Bayesian Inference (II)

Adrian Brasoveanu

[based on slides by Michael D. Lee & Eric-Jan Wagenmakers]

Winter 2012 · Seminar in Semantics · UCSC Linguistics

What is Bayesian Inference?

“Common sense expressed in numbers”

“The only statistical procedure that is *coherent*, meaning that it avoids statements that are internally inconsistent.”

“The only good statistics”

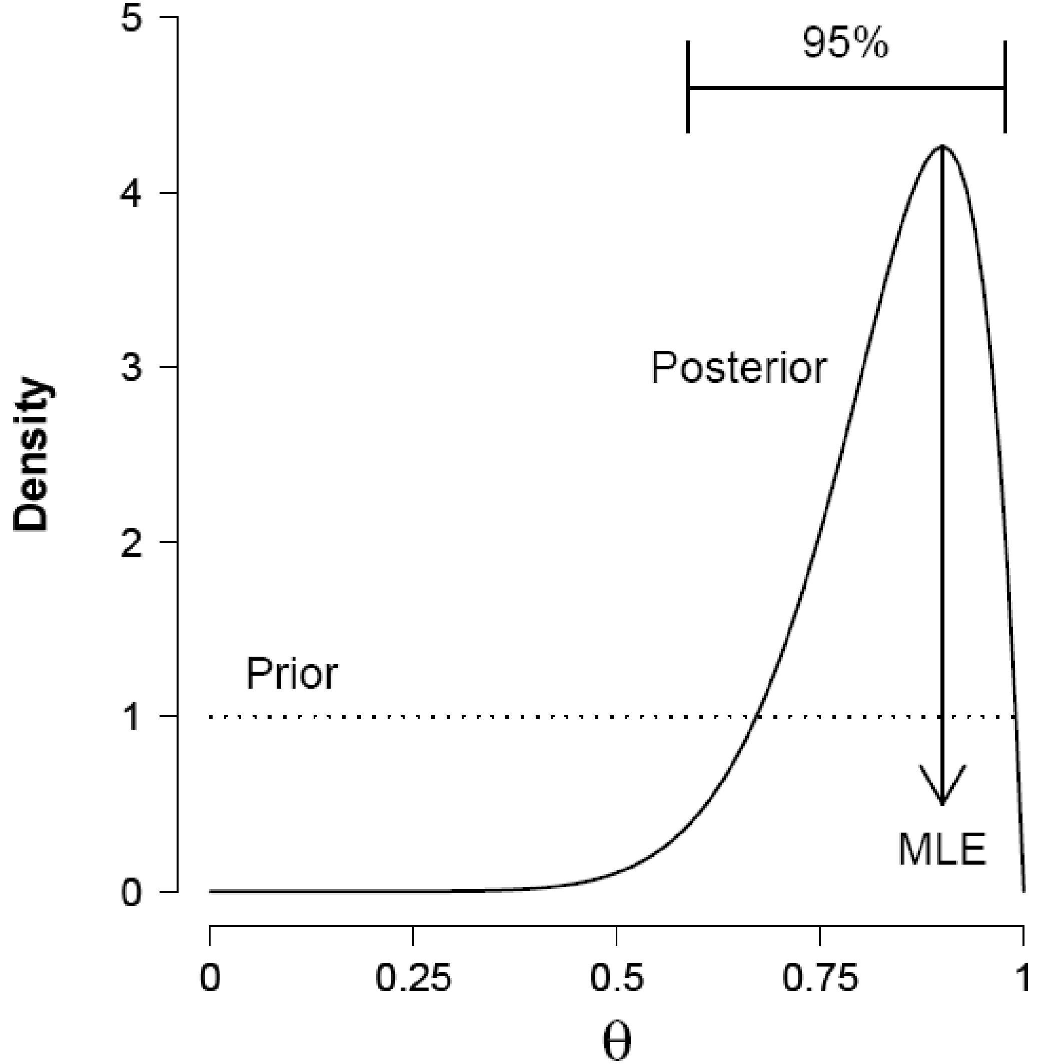
Bayesian Inference in a Nutshell

- In Bayesian inference, uncertainty or degree of belief is quantified by probability.
- Prior beliefs are updated by means of the data to yield posterior beliefs.

Bayesian Parameter Estimation: Example

- We prepare for you a series of 10 factual questions of equal difficulty.
- You answer 9 out of 10 questions correctly.
- What is your latent probability θ of answering any one question correctly?

- We start with a prior distribution for θ. This reflect all we know about θ prior to the experiment. Here we make a standard choice and assume that all values of θ are equally likely *a priori*.
- We then update the prior distribution by means of the data (technically, the *likelihood*) to arrive at a posterior distribution.
- The posterior distribution is a compromise between what we knew before the experiment and what we have learned from the experiment. The posterior distribution reflects all that we know about θ.

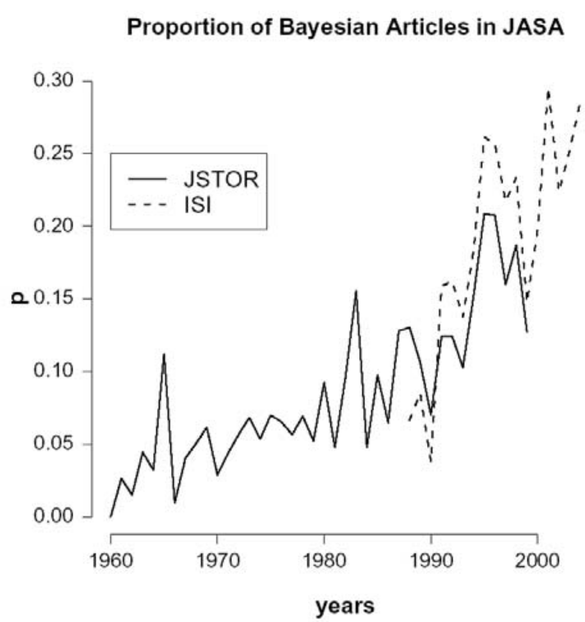


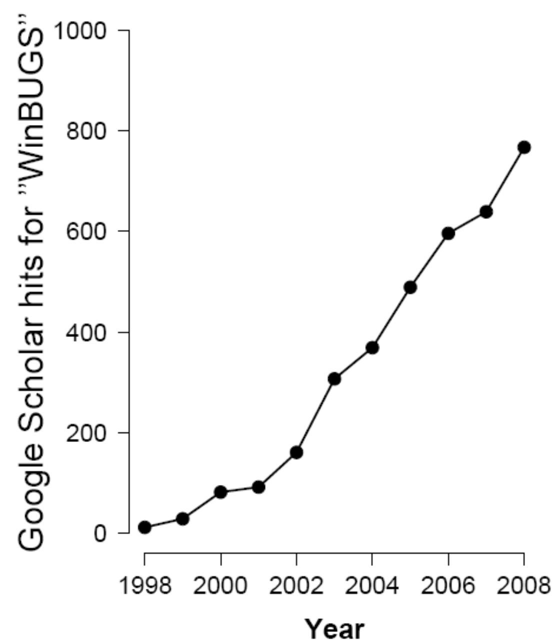
Mode = 0.9

95% confidence interval: (0.59, 0.98)

The Bayesian Revolution

- Until about 1990, Bayesian statistics could only be applied to a select subset of very simple models.
- Only recently, Bayesian statistics has undergone a transformation; with current numerical techniques, Bayesian models are "*limited only by the user's imagination*."

