Bayesian Inference (II)

Intro to Bayesian Data Analysis & Cognitive Modeling

Adrian Brasoveanu

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

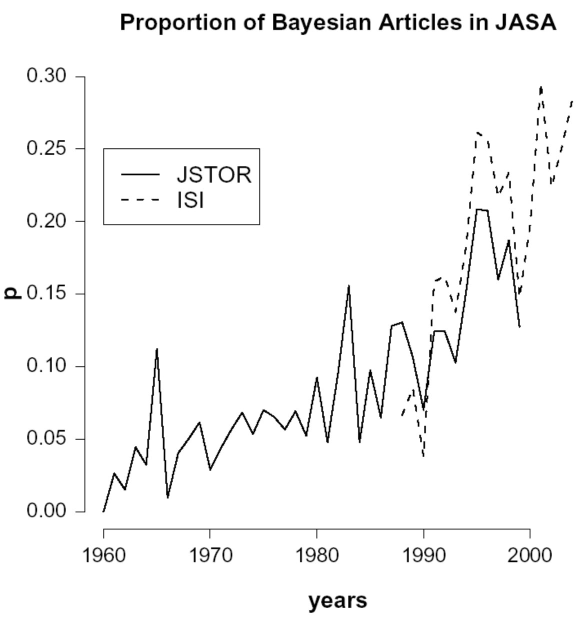
Fall 2012 • UCSC Linguistics

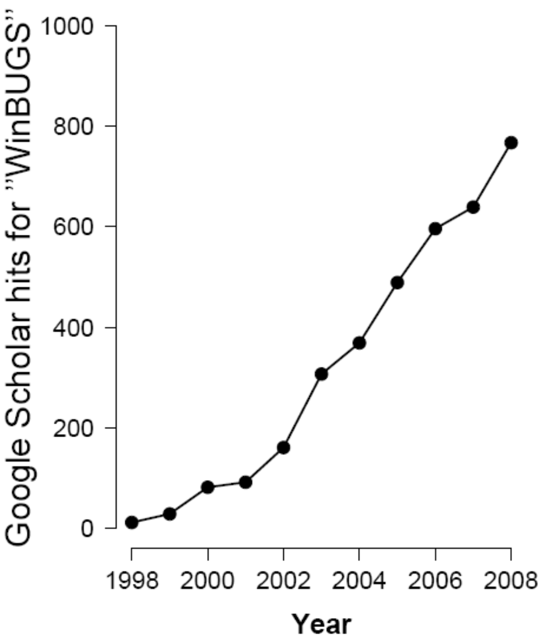
Bayesian Inference in a Nutshell (Again)

- 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.
- We will spend a lot of time talking about coins: our beliefs about their (latent/unobserved) bias, and how to update our beliefs when we gather data (observed coin flips)
- These slides provide more examples of actual phenomena that have the same structure as coin flips – this helps us see that we are talking about useful things

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*.”





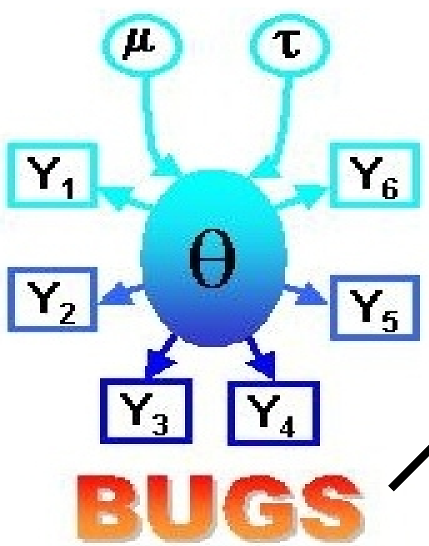
Why Bayes is Now Popular

Markov Chain Monte Carlo (MCMC)!

- Instead of calculating the posterior analytically, numerical techniques such as MCMC approximate the posterior by drawing samples from it.

