Explaining variability and conservatism in causal inference: the Bayesian Mutation Sampler

Ivar Kolvoort

i.r.kolvoort@uva.nl
Department of Psychological Methods & Institute for Logic, Language, and Computation
University of Amsterdam

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 44 (2020) e12839

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12839

A Process Model of Causal Reasoning

Zachary J. Davis, Bob Rehder

Department of Psychology, New York University

Received 16 October 2018; received in revised form 16 March 2020; accepted 6 April 2020

Abstract

How do we make causal judgments? Many studies have demonstrated that people are capable causal reasoners, achieving success on tasks from reasoning to categorization to interventions. However, less is known about the mental processes used to achieve such sophisticated judgments. We propose a new process model—the *mutation sampler*—that models causal judgments as based on a *sample* of possible states of the causal system generated using the Metropolis–Hastings sampling algorithm. Across a diverse array of tasks and conditions encompassing over 1,700 participants, we found that our model provided a consistently closer fit to participant judgments than standard causal graphical models. In particular, we found that the biases introduced by mutation sampling accounted for people's consistent, predictable errors that the normative model by definition could not. Moreover, using a novel experimental methodology, we found that those biases appeared in the samples that participants explicitly judged to be representative of a causal system. We conclude by advocating sampling methods as plausible process-level accounts of the computations specified by the causal graphical model framework and highlight opportunities for future research to identify not just *what* reasoners compute when drawing causal inferences, but also *how* they compute it.

Keywords: Sampling; Causal representation; Causal reasoning; Process models

American Psychological Association

Psychological Review

ISSN: 0033-295X

http://dx.doi.org/10.1037/rev0000190

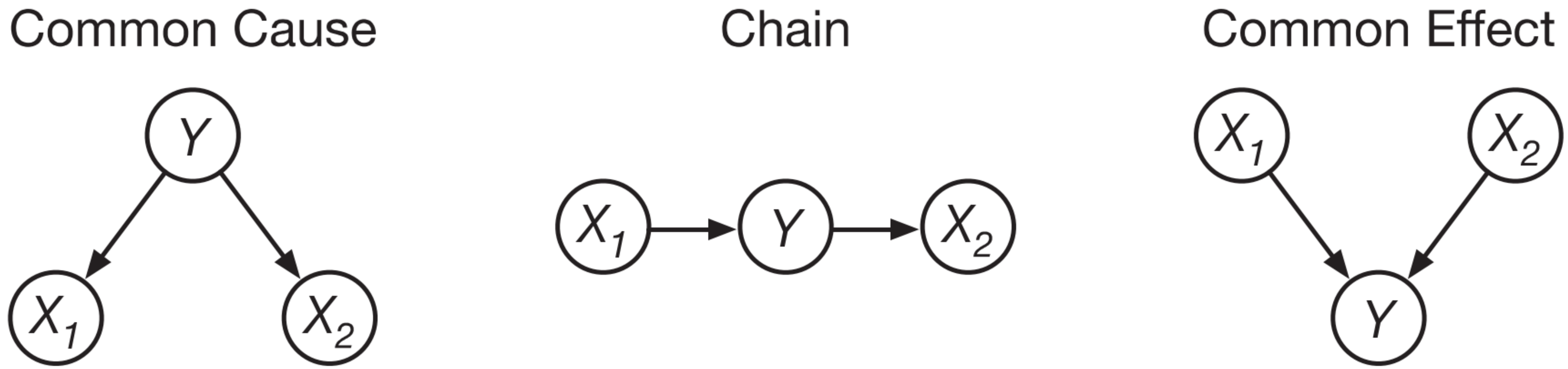
The Bayesian Sampler: Generic Bayesian Inference Causes Incoherence in Human Probability Judgments

Jian-Qiao Zhu, Adam N. Sanborn, and Nick Chater
University of Warwick

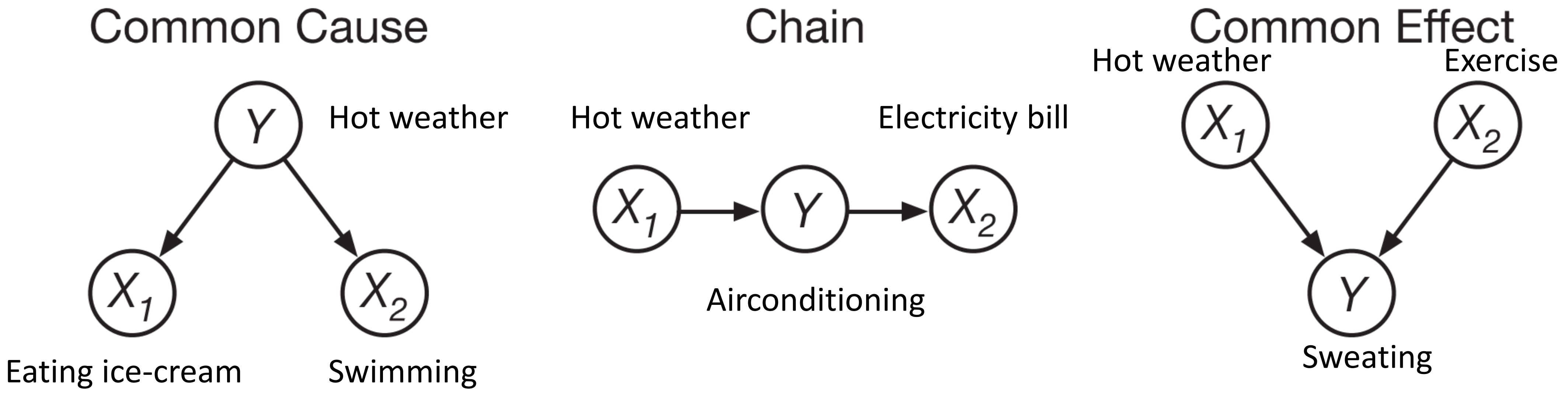
Human probability judgments are systematically biased, in apparent tension with Bayesian models of cognition. But perhaps the brain does not represent probabilities explicitly, but approximates probabilistic calculations through a process of sampling, as used in computational probabilistic models in statistics. Naïve probability estimates can be obtained by calculating the relative frequency of an event within a sample, but these estimates tend to be extreme when the sample size is small. We propose instead that people use a generic prior to improve the accuracy of their probability estimates based on samples, and we call this model the Bayesian sampler. The Bayesian sampler trades off the coherence of probabilistic judgments for improved accuracy, and provides a single framework for explaining phenomena associated with diverse biases and heuristics such as conservatism and the conjunction fallacy. The approach turns out to provide a rational reinterpretation of "noise" in an important recent model of probability judgment, the *probability theory plus noise* model (Costello & Watts, 2014, 2016a, 2017; Costello & Watts, 2019; Costello, Watts, & Fisher, 2018), making equivalent average predictions for simple events, conjunctions, and disjunctions. The Bayesian sampler does, however, make distinct predictions for conditional probabilities and distributions of probability estimates. We show in 2 new experiments that this model better captures these mean judgments both qualitatively and quantitatively; which model best fits individual distributions of responses depends on the assumed size of the cognitive sample.

Keywords: sampling, approximation, biases, Bayes, noise

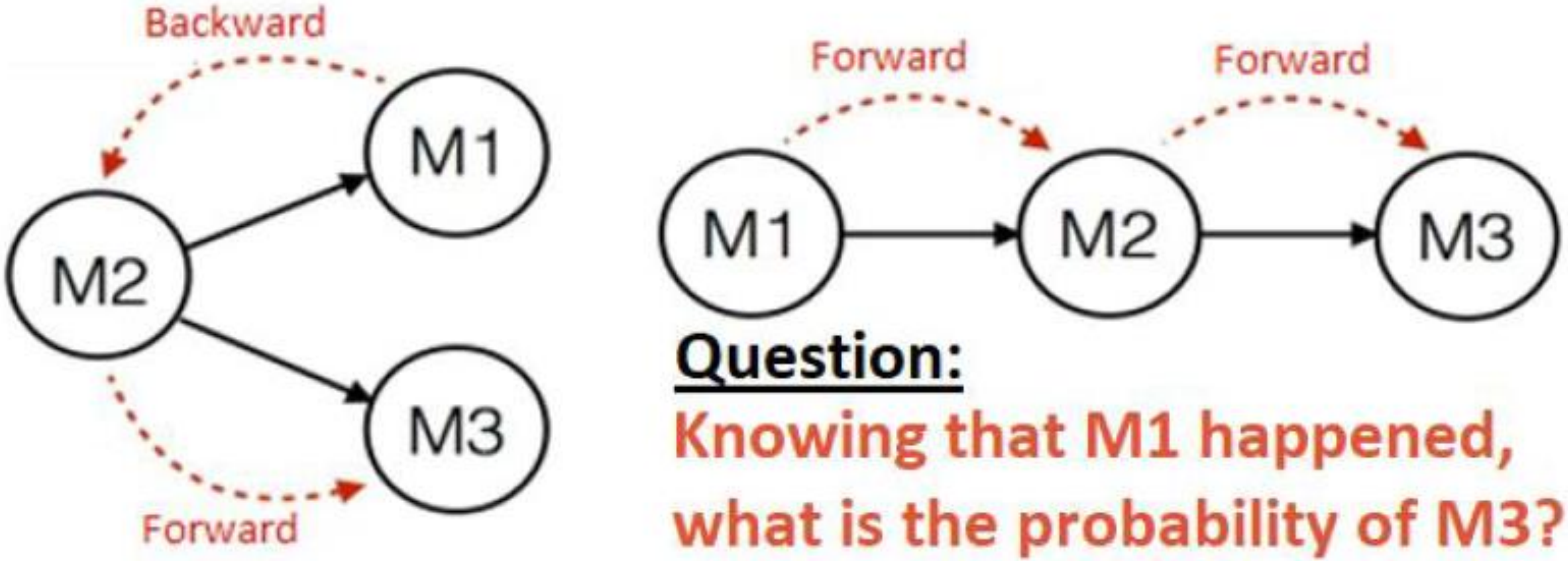
3 variable causal networks

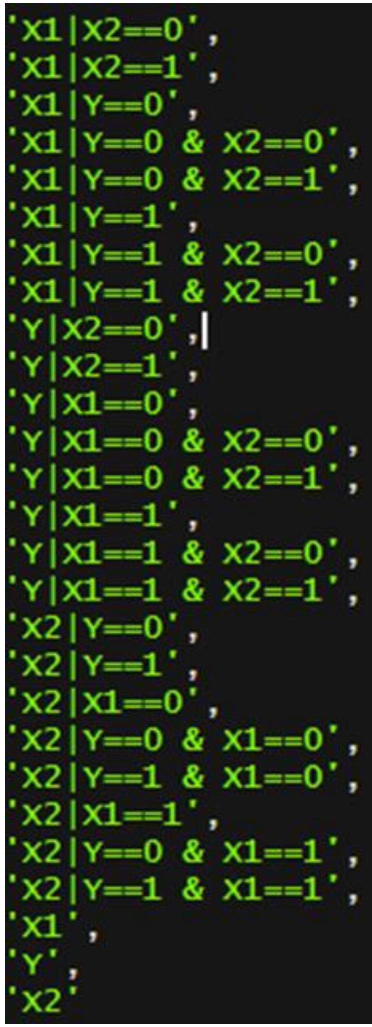


3 variable causal networks



Causal reasoning from known structure





```
'X1|X2==0',
'X1|X2==1',
'X1|Y==0',
'X1|Y==0 & X2==0',
'X1|Y==0 & X2==1',
'X1|Y==1',
'X1|Y==1 & X2==0',
'X1|Y==1 & X2==1',
'Y|X2==0',
'Y|X2==1',
'Y|X1==0',
'Y|X1==0 & X2==0',
'Y|X1==0 & X2==1',
'Y|X1==1',
'Y|X1==1 & X2==0',
'Y|X1==1 & X2==1',
'X2|Y==0',
'X2|Y==1',
'X2|X1==0',
'X2|Y==0 & X1==0',
'X2|Y==1 & X1==0',
'X2|X1==1',
'X2|Y==0 & X1==1',
'X2|Y==1 & X1==1',
'X1',
'Y',
'X2'
```

Causal reasoning by sampling

- We sample from memory or probabilistic generative models, base inferences or judgments on frequencies in obtained samples
- This sampling is the process of <u>thinking of concrete cases</u>, we generate a single chain of concrete cases and base our response on that.