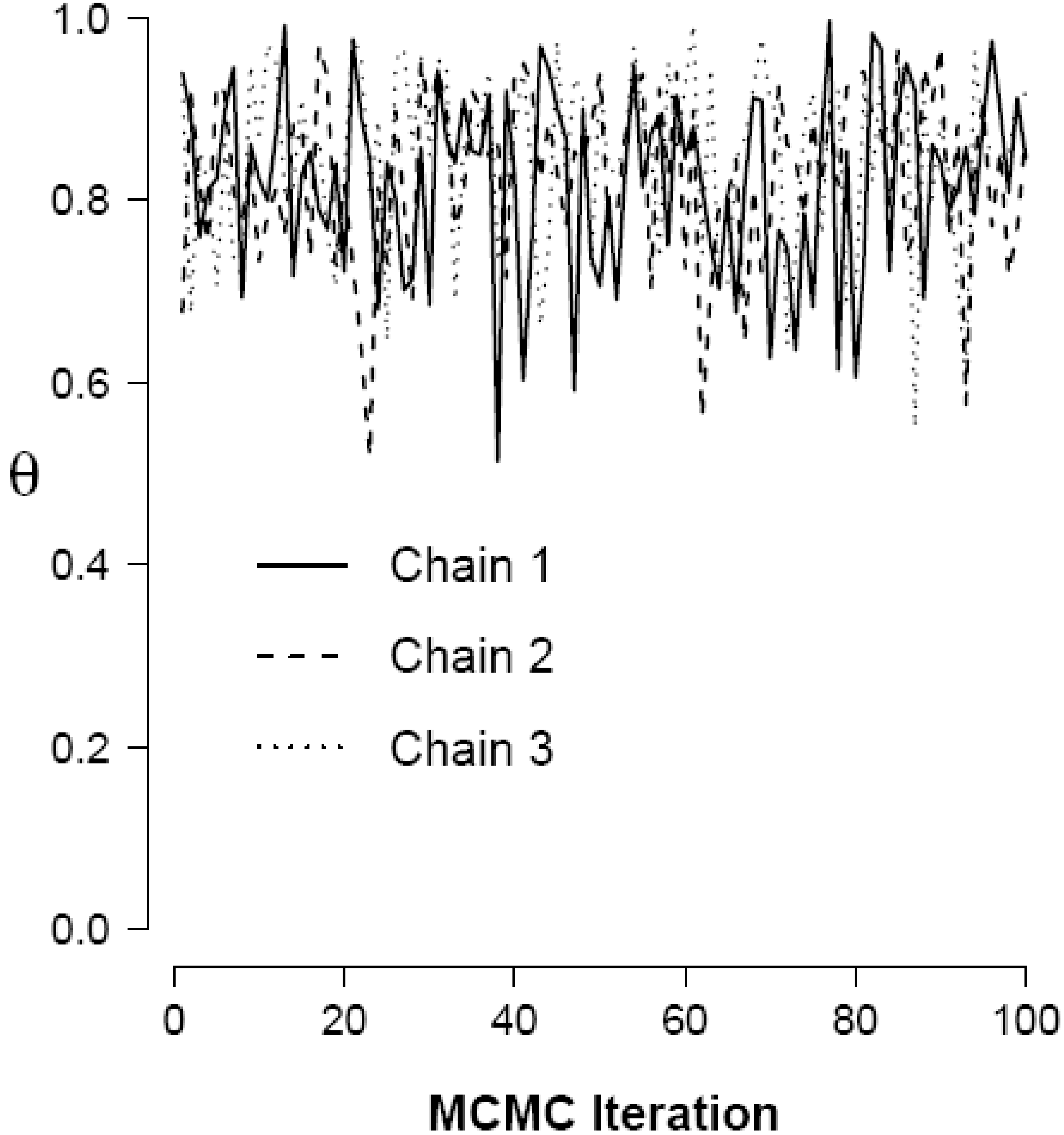


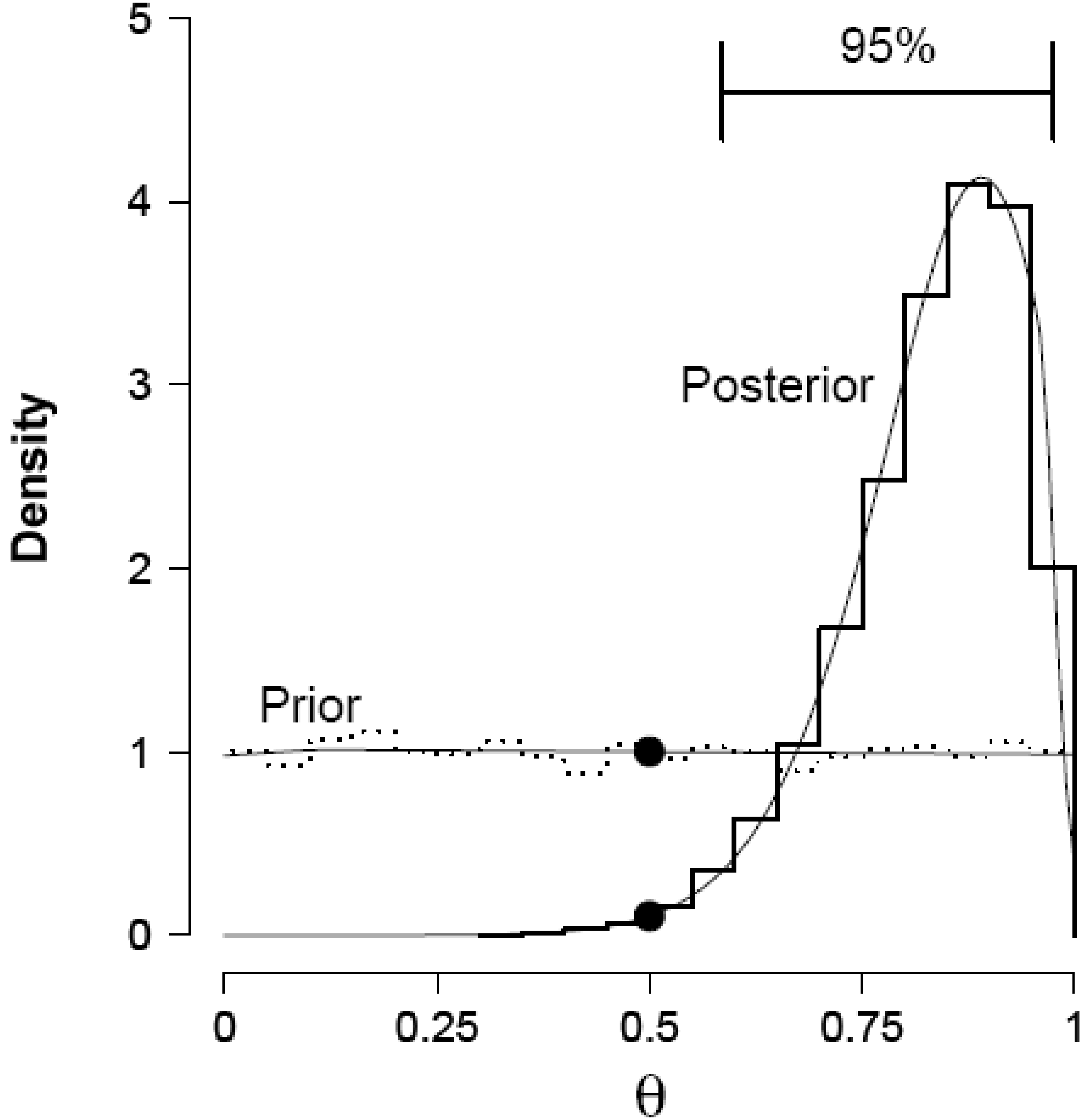


Why Bayes is Now Popular

Markov Chain Monte Carlo!