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 yourself (we'll do a little bit of that in the beginning to better understand what's going on)
  - Option 2: use JAGS/WinBUGS/OpenBUGS (we'll do a lot of that for all realistic models we would want to use)



Bayesian inference **U**sing **G**ibbs **S**ampling

JAGS:
Just **A**nother **G**ibbs **S**ampler

JAGS/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

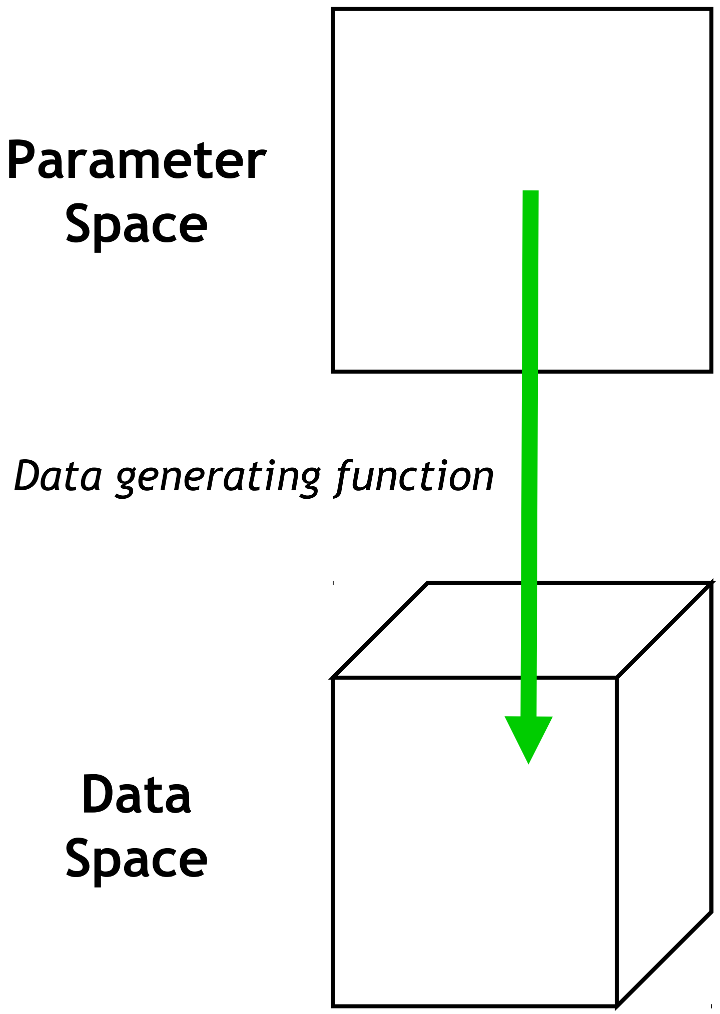
Dennis Lindley

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

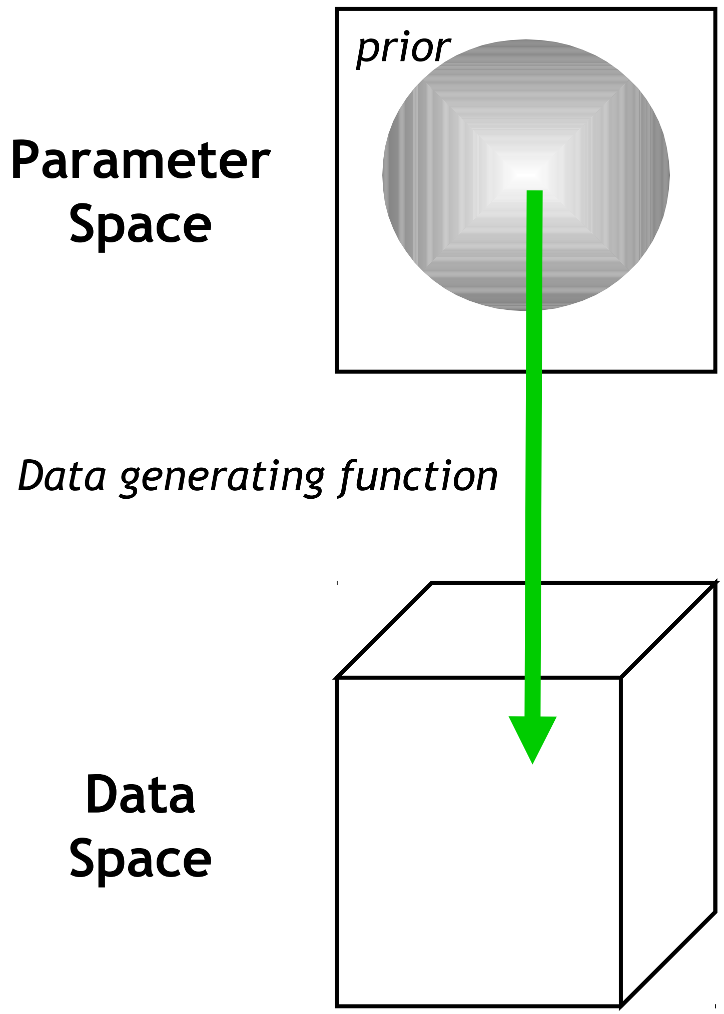
Psychological Models in Bayesian Framework