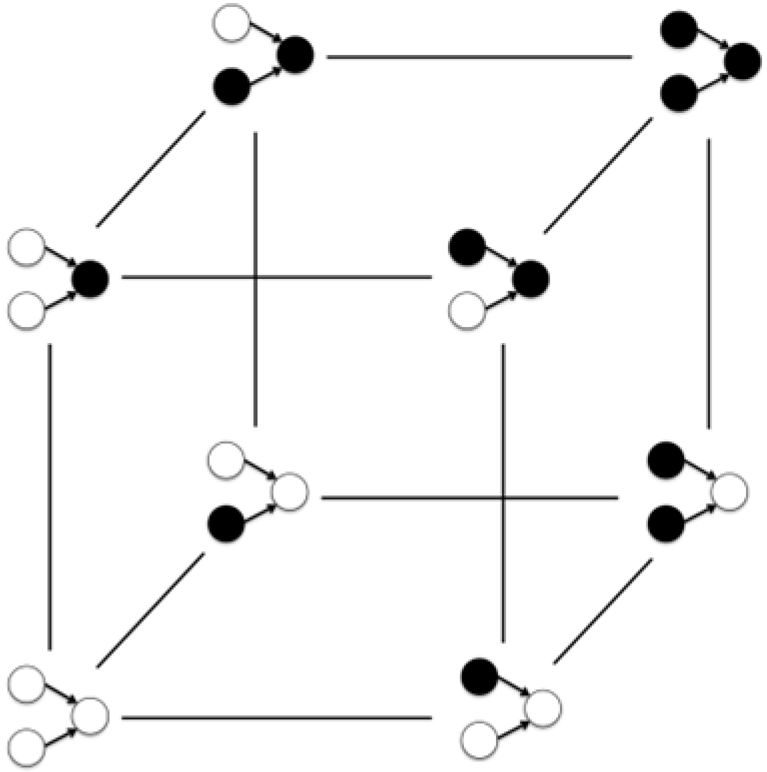
The Mutation Sampler: MH Sampling of concrete cases

- Metropolis-Hastings MCMC sampling over discrete states of causal graph
- MH: construct sequence of samples, where sample n depends on sample n-1.

  - Transition probability $a(q'|q)=\min(1, \frac{\pi(q')}{\pi(q)})$

  Where q is current state, q' is proposal state, π() is joint probability of the state.

  Only relative probability of two states is required, not full joint distribution.



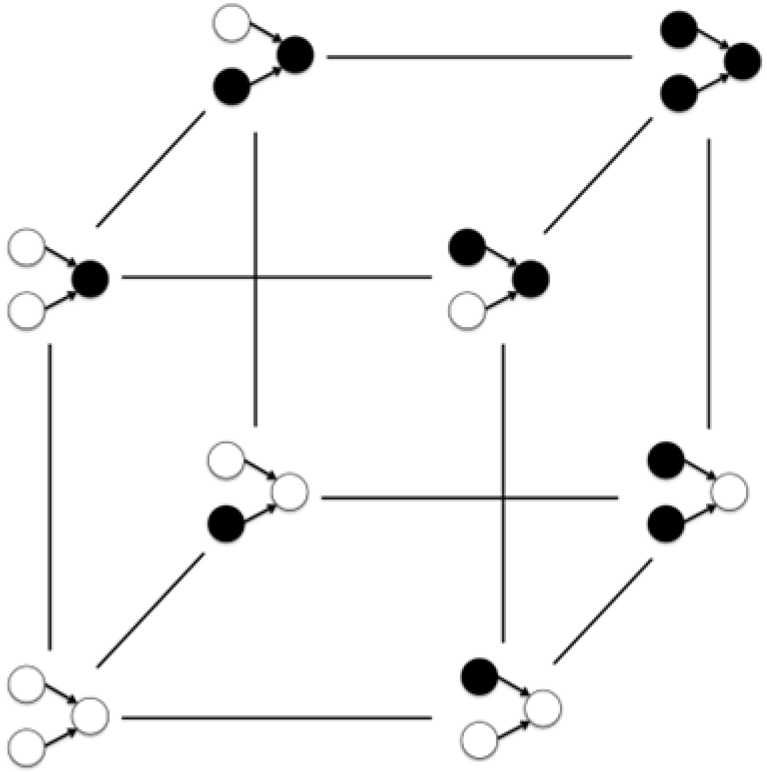
The Mutation Sampler: Proposal distribution

- Which network/graph state should be proposed as next state in chain: Mutation!
- Potential proposals are those network states which differ in the value of only one variable. Selected with equal probability.

- $\frac{\pi(q')}{\pi(q)}$ simplifies to

$$\frac{\pi(v_i', v_{-i})}{\pi(v_i, v_{-i})} = \frac{\pi(v_i'|v_{-i})\pi(v_{-i})}{\pi(v_i|v_{-i})\pi(v_{-i})} = \frac{\pi(v_i'|v_{-i})}{\pi(v_i|v_{-i})} = \frac{\pi(v_i'|u_i)}{\pi(v_i|u_i)}$$

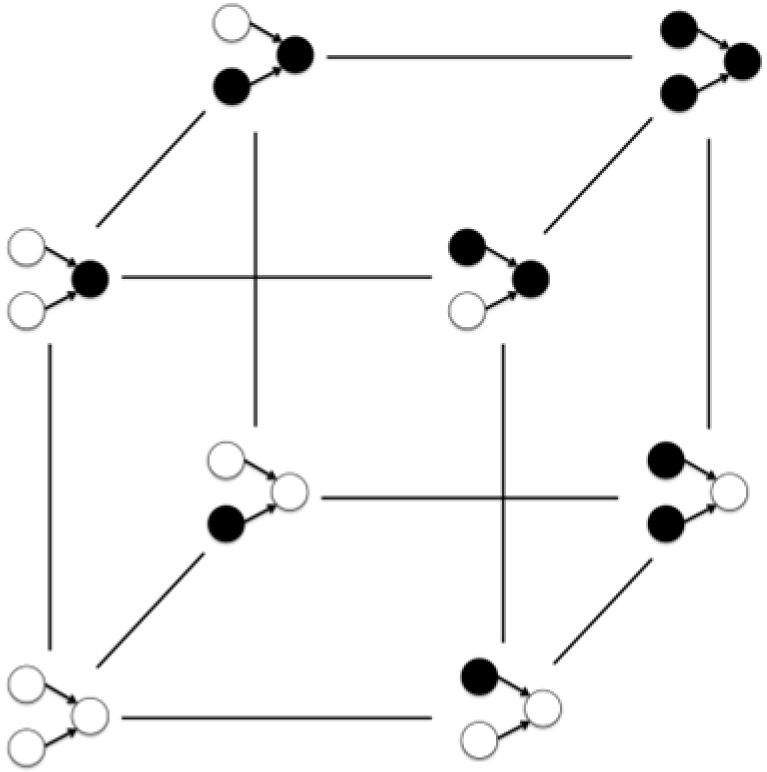
Where vi is value mutated variable, ui are variables in vi's Markov blanket

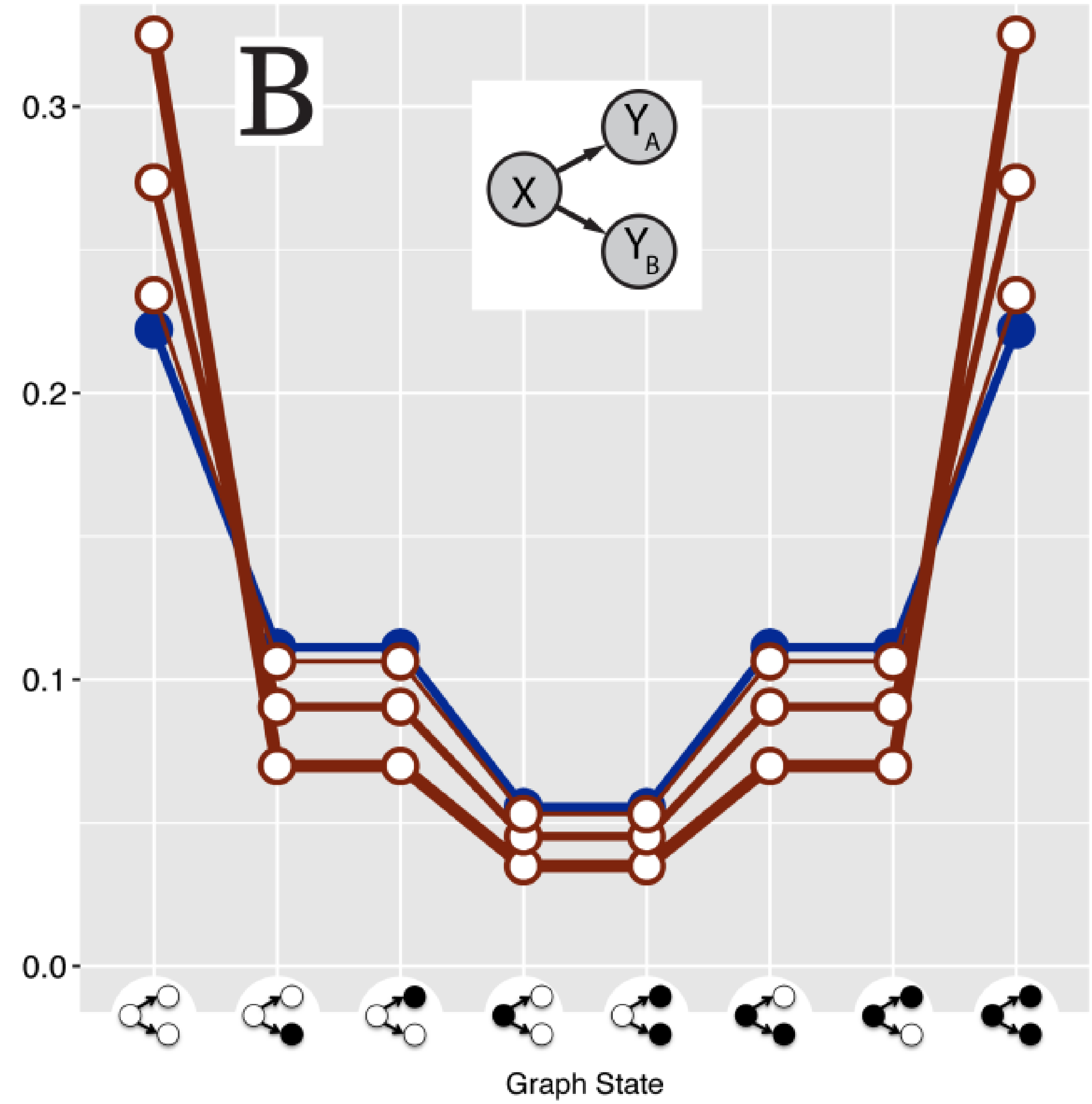


The Mutation Sampler: Biased starting points and limited capacity

- **Biased starting points**: sampling starts at prototypical states.
- **Limited capacity**: fixed capacity, but vary in number of samples taken for a judgement.