- Instead of calculating the posterior analytically, numerical techniques such as MCMC approximate the posterior by drawing samples from it.
- Consider again our earlier example…

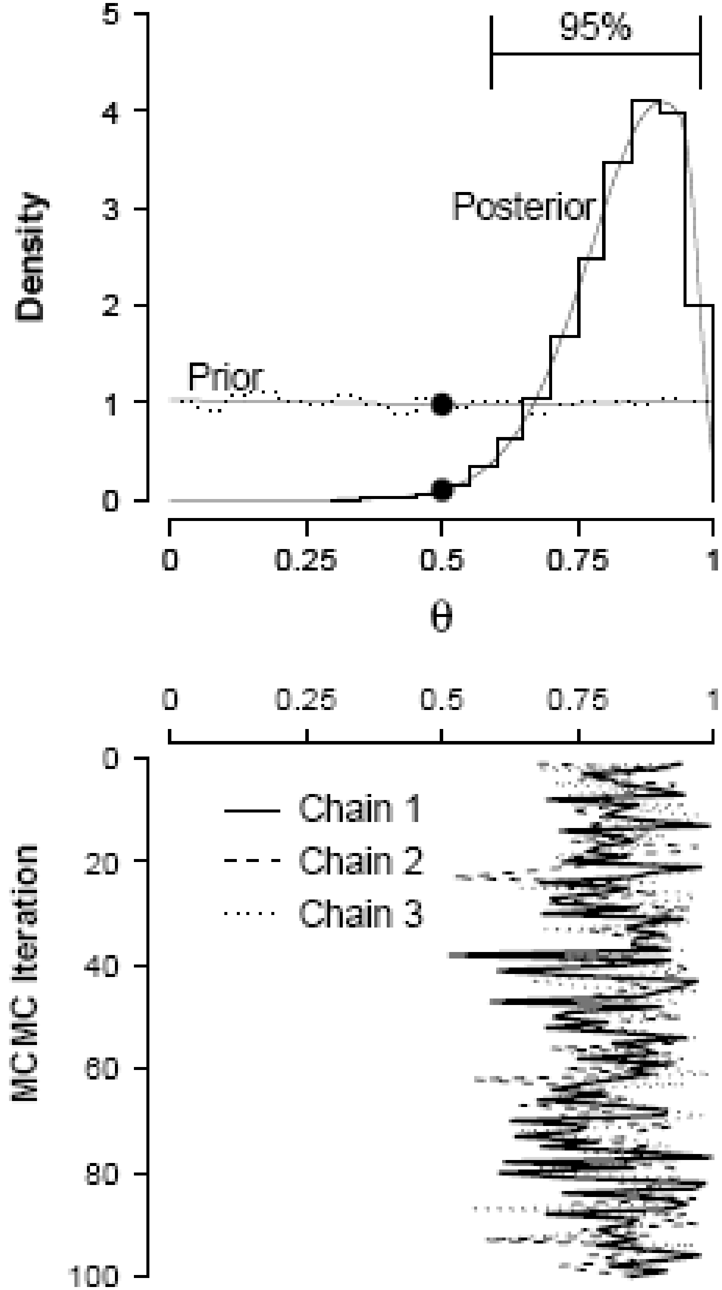

1.0
0.8
0.6
0.4
0.2
0.0
θ
Chain 1
Chain 2
Chain 3
0
20
40
60
80
100
MCMC Iteration



Mode = 0.89

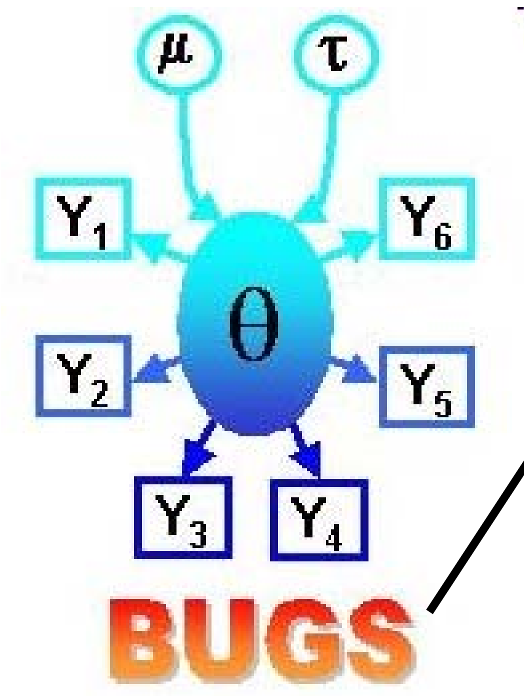
95% confidence interval: (0.59, 0.98)

With 9000 samples, almost identical to analytical result.


95%
Posterior
Prior
Density
0
1
2
3
4
5
0
0.25
0.5
0.75
1
θ
Chain 1
Chain 2
Chain 3
MCMC Iteration
0
20
40
60
80
100

MCMC

- With MCMC, the models you can build and estimate are said to be "limited only by the user's imagination".
- But how do you get MCMC to work?
  - Option 1: write the code it yourself.
  - Option 2: use WinBUGS!



Bayesian inference
Using
Gibbs Sampling

WinBUGS

- Knows many probability distributions (likelihoods), e.g., the binomial distribution, the Gaussian distribution, the Poisson distribution;
- These distributions form the elementary building blocks from which you may construct infinitely many models.
- Allows you to specify a model;
- Allows you to specify priors;
- Will then automatically run the MCMC sampling routines and produce output.