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



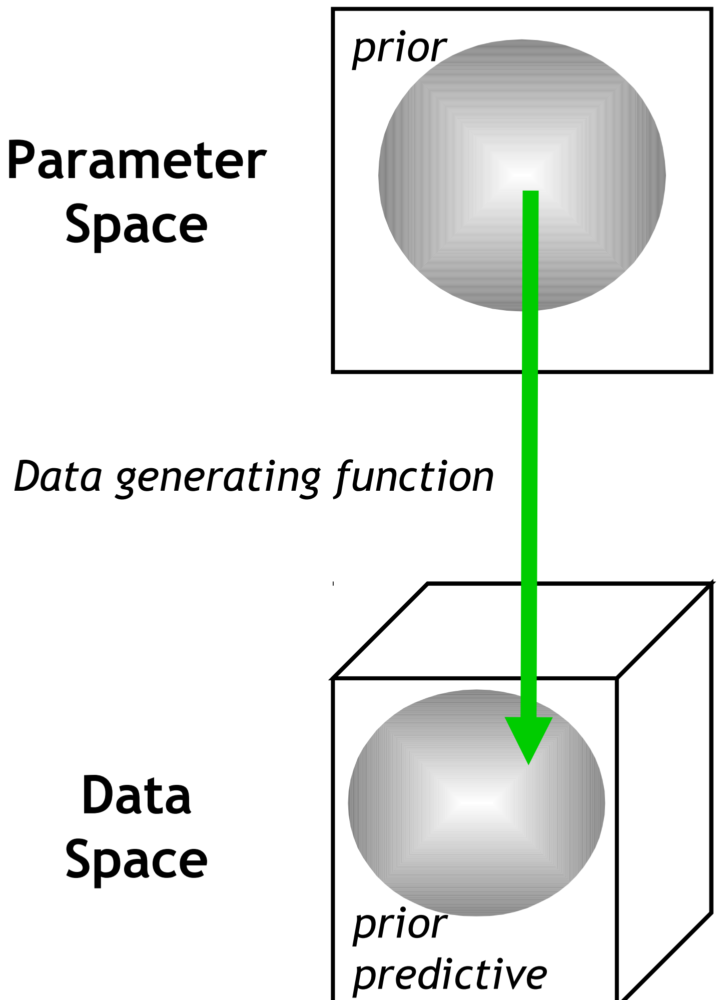
Psychological Models in Bayesian Framework