- Together, MS approximates joint probability of causal graph states if n samples grows large





B
X
Y_A
Y_B
0.3
0.2
0.1
0.0
Graph State

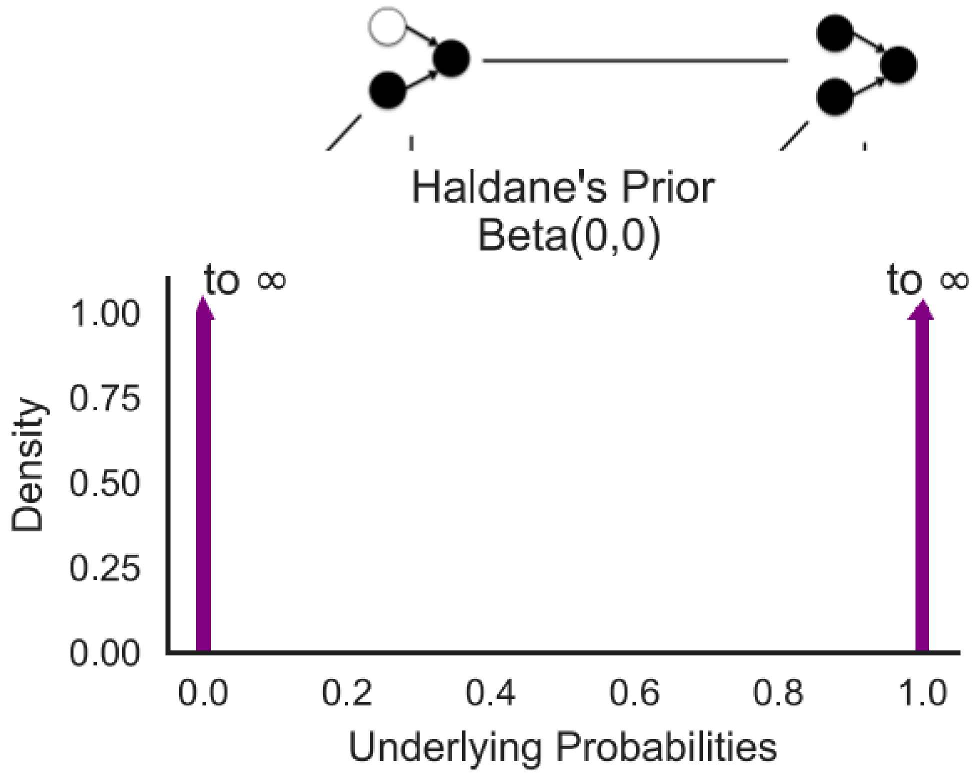
The Mutation Sampler: Inference

1. Calculate relative frequency of queried variable in samples

2. (if conditional query, calculate relative frequency within samples where antecedent is true)

Ex: $p(\mathrm{x} = 1|\mathrm{y} = 1) = \frac{freq(x=1,y=1)}{freq(x=0,y=1)+freq(x=1,y=1)}$

3. Guess 50% if required states are not sampled (i.e. no samples where antecedent is true)



* From Zhu, Sanborn, Chater (2020) used without permission

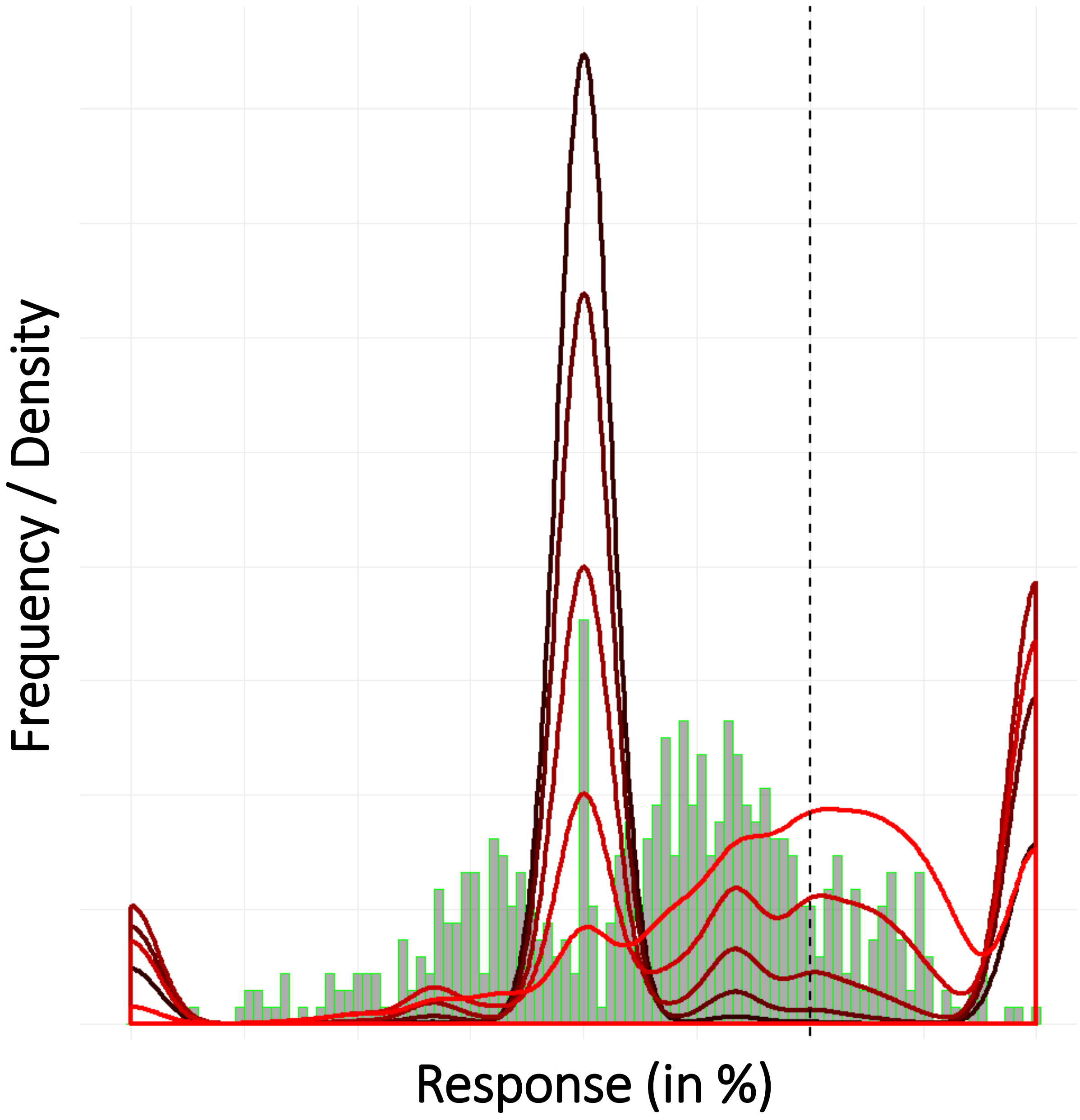
They tested the model: works great

- Fitted it to many causal reasoning studies, it performed better than all other models.
- Moreover, it explained 3 big non-normative patterns in causal reasoning data:
  - Markov violations
  - Failures to explain away
  - Conservative inferences

- So I thought: great! I want to play around with this model.

They tested the model: works great

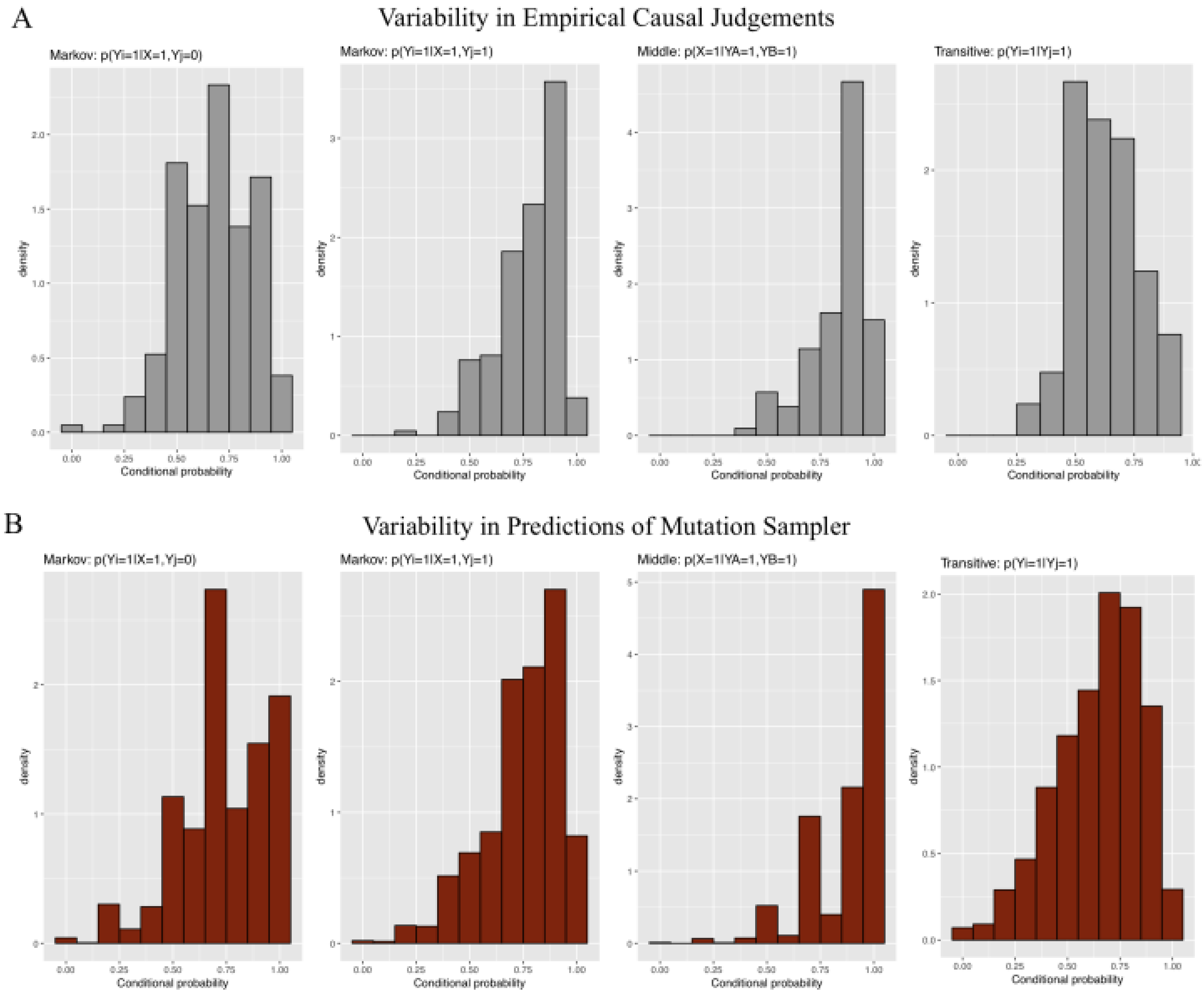
- Fitted it to many causal reasoning studies, it performed better than all other models.
- Moreover, it explained 3 big non-normative patterns in causal reasoning data:
  - Markov violations
  - Failures to explain away
  - **Conservative inferences**
- So I thought: great! Let’s play around with this model.



Participant data

Model predictions
darker = fewer samples
(4, 8, 16, 32, 64)

Black dashed line is
normative answer


A
Variability in Empirical Causal Judgements
Markov: p(Yi=1|X=1,Yj=0)
Markov: p(Yi=1|X=1,Yj=1)
Middle: p(X=1|YA=1,YB=1)
Transitive: p(Yi=1|Yj=1)
density
Conditional probability
B
Variability in Predictions of Mutation Sampler
Markov: p(Yi=1|X=1,Yj=0)
Markov: p(Yi=1|X=1,Yj=1)
Middle: p(X=1|YA=1,YB=1)
Transitive: p(Yi=1|Yj=1)
density
Conditional probability

Can we solve these issues of the mutation sampler?