Inside every Non-Bayesian,
there is a Bayesian
struggling to get out

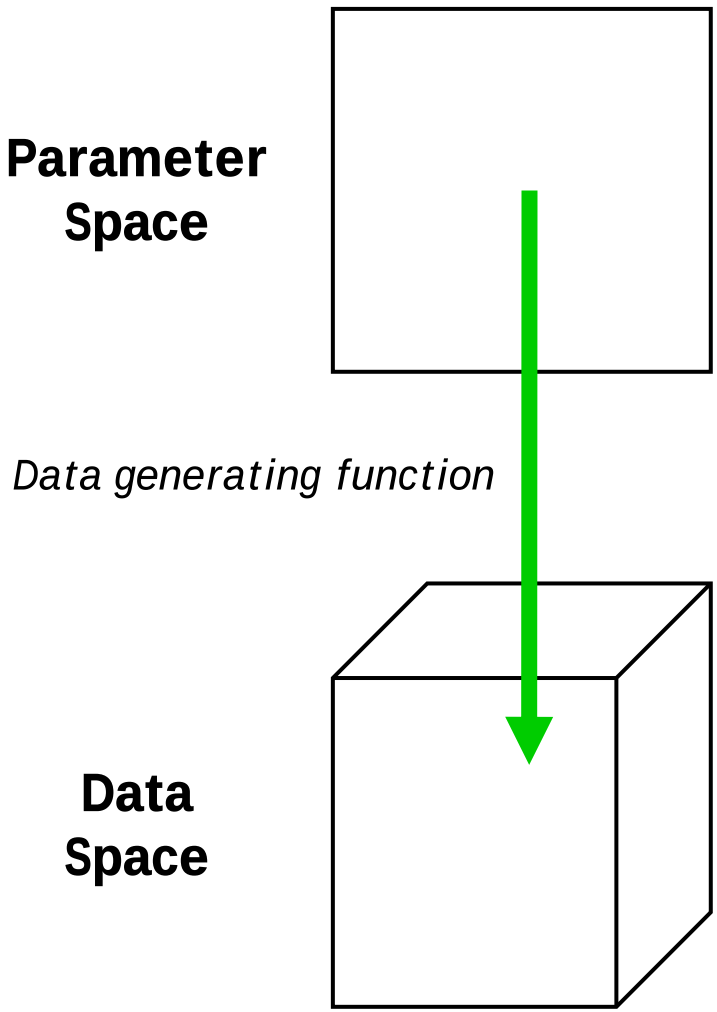
Dennis Lindley

Bayes for Cognitive Science

- Bayesian methods are becoming very important in the cognitive sciences
- Bayesian statistics is a framework for doing inference, in a principled way, based on probability theory
- Three types of application
  - **Bayes in the head:** Use Bayes as a theoretical metaphor, assuming when people make inferences they apply (at some level) Bayesian methods (Tenenbaum, Griffiths, Yuille, Chater, Kemp, …)
  - **Bayes for data analysis:** Instead of using frequentist estimation, null hypothesis testing, and so on, use Bayesian inference to analyze data (Kruschke)
  - **Bayes for modeling**: Use Bayesian inference to relate models of psychological processes to behavioral data

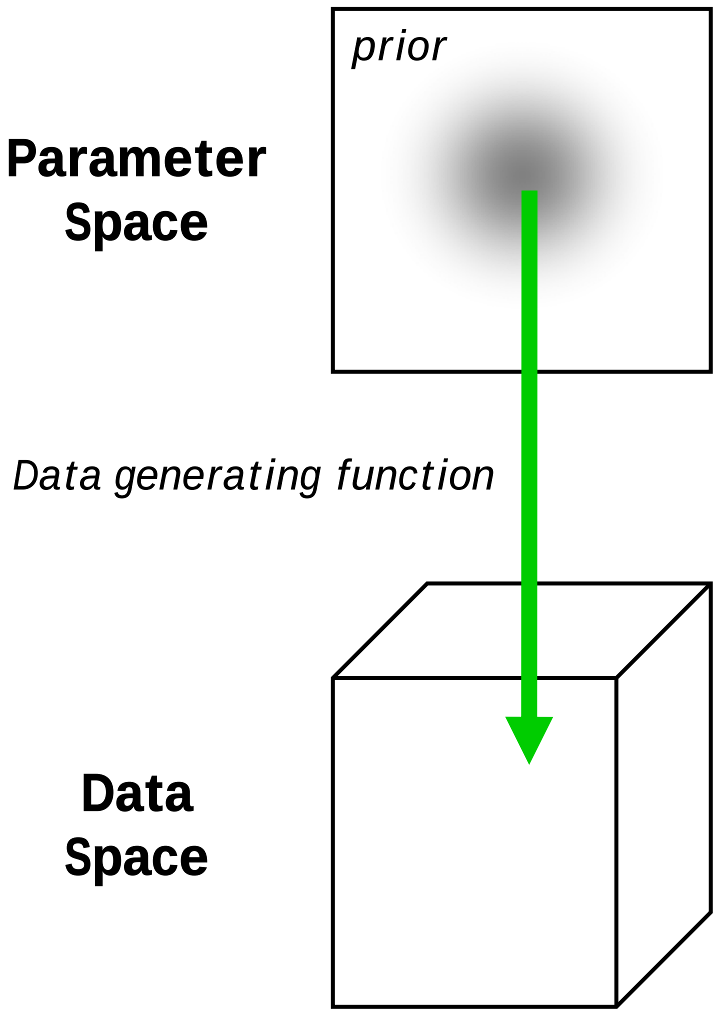
Psychological Models in Bayesian Framework

- Psychological models can be thought of as generative statistical processes, mapping latent parameters to observed data



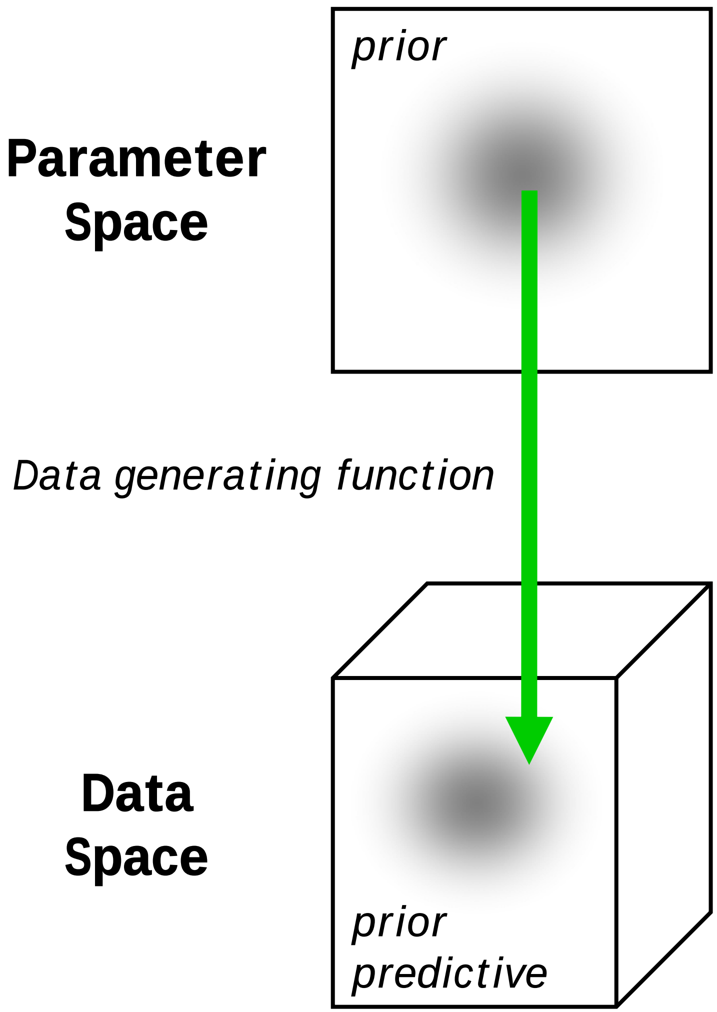
Psychological Models in Bayesian Framework

- The data generating function (primarily) and the prior distribution on parameters (under-used) formalize the model