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



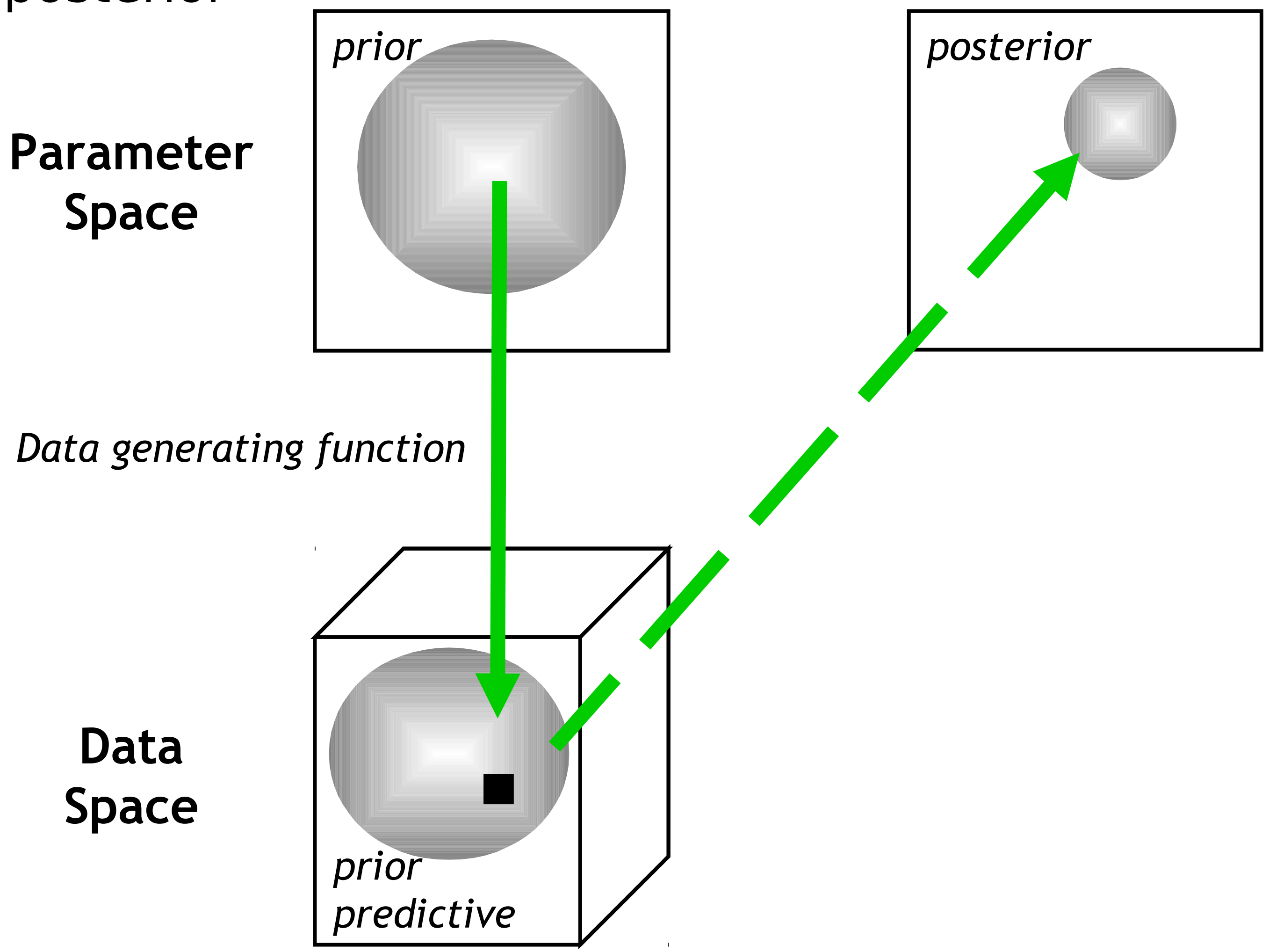
Psychological Models in Bayesian Framework