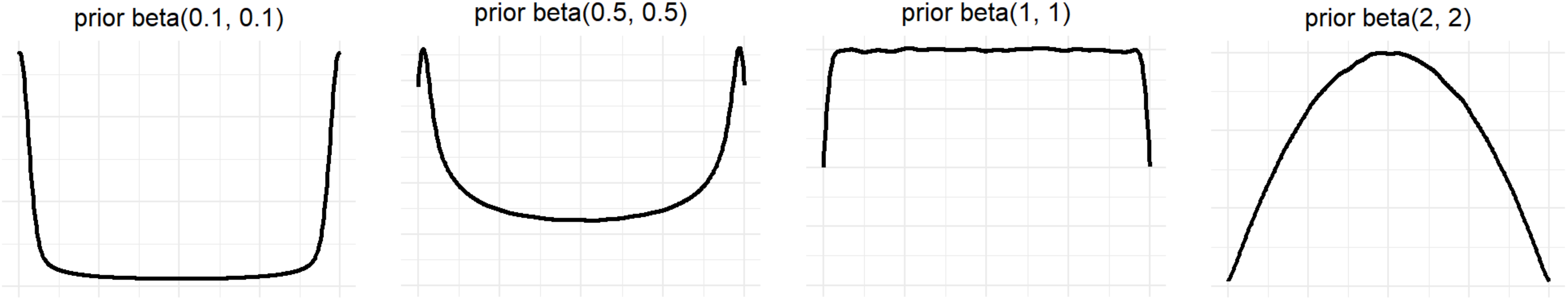
- The model should predict conservative inferences, but not by peaks at 50%.
- The model should not predict extreme responses at 0 and 100%

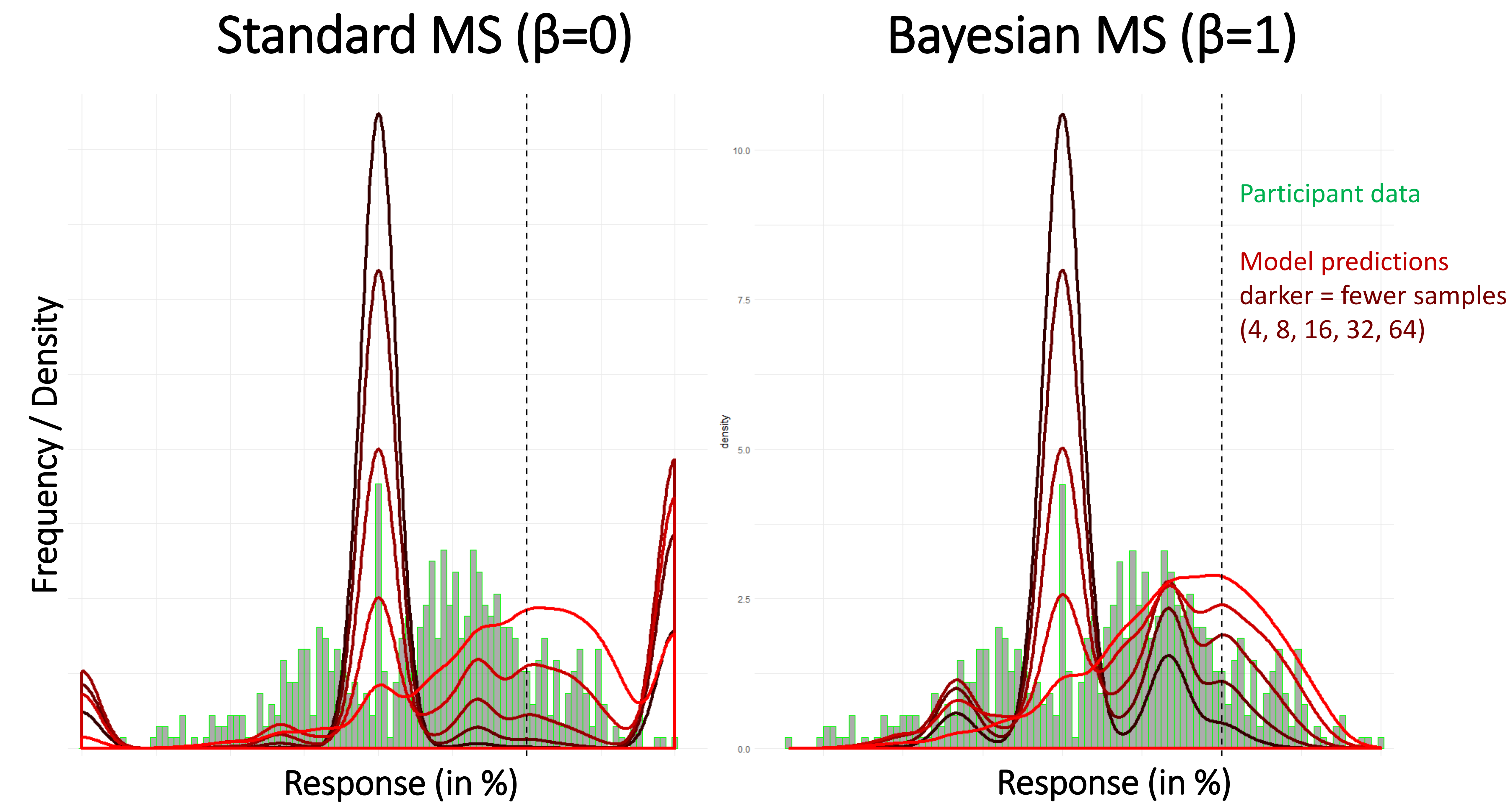
- Those conservative inferences, maybe they are not a result of an fascination with 50% or a rounding habit, maybe it is due to a ‘rational adjustment for small sample sizes’?

All this points toward using a (generic) prior

- Prior dominates with small sample sizes: no extreme responses
- Prior pushes judgments towards 50%: conservative inferences

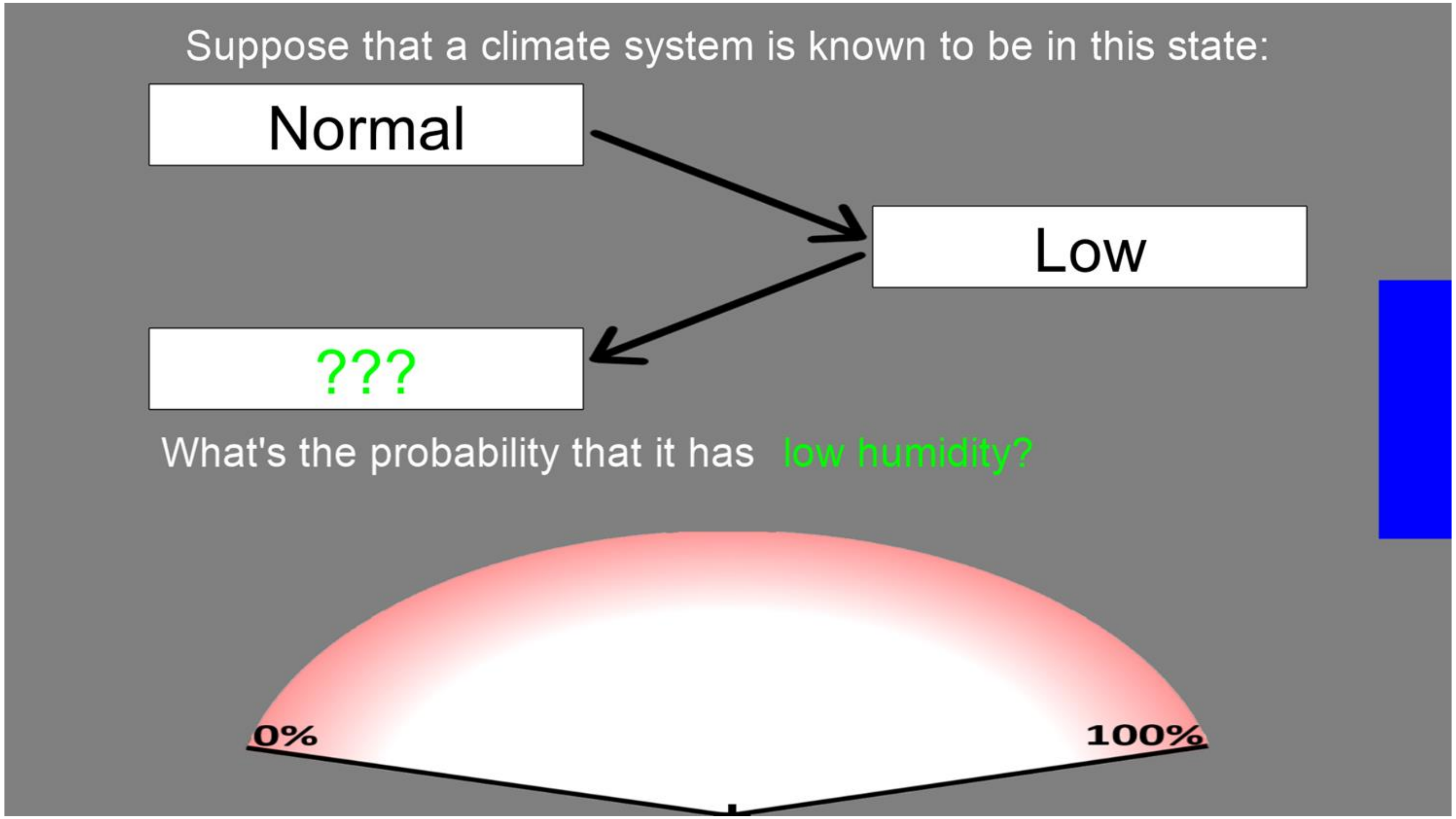
- Adapt the Mutation Sampler into a Bayesian Mutation Sampler by using symmetric beta priors: incorporate pseudocount β, response is expected value