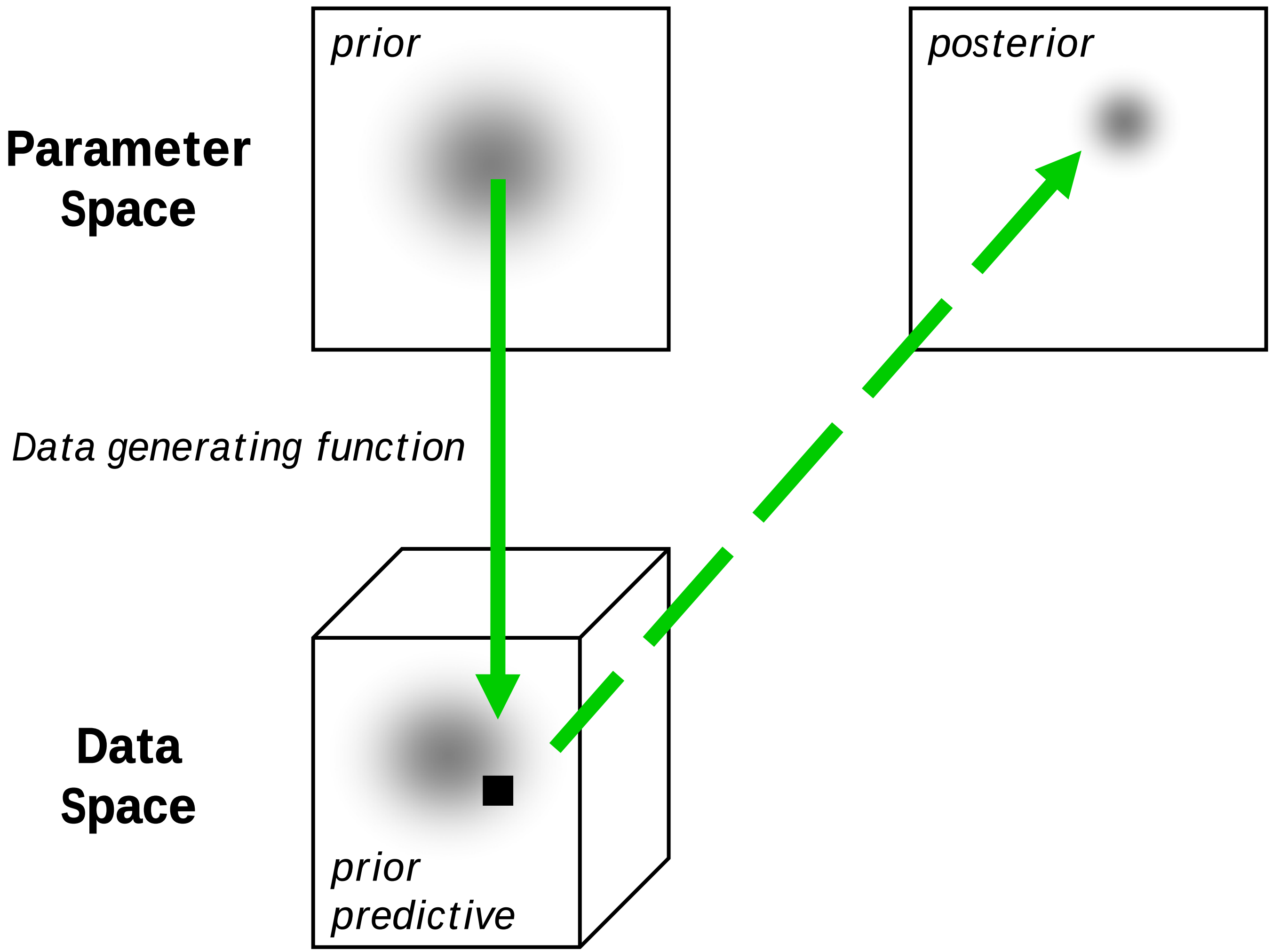
Psychological Models in Bayesian Framework

- This model, the prior plus data generating function (aka likelihood function), predict the nature of observed data



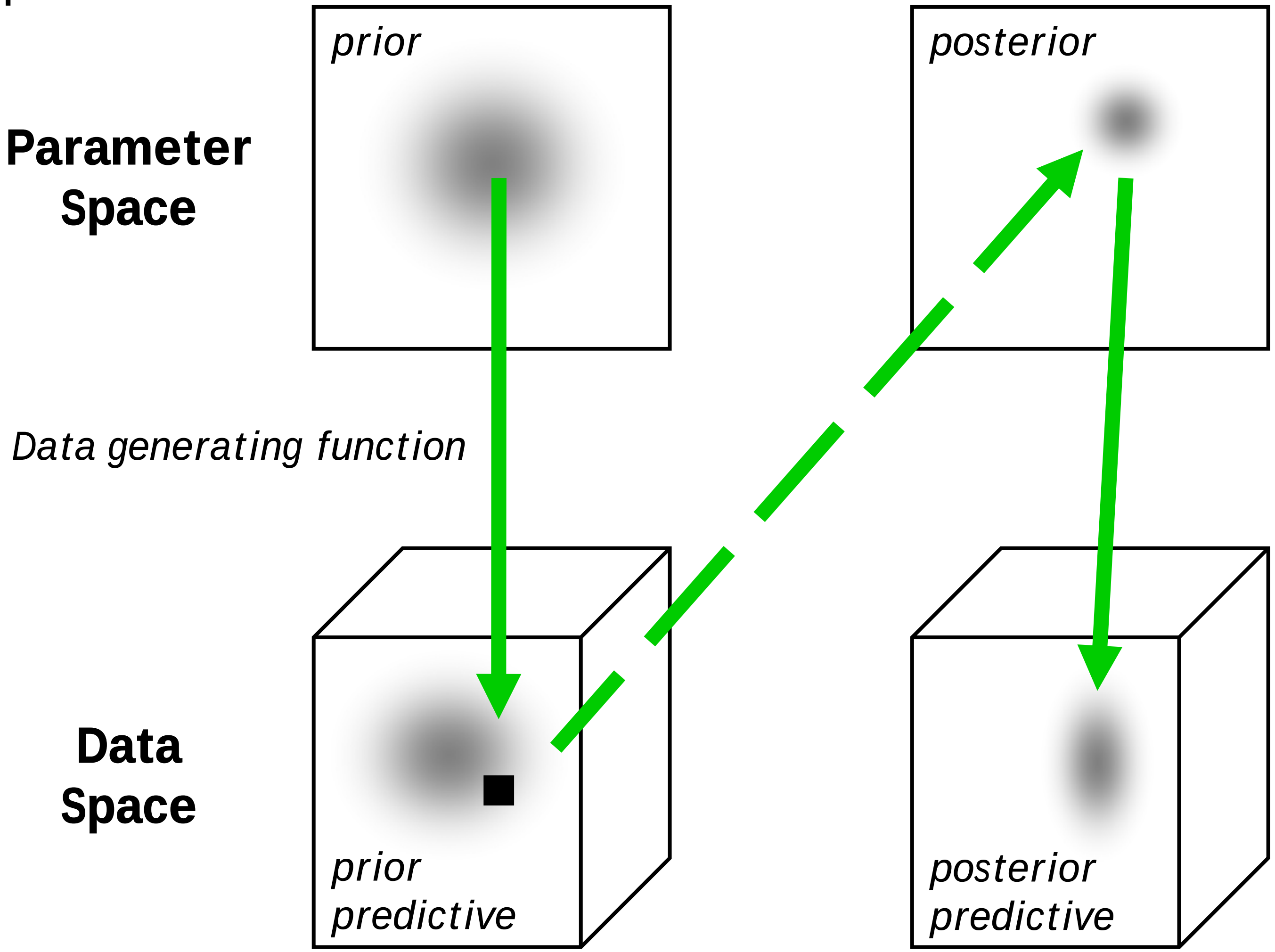
Psychological Models in Bayesian Framework

- Once data are observed, probability theory (via Bayes theorem) allows the prior over parameters to be updated to a posterior