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



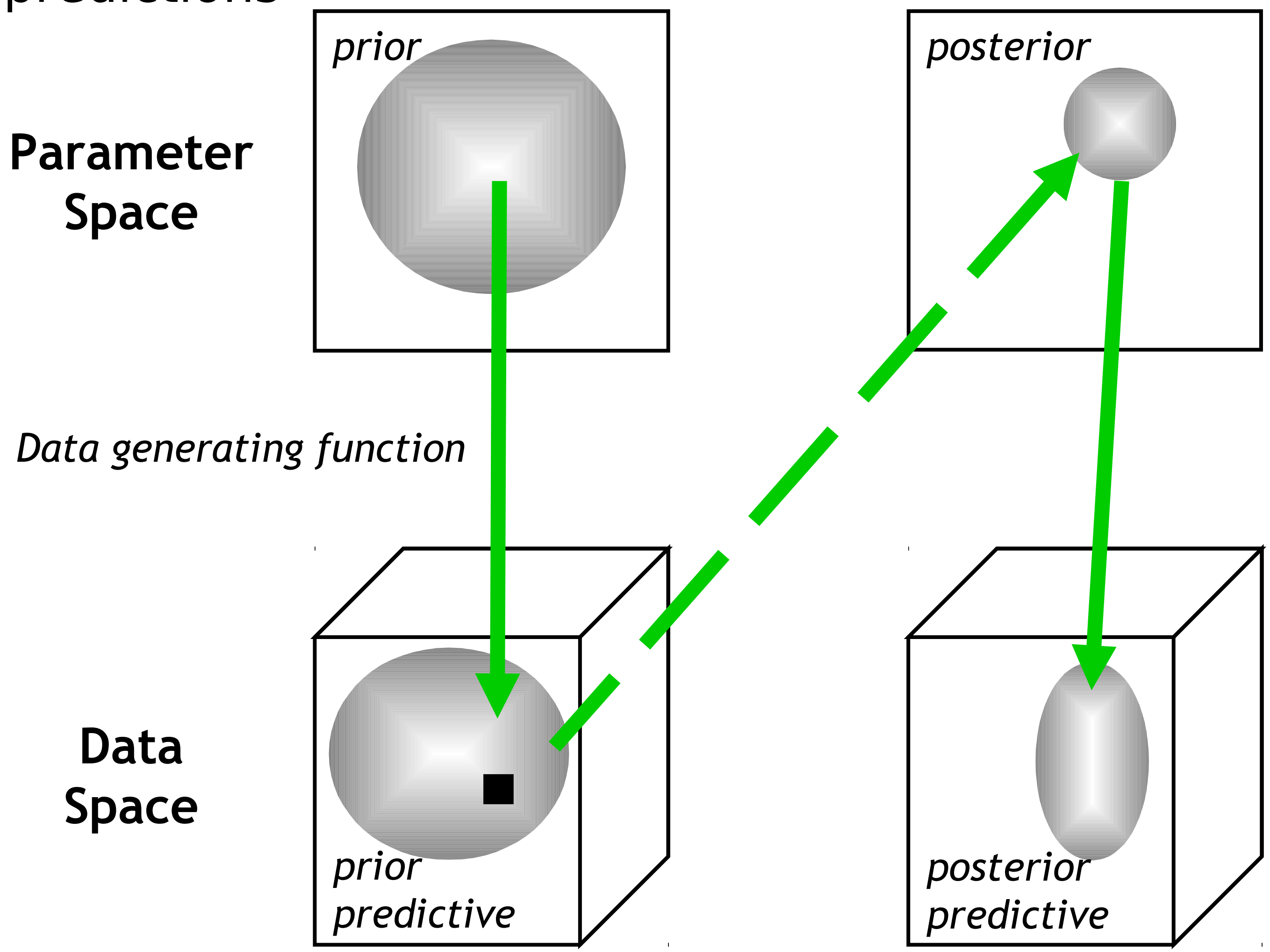
Psychological Models in Bayesian Framework

