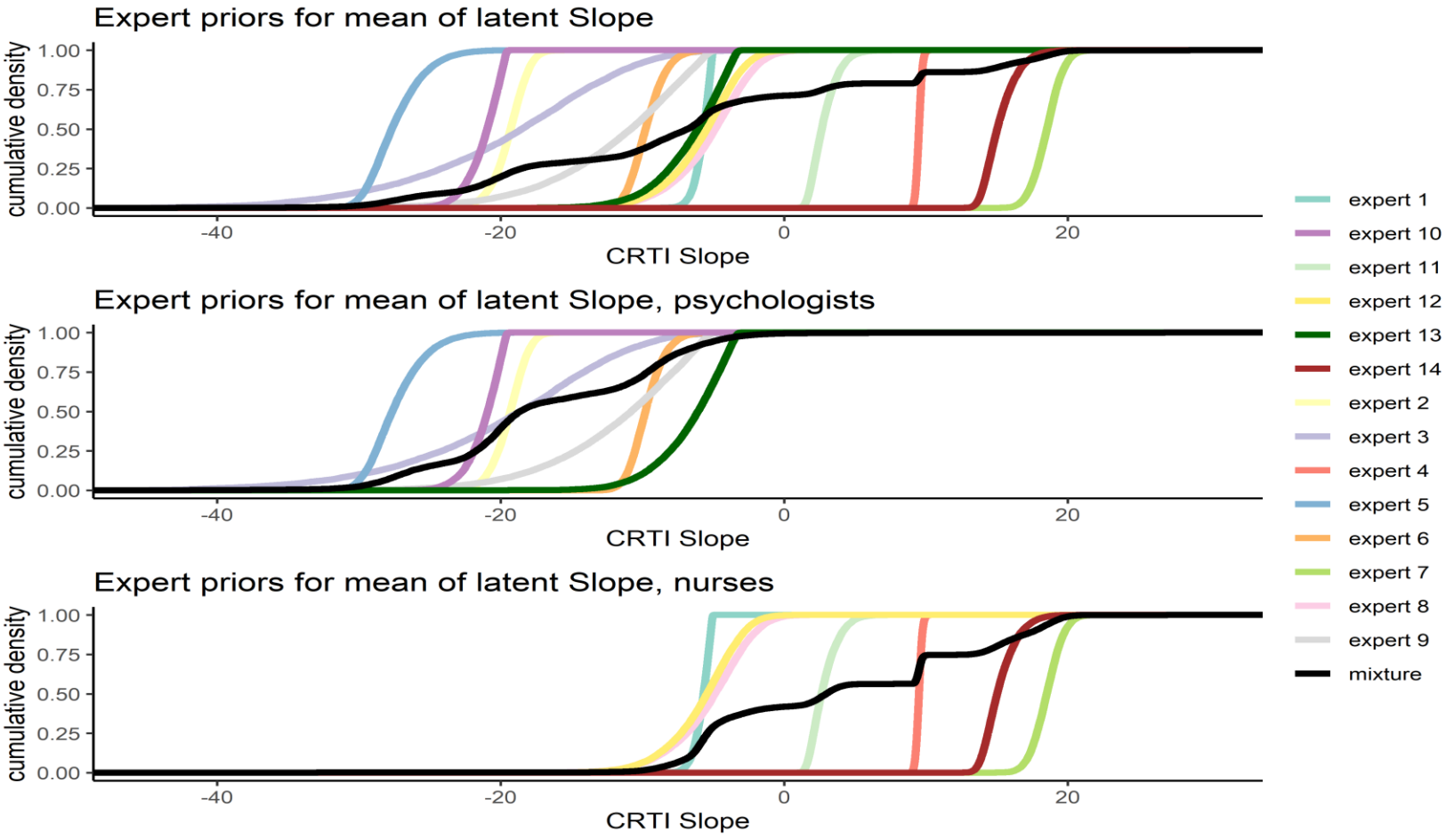


Universiteit Utrecht

Results

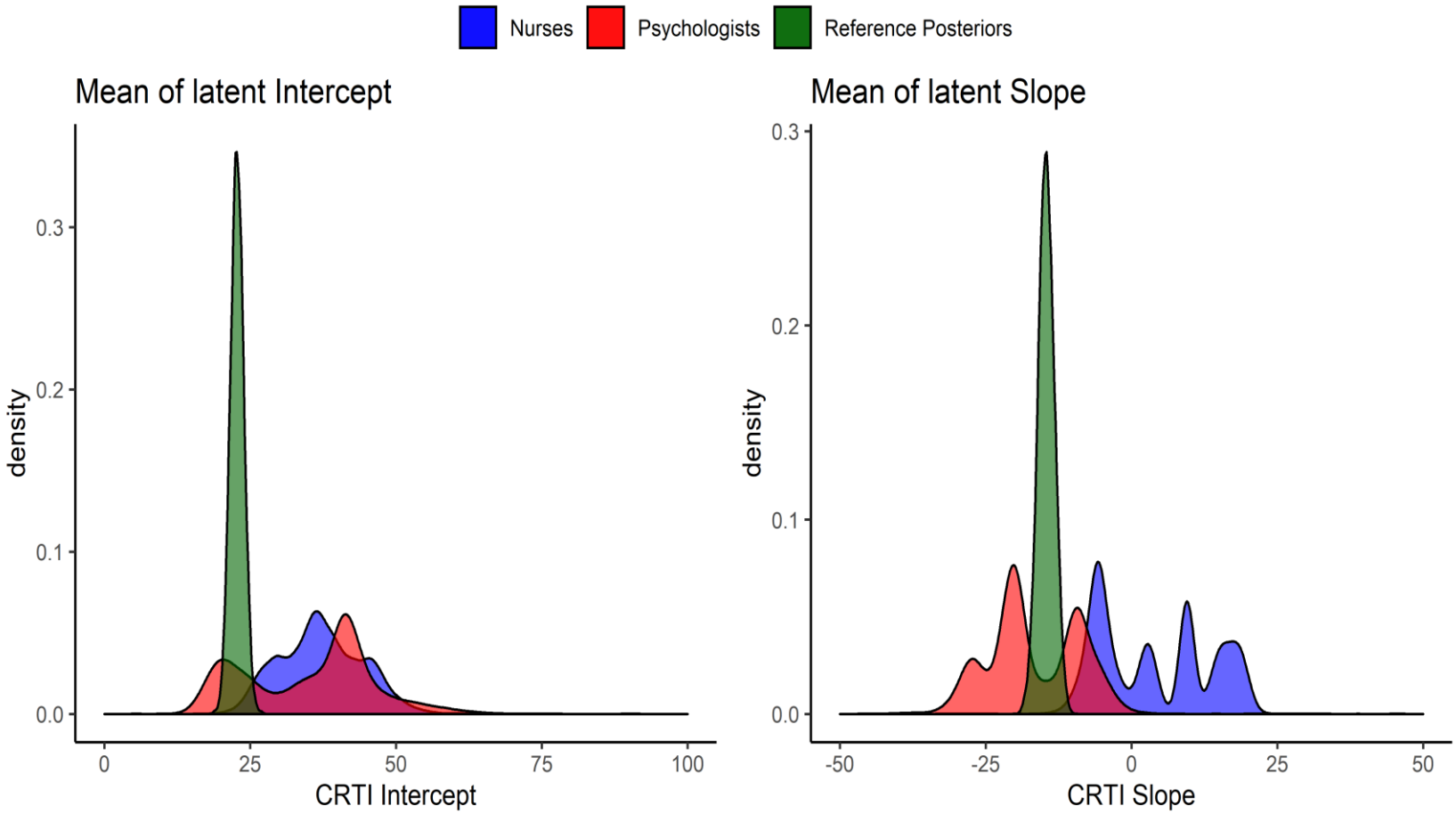

All experts
Nurses
Psychologists
Priors Intercept
density
CRTI Intercept
Priors Slope
CRTI Slope
Priors Intercept
density
CRTI Intercept
Priors Slope
density
CRTI Slope
expert 1
expert 10
expert 11
expert 12
expert 13
expert 14
expert 2
expert 3
expert 4
expert 5
expert 6
expert 7
expert 8
expert 9

Results



Universiteit Utrecht

Results



Universiteit Utrecht

Results – KL divergences

	Intercept	Slope
Benchmark 1	3.04	3.56
Benchmark 2	8.56	8.39
Nurses	8.19	5.88
Psychologists	1.99	2.18
All	2.72	2.63
Expert 1	42.87	59.18
Expert 2	45.16	25.87
Expert 3	6.71	1.23
Expert 4	72.86	55.38
Expert 5	5.66	98.32
Expert 6	2.1	22.17
Expert 7	79.2	59.61
Expert 8	46.97	4.37
Expert 9	2.48	1.28
Expert 10	43.74	67.55
Expert 11	12.78	64.56
Expert 12	99.94	4.88
Expert 13	0.35	3.62
Expert 14	75	74.11

Results – Audio recordings

- Referring specifically to (concepts of) PTSS
  - All psychologists
  - Only two nurses, though lost of mention of stress
- Expressing sentiment of more severe cases come to mind
  - 5 nurses – 1 psychologist
- Three psychologists reflected on linearity assumption of model

Results – Audio recordings

- Thee experts actively reflected based on visual feedback and adjusted their input
  - One psychologists and two nurses
- One experts stated that although they were sure about the direction of the trajectory, they felt unsure about the associated numerical representation
- Finally, one expert repeatedly mentioned that they found the task hard to do

What would you conclude?