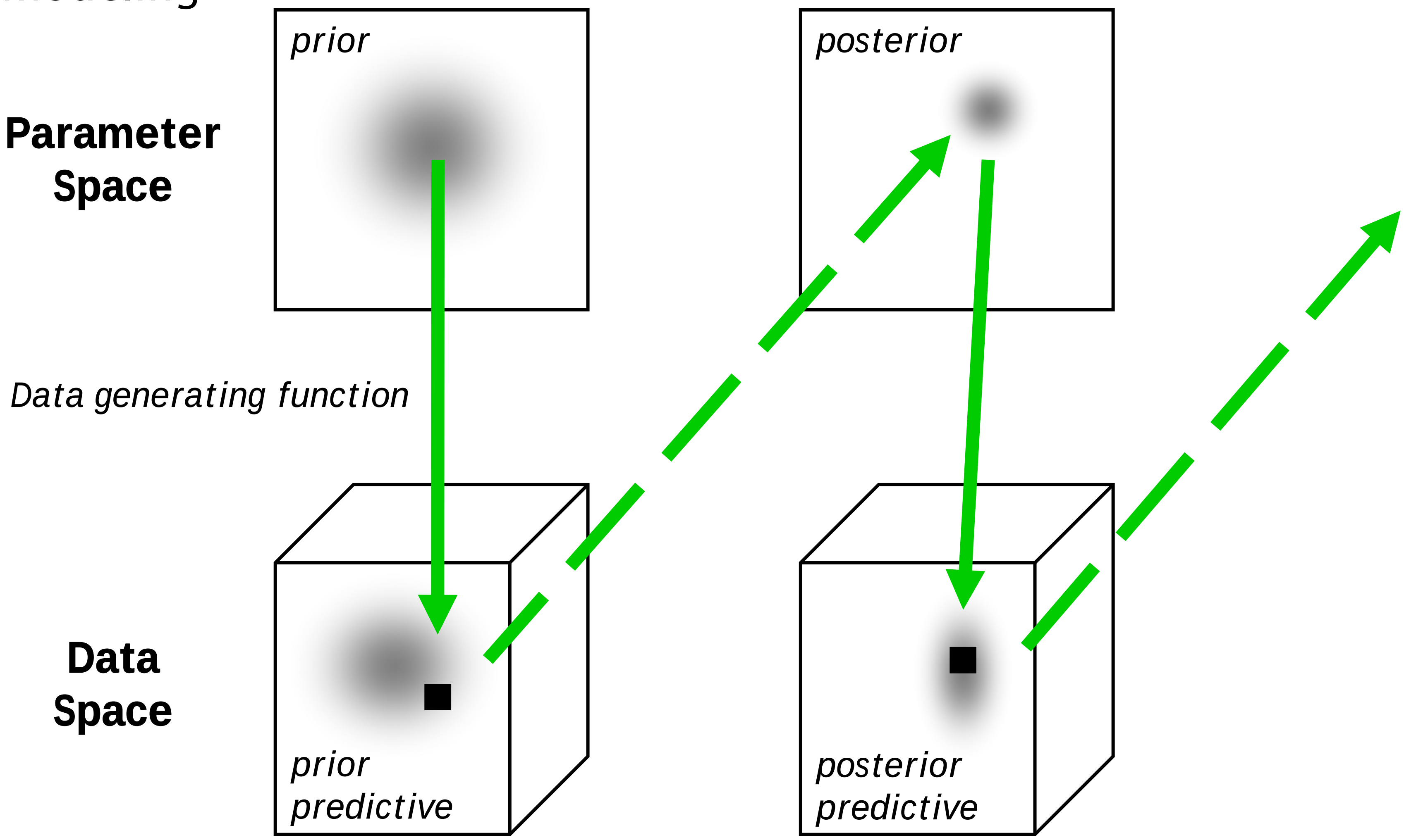
Psychological Models in Bayesian Framework

- The posterior distribution over parameters quantifies uncertainty about what is know and unknown, and makes predictions



Psychological Models in Bayesian Framework

- Bayesian inference is a complete framework for representing and incorporating information, in the context of psychological modeling

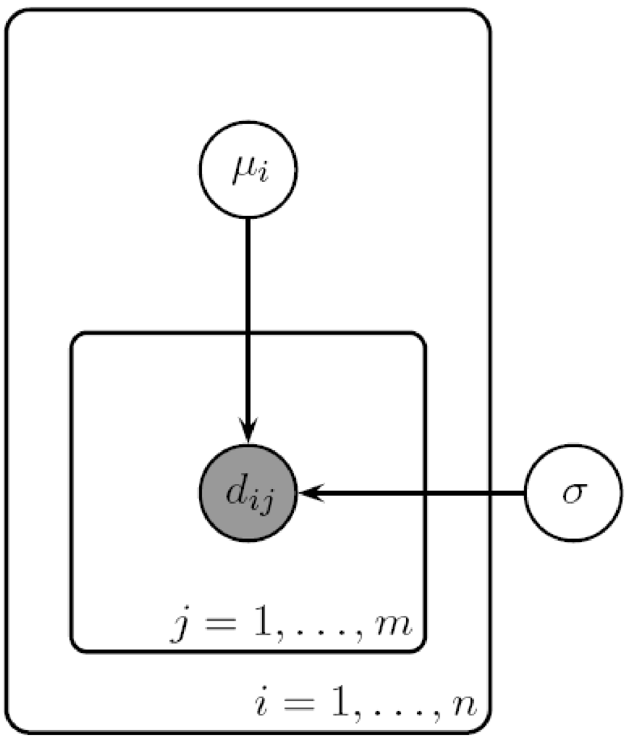


Example: Repeated Measurement of IQ (Lee & Wagenmakers to appear, ch. 6)

An example of the role of information, in the prior, or in the data, or both, in influencing estimation

Repeated IQ Measures

- Three people each have their IQ assessed 3 times by repeated versions of the same test
- The goals are
  - To infer each person's IQ
  - To infer the accuracy or reliability of the testing instrument



$$\mu_i \sim \text{Uniform}(0, 300)$$

$$\sigma \sim \text{Uniform}(0, 100)$$

$$d_{ij} \sim \text{Gaussian}(\mu_i, \tfrac{1}{\sigma^2})$$

Four Scenarios

- We do the inference four times
  - Their scores are either
    - **Imprecise test:** (90,95,100), (105,110,115), and (150,155,160)
    - **Precise test:** (94,95,96), (109,110,111) and (154,155,156)
  - The prior placed on each person's IQ is either
    - **Vague prior:** A flat prior from 0 to 300
    - **Informed prior:** A Gaussian prior with a mean of 100 and standard deviation of 15