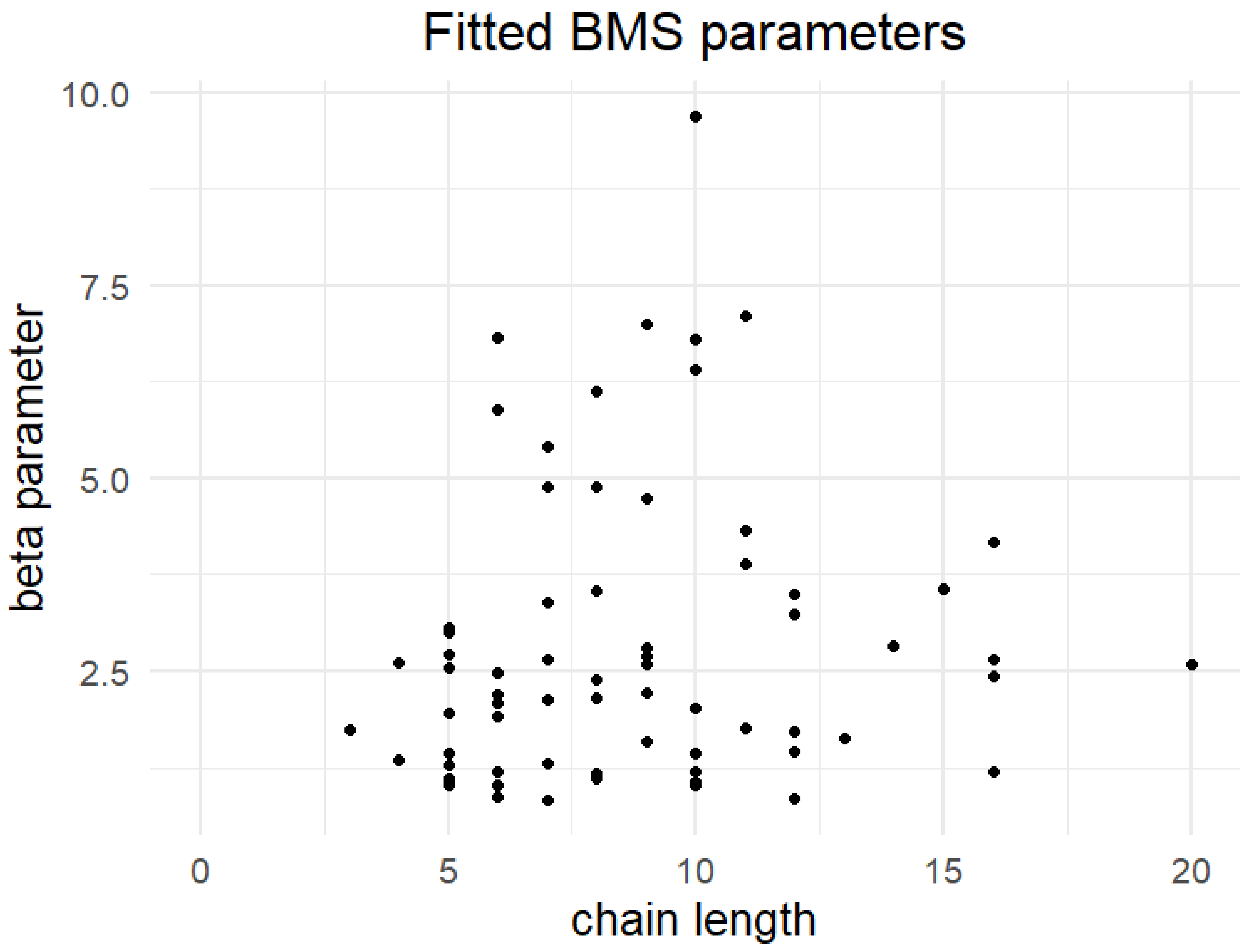




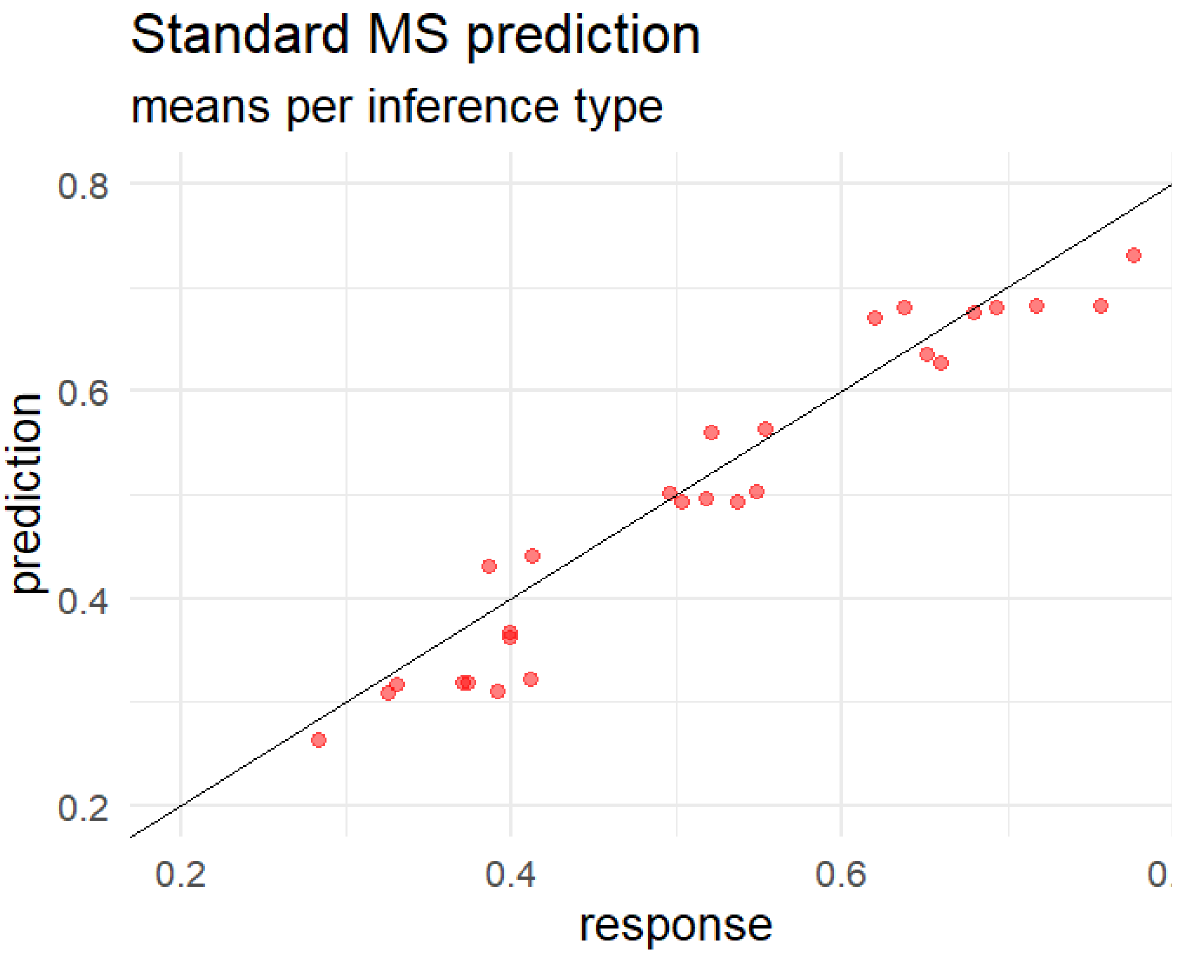
Standard MS (β=0)
Bayesian MS (β=1)
Frequency / Density
Response (in %)
Response (in %)
density
10.0
7.5
5.0
2.5
0.0
Participant data
Model predictions
darker = fewer samples
(4, 8, 16, 32, 64)

- Experiment using joystick to respond
- Fit using Probability Density Approximation (PDA) method (Turner & Sederberg, 2014)

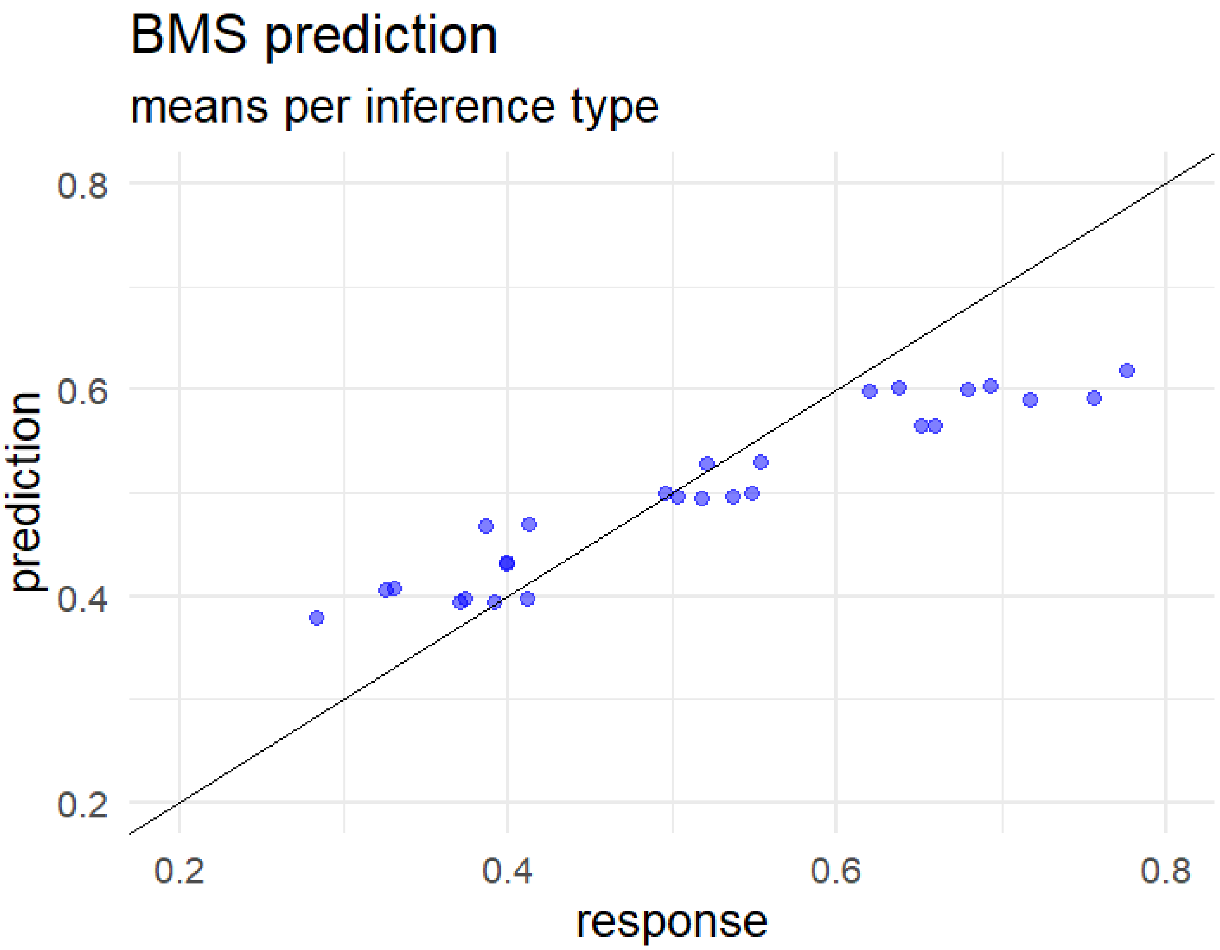



Fitted BMS parameters
beta parameter
10.0
7.5
5.0
2.5
0
5
10
15
20
chain length

X_1
Y
X_2

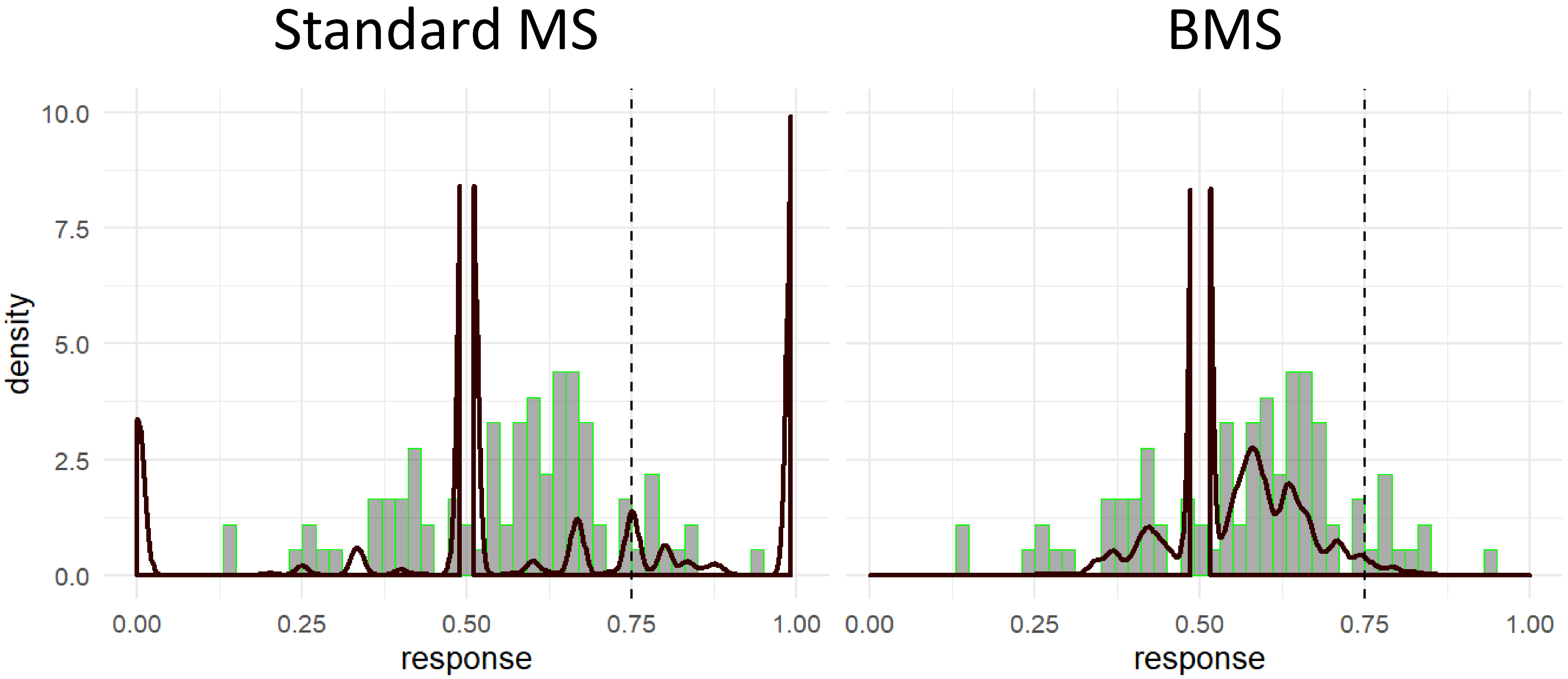


Standard MS prediction
means per inference type
prediction
response
0.2
0.4
0.6
0.8