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



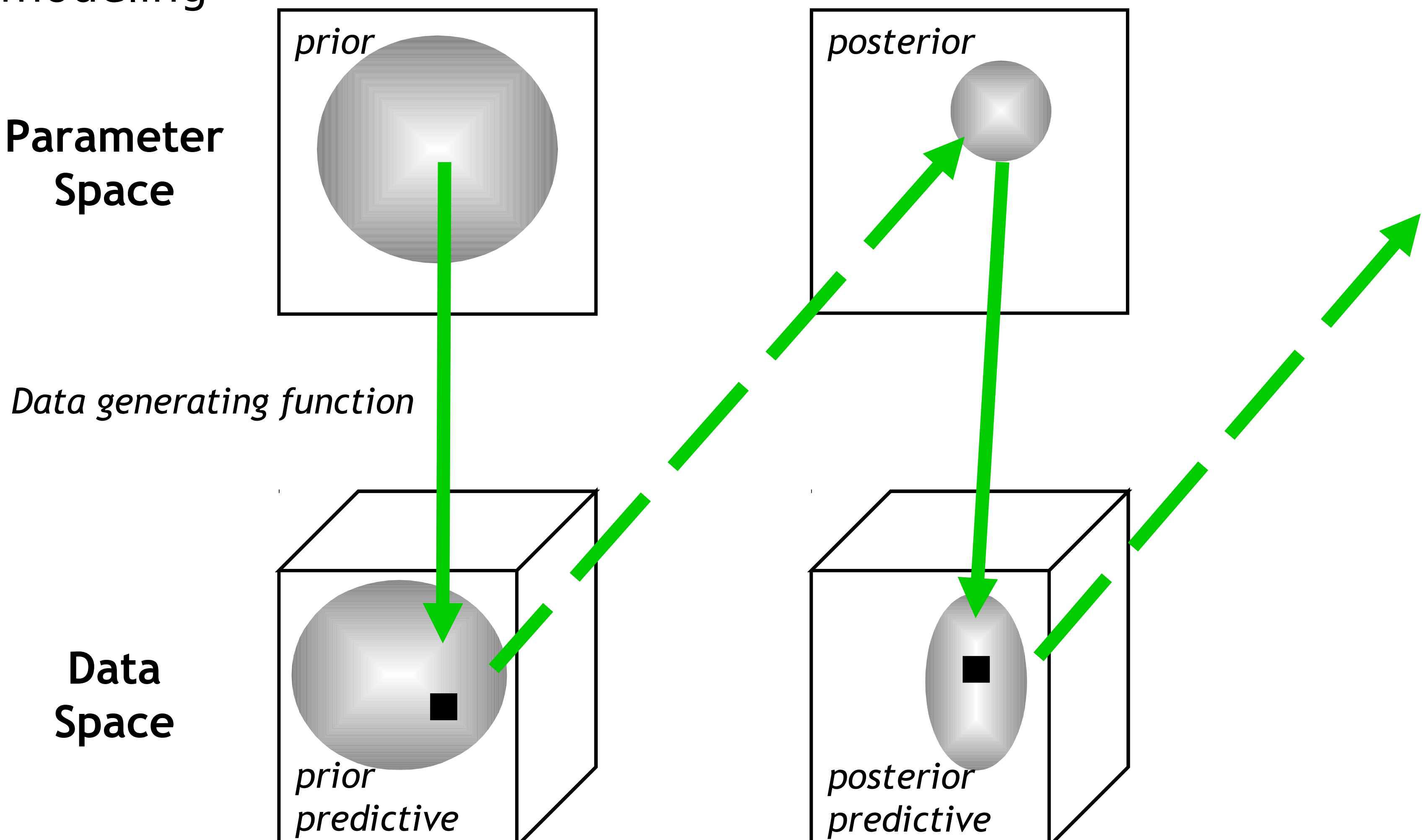
Psychological Models in Bayesian Framework

- 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, the data, or both) in influencing estimation

- Three people each have their IQ assessed 3 times by repeated versions of the same test
- The goal is to infer the IQ of each person

Four Scenarios

- We do the inference four times

  - 2 options for their scores – 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)

  - 2 options for the prior – 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