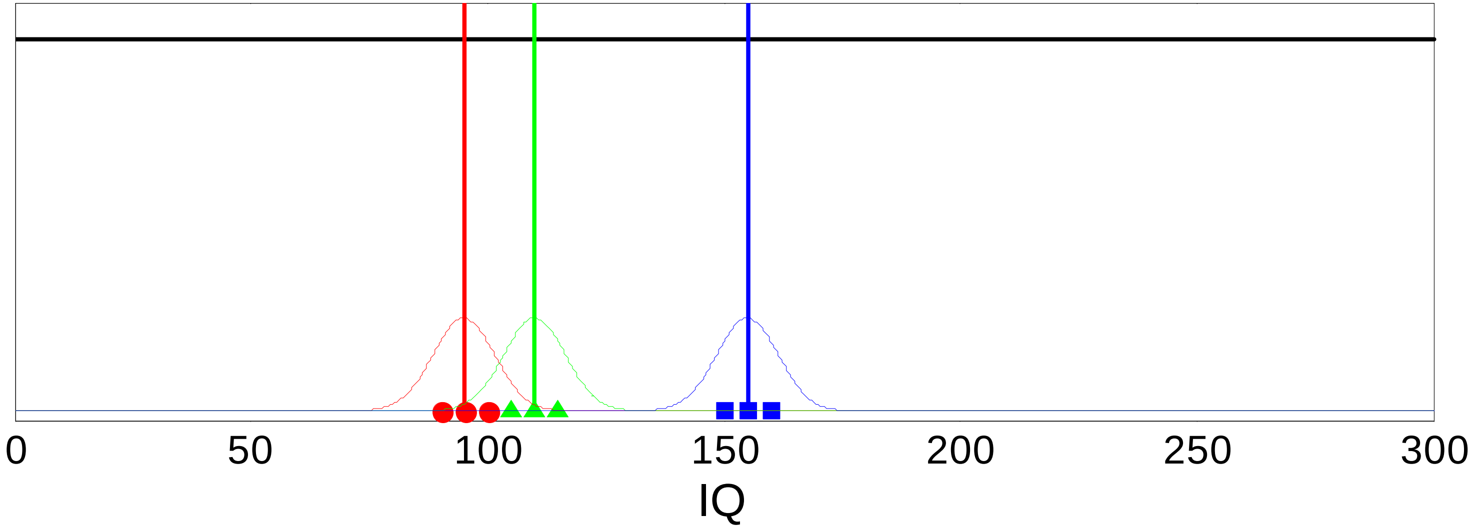
Results Summary

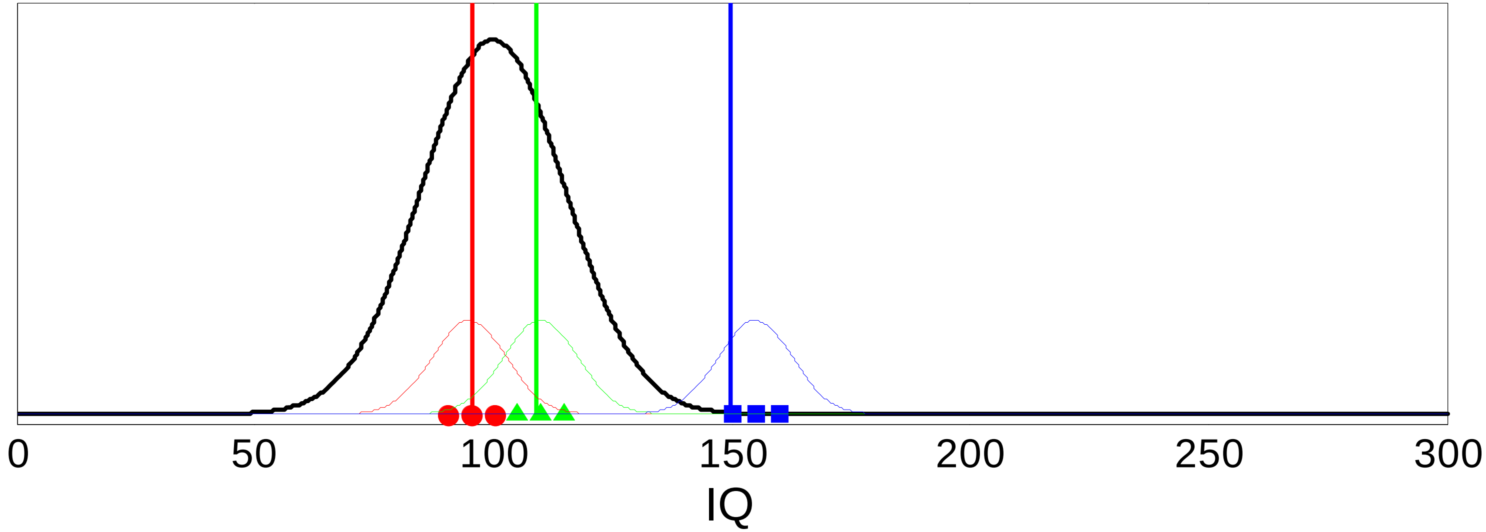
- The expectations of the posterior IQ distributions in each case are approximately

Data	Vague Prior	Informed Prior
(90,95,100)	95	95.5
(105,110,115)	110	109
(150,155,160)	155	150
(94,95,96)	95	95
(109,110,111)	110	110
(154,155,156)	155	154.9

Imprecise Test

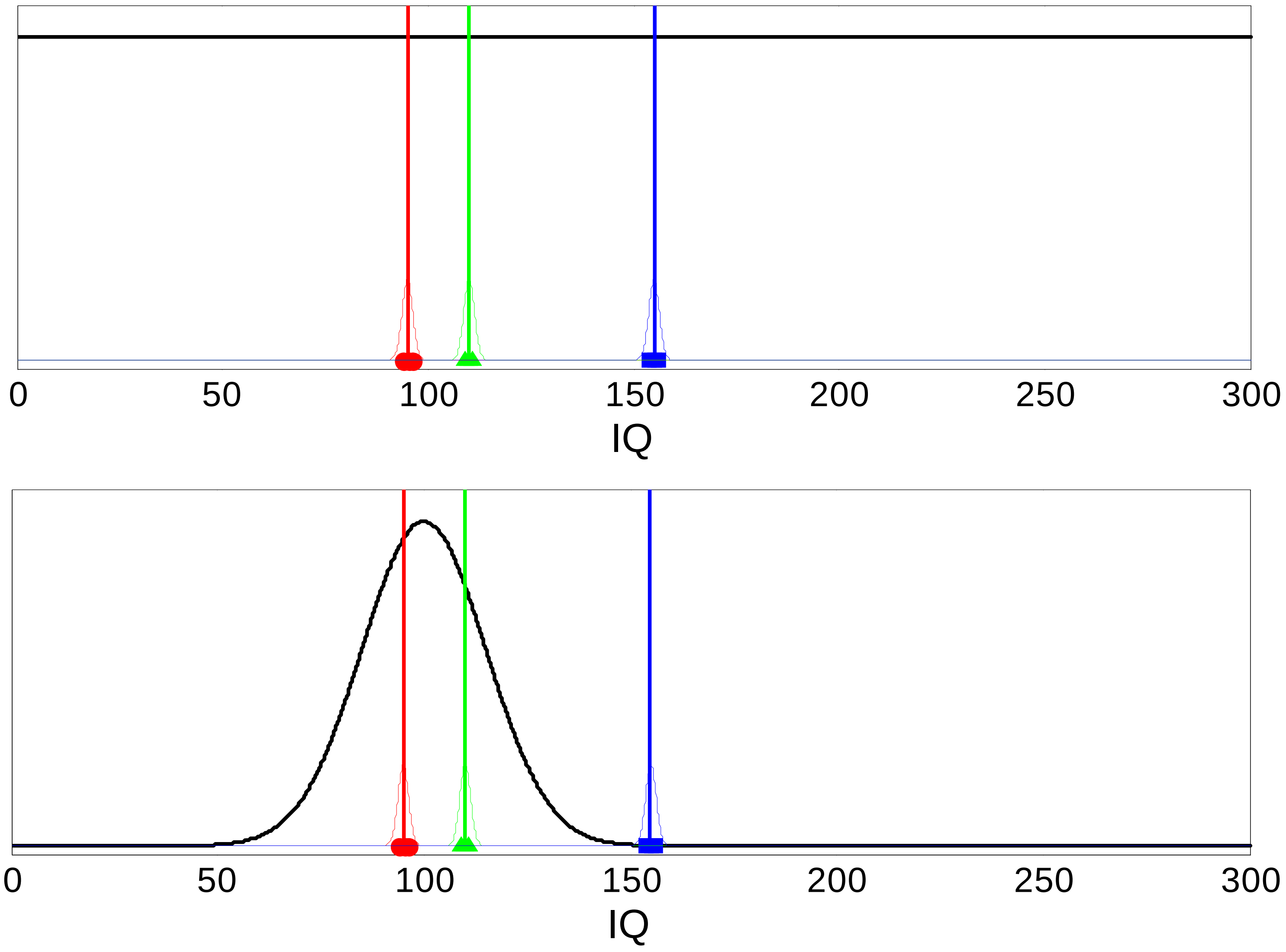
- The informed prior changes the estimate of the extreme case