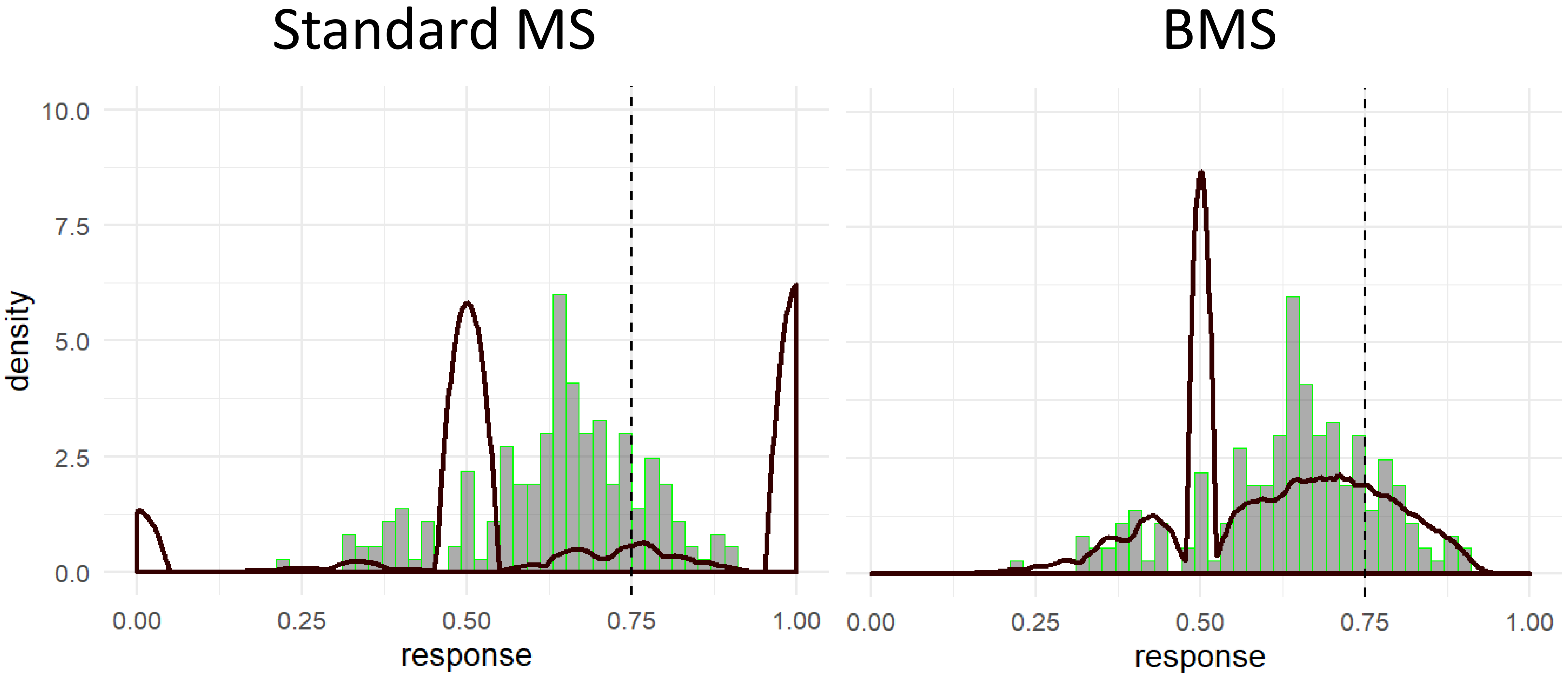


BMS prediction
means per inference type
prediction
response
0.2
0.4
0.6
0.8

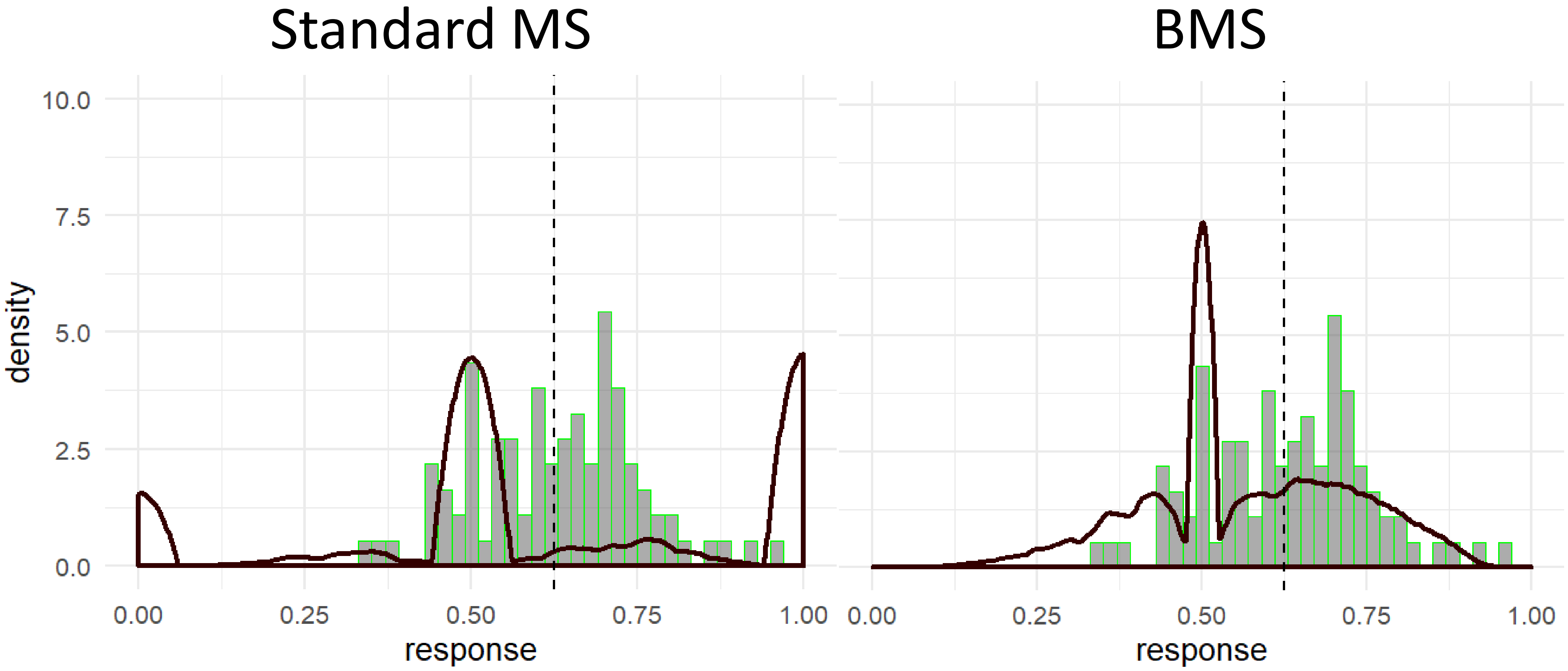
Conflict trials, e.g. p(x1|y=1,x2=0)

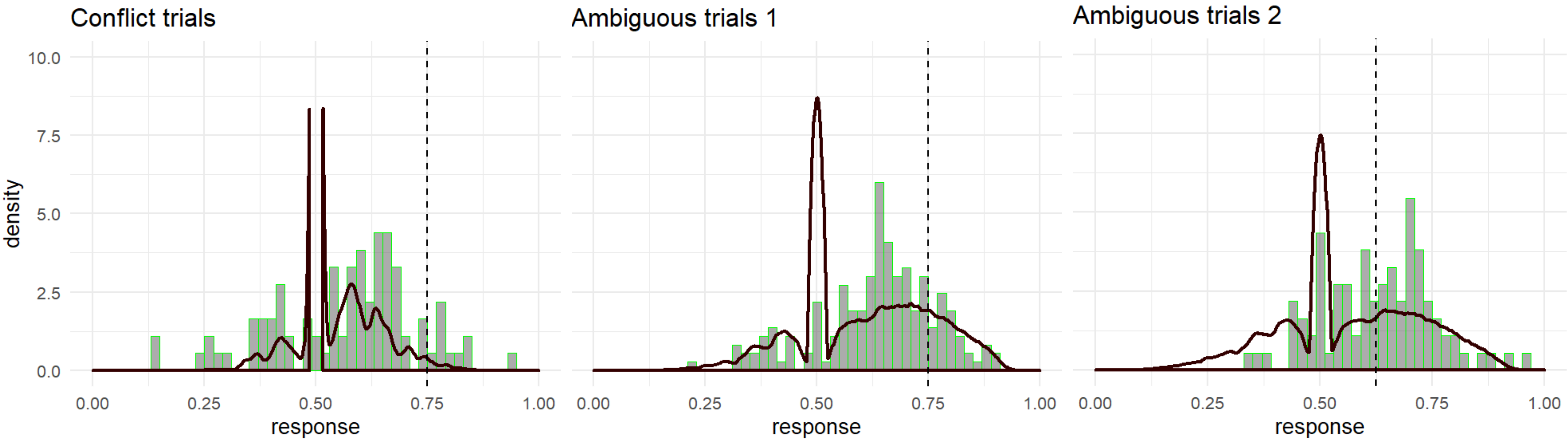


Ambiguous trials 1, e.g. p(y|x1=0)



Ambiguous trials 2, e.g. p(x1|x2=1)





- BMS predicted distributions promising, except for peak at 50%

- Theoretical advantages
  - Information on the amount of samples taken is actually used by agent
  - Better matches phenomenology of making probability estimates

The end

- References

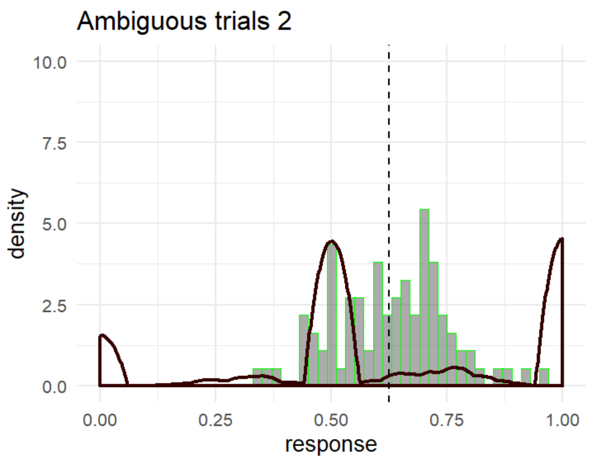
  - Davis, Z. J., & Rehder, B. (2020). A Process Model of Causal Reasoning. *Cognitive Science*, *44*(5), e12839.
  - Zhu, J. Q., Sanborn, A. N., & Chater, N. (2020). The Bayesian sampler: Generic Bayesian inference causes incoherence in human probability judgments. *Psychological Review.* Advance online publication.
  - Turner, B. M., & Sederberg, P. B. (2014). A generalized, likelihood-free method for posterior estimation. *Psychonomic bulletin & review*, *21*(2), 227-250.

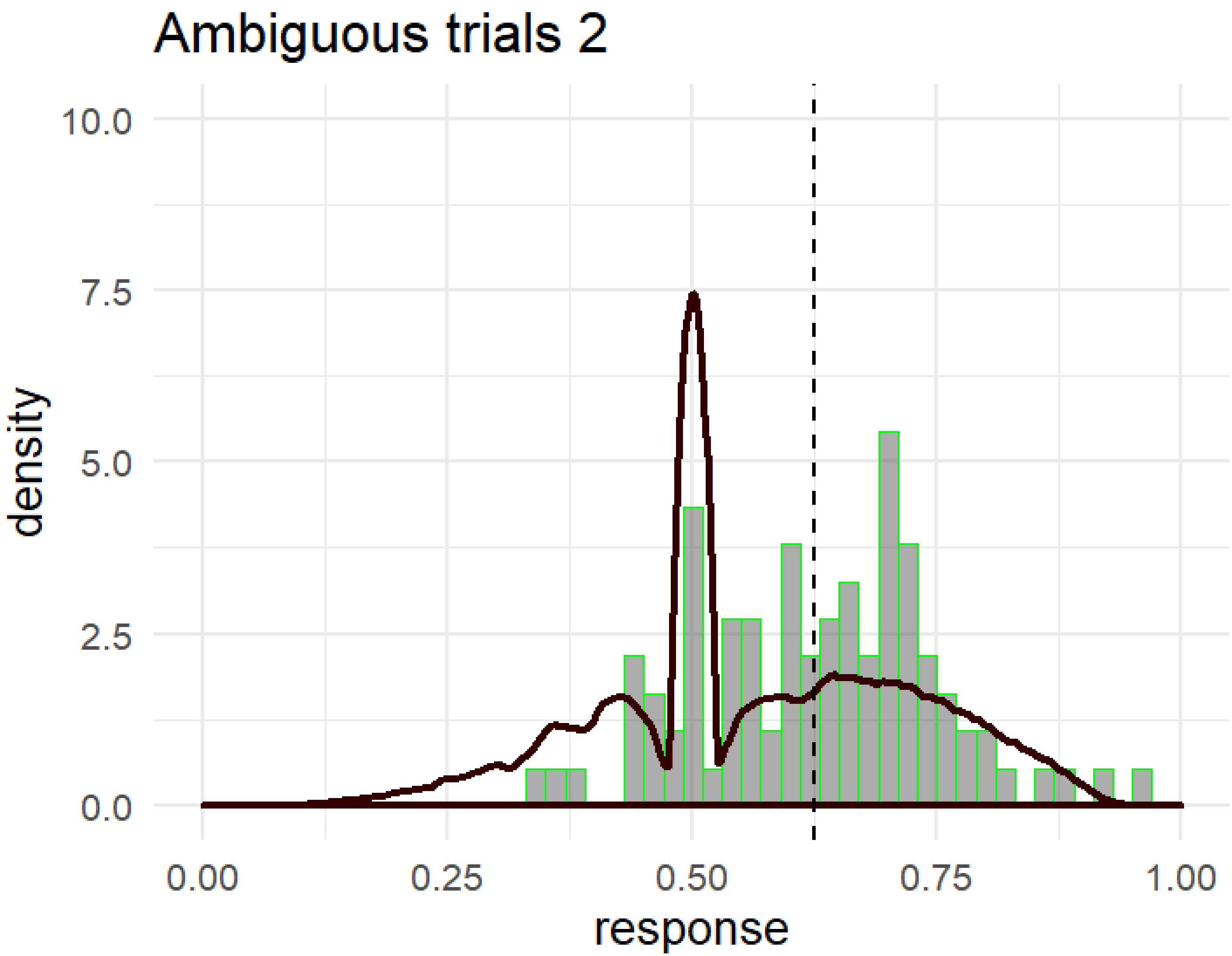
- PhD supervision
  - Leendert van Maanen (Experimental psychology, Utrecht University)
  - Robert van Rooij (ILLC, University of Amsterdam)
  - Katrin Schulz (ILLC, University of Amsterdam)

Current and future work

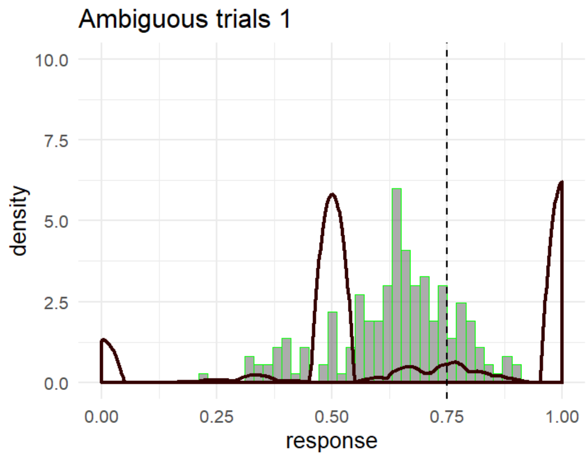
- Modelling:
  - Starting a queried state
  - Only sample states consistent with conditional
  - Amortized inference
  - Model guessing as distribution
- When does one stop sampling? When one is certain of your answer? Study currently running.
- Inter or intra-subject variability? Larger study on variability in causal inference with Zach Davis and Bob Rehder (NYU)

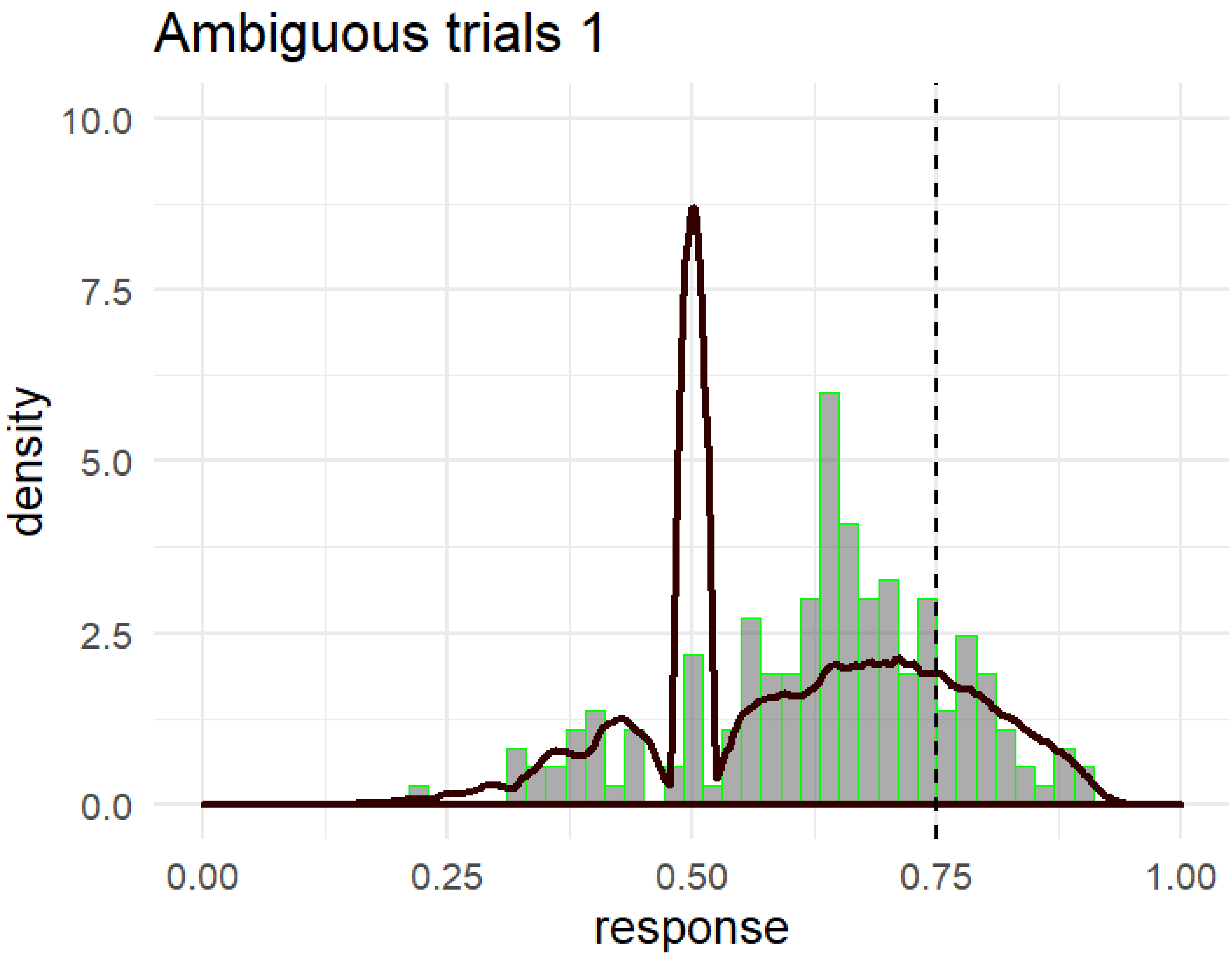
- Ambiguous trials 2
- Inference conditioning on 1 variable, i.e. the third variable is unknown
- E.g. p(y|x2=1)



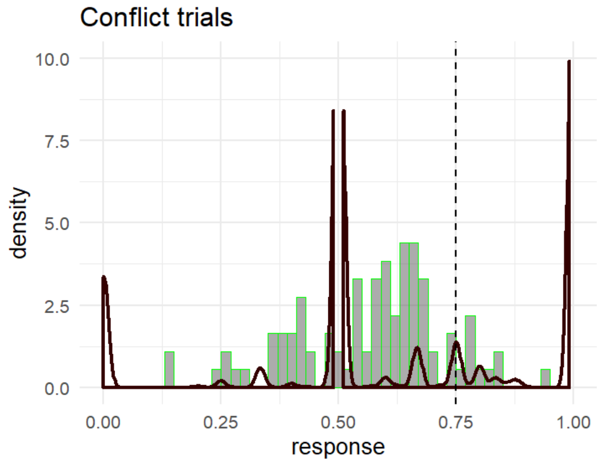


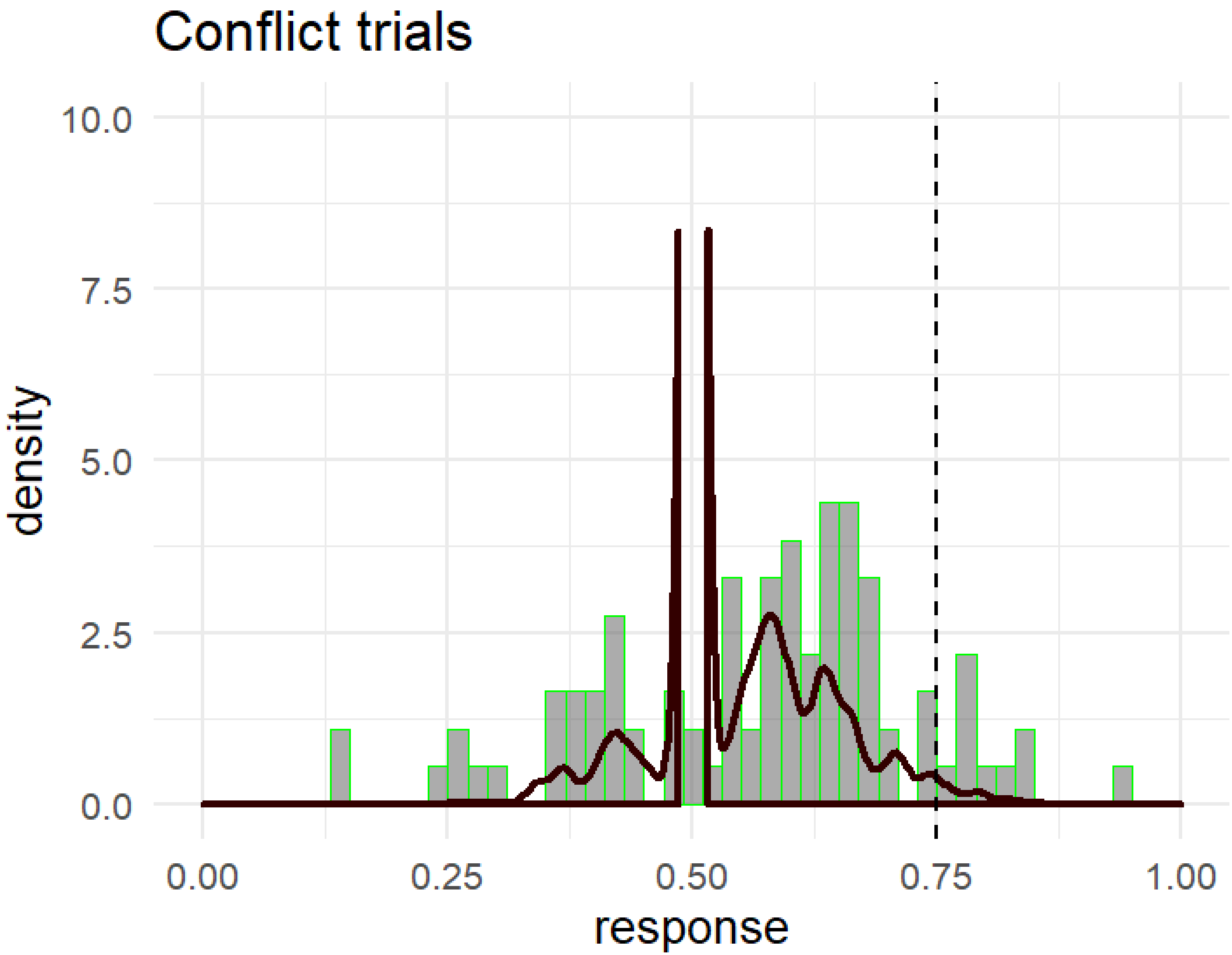
- Ambiguous trials 1
- Inference conditioning on 1 variable, i.e. the third variable is unknown
- E.g. p(y|x2=1)