Precise Test

- The data provide information that overwhelms the priors



Main Messages

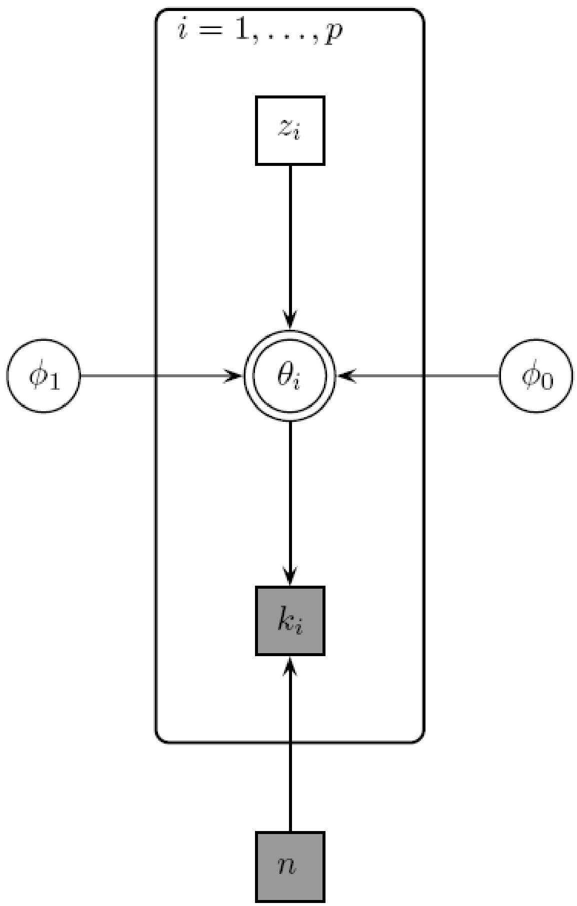
- Bayesian methods are naturally able to incorporate relevant prior information
  - This must improve inference, because the prior contains additional information that we are now able to use
  - The IQ example shows how inferences from an imprecise test can be influenced by prior knowledge about IQ distributions
- There is a familiar slogan that "with enough data, the influence of the prior will disappear"
  - This is often true, but not the best way to think about things
    - Irrelevant data will not update knowledge of a psychological parameter
    - The same number of observations will lessen the influence of the prior if the observations provide more information

Another example: Exams and Quizzes (Lee & Wagenmakers to appear, ch. 6)

An example of using latent mixture models to explain data as coming from more than one type of cognitive process

Exam Scores

- 16 people take a 40-item true-or-false test, and score 17, 18, 21, 21, 22, 28, 31, 31, 34, 34, 35, 35, 35, 36, 36, 39
- Model as a latent mixture of guessing and knowledge groups



$$z_i \sim \text{Bernoulli}(1/2)$$

$$\phi_0 = 1/2$$

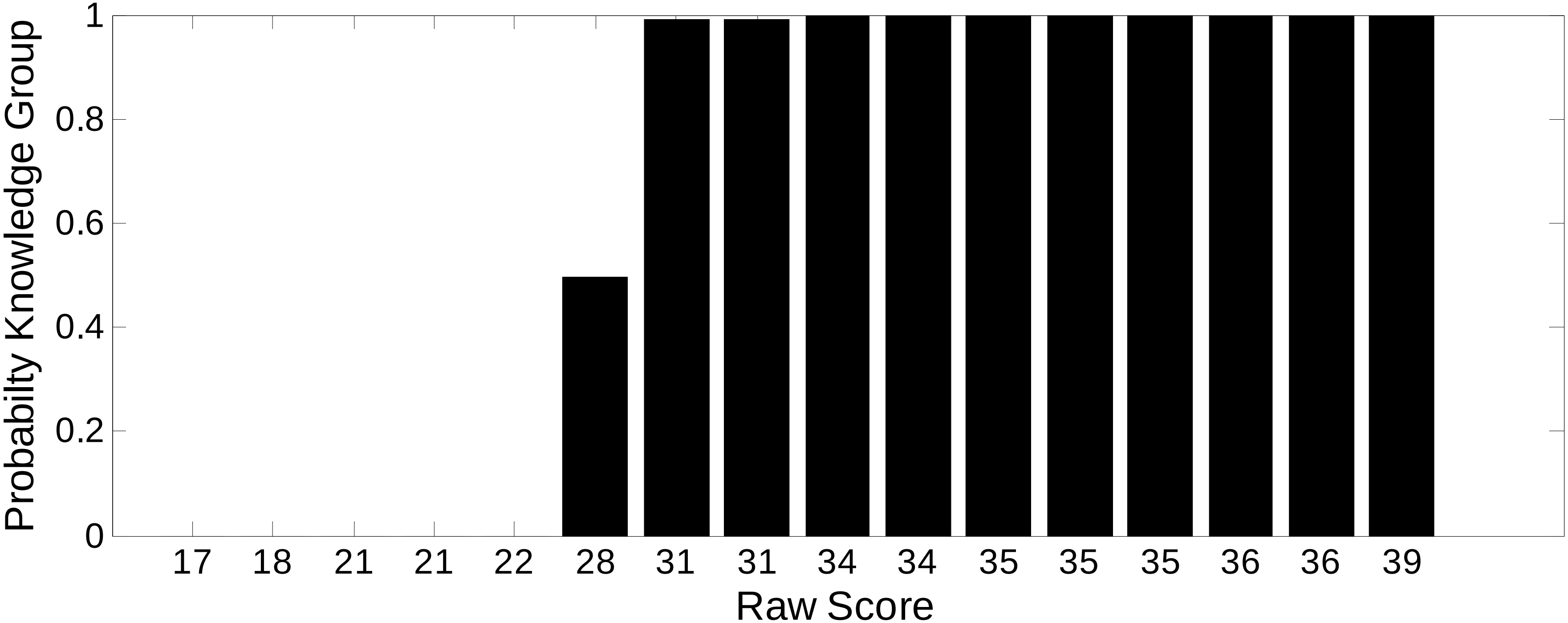
$$\phi_1 \sim \text{Uniform}(0.5, 1)$$

$$\theta_i = \phi_{z_i}$$

$$k_i \sim \text{Binomial}(\theta_i, n)$$

Latent Assignment Results

- The people who
  - Scored 17-22 are all classified as “guessers” with certainty
  - Scored 31+ are all classified as “knowing” with certainty
- There is uncertainty about the classification of the person who scored 28



Extensions and Main Messages

- Extensions of this basic latent mixture model to make it more psychologically interesting and plausible:
  - Allow for individual differences in the “knowledge” group
  - Allow for the base-rate of guessers vs knowers to be inferred (which in turn influences inference)

- Latent mixtures are a basic but probably under-used tool for cognitive science
  - Account for data as hierarchical mixtures of quantitatively and qualitatively different processes