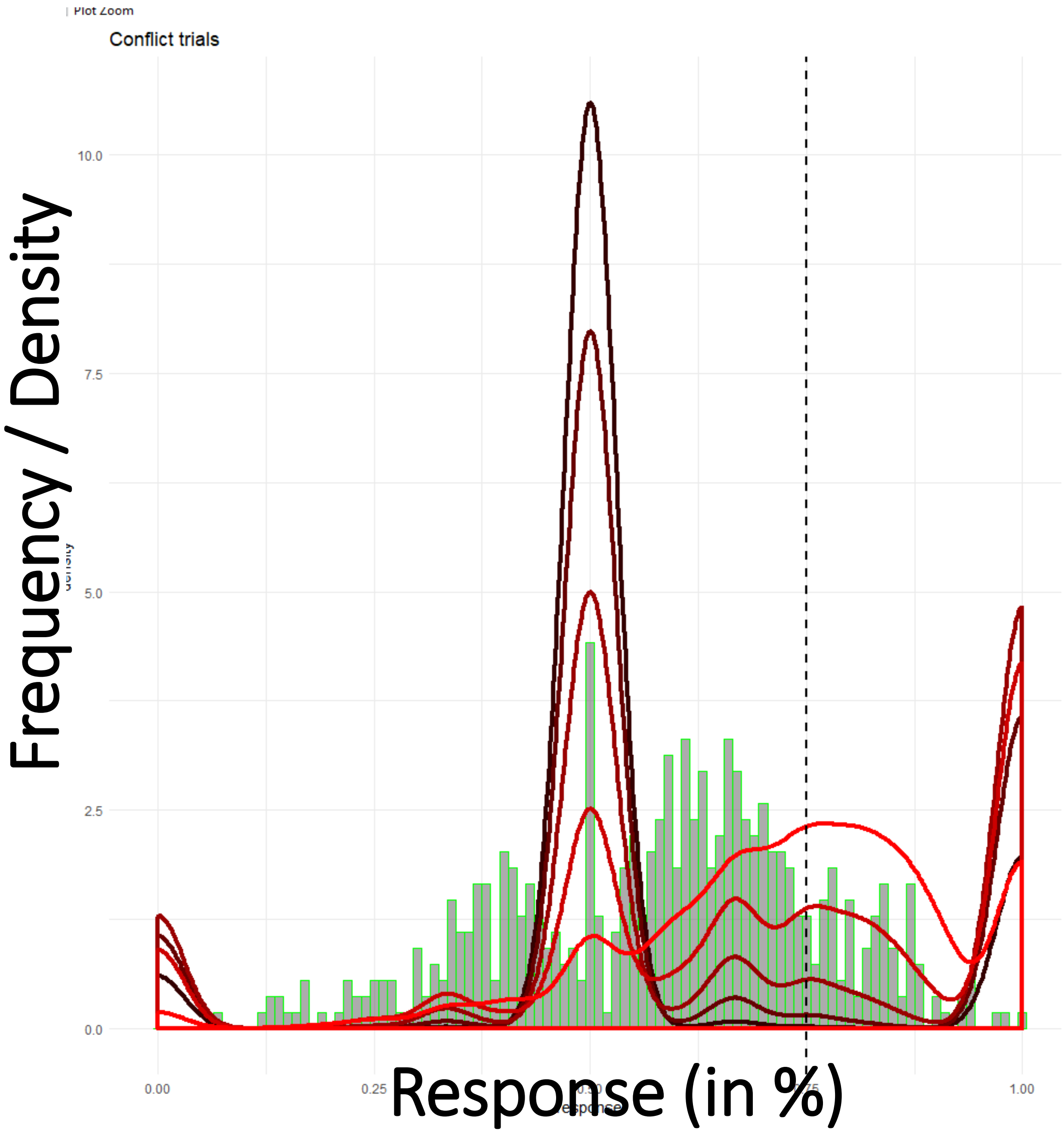




- Conflict trials
- Inference conditioning on two variable states that are 'inconsistent'
- E.g. $p(y|x1=1,x2=0)$



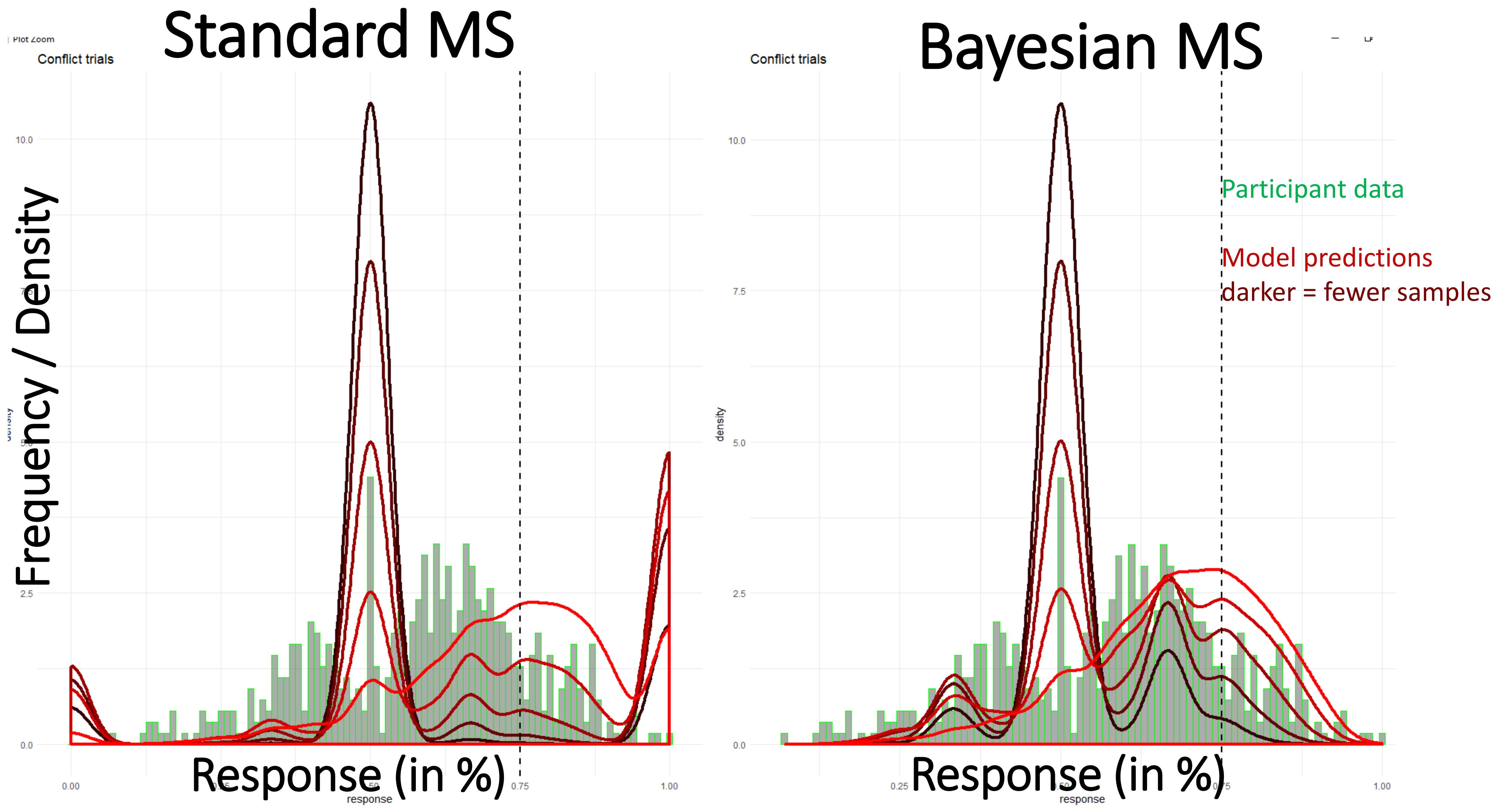




Participant data

Model predictions
darker = fewer samples
(4, 8, 16, 32, 64)

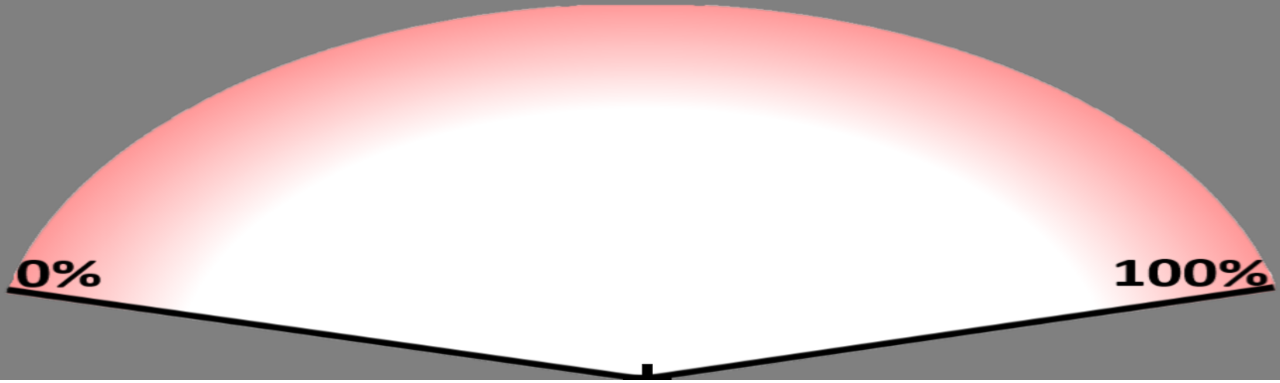
Black dashed line is
normative answer

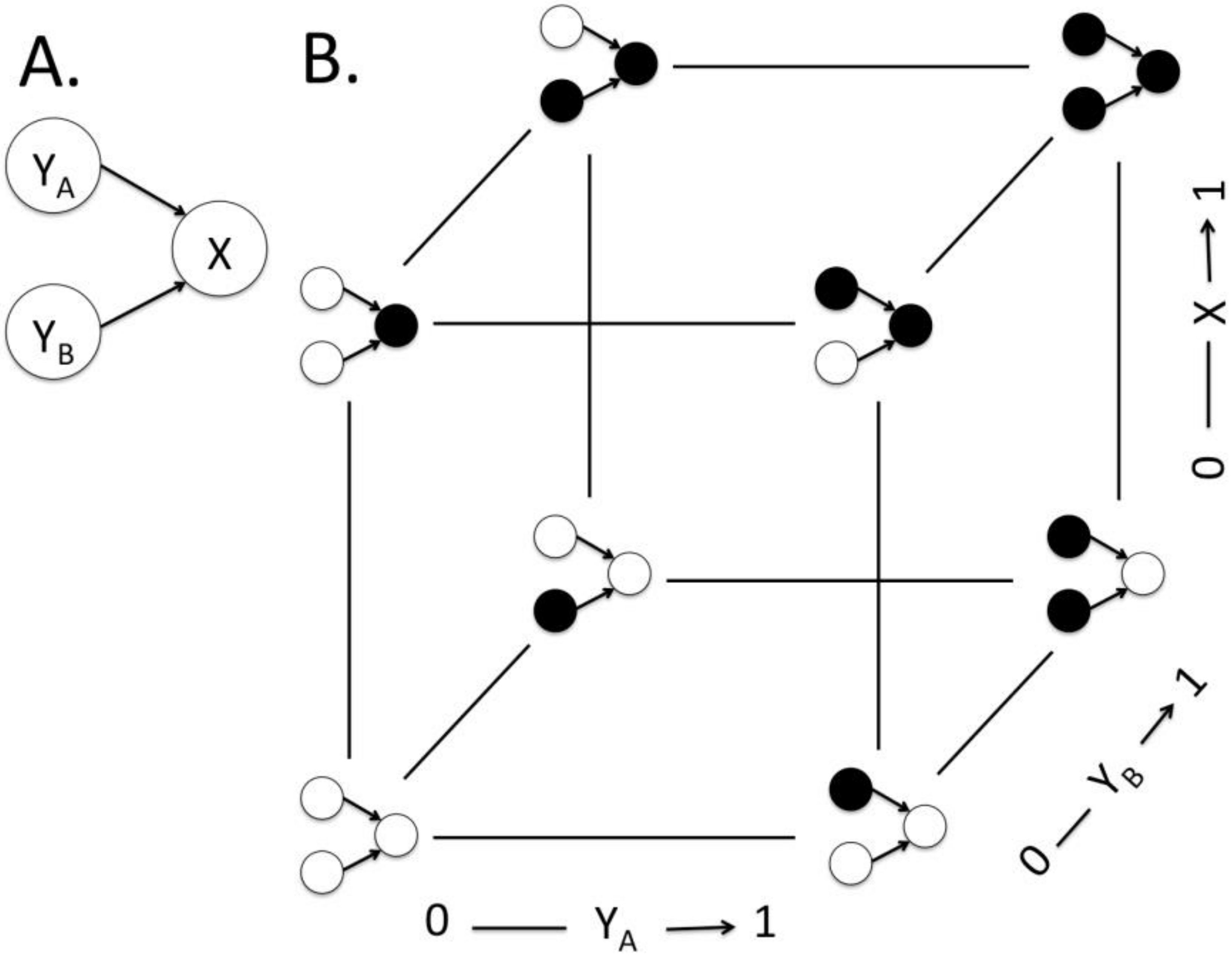

Standard MS
Bayesian MS
Plot Zoom
Conflict trials
Conflict trials
Frequency / Density
density
10.0
7.5
5.0
2.5
0.0
Participant data
Model predictions
darker = fewer samples
Response (in %)
Response (in %)
response
response
0.00
0.25
0.50
0.75
1.00

Suppose that a climate system is known to be in this state:

Normal → Low → ???

What's the probability that it has low humidity?





A.
B.
Y_A
Y_B
X
$0 — Y_A \rightarrow 1$
$0 — Y_B \rightarrow 1$
$0 — X \rightarrow 1$