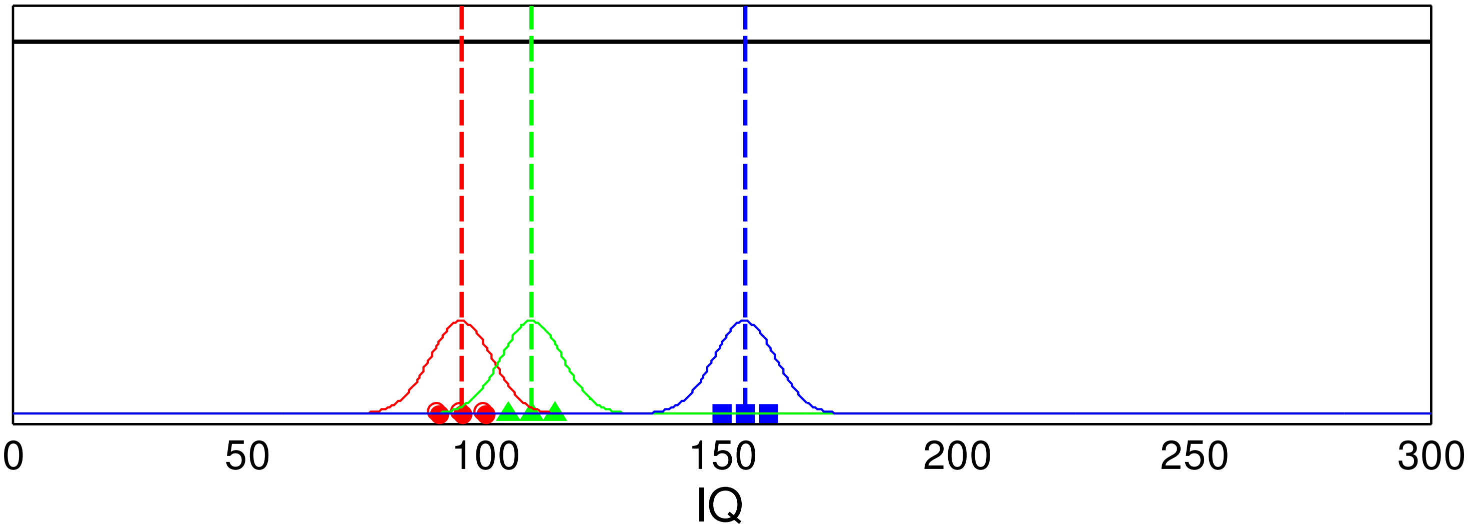
- The expectations of the posterior IQ distributions in each case are approximately

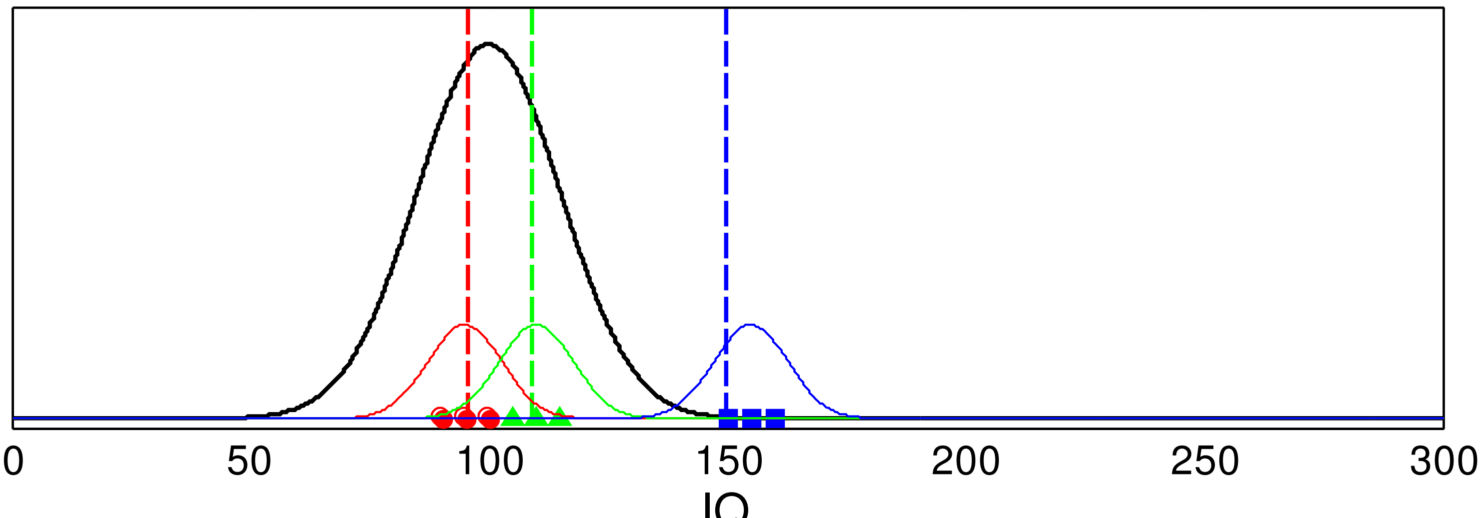
Imprecise Test Data	Vague Prior	Informed Prior
(90,95,100)	95	95.5
(105,110,115)	110	109
(150,155,160)	155	150

Precise Test Data	Vague Prior	Informed Prior
(94,95,96)	95	95
(109,110,111)	110	110
(154,155,156)	155	154.9

Imprecise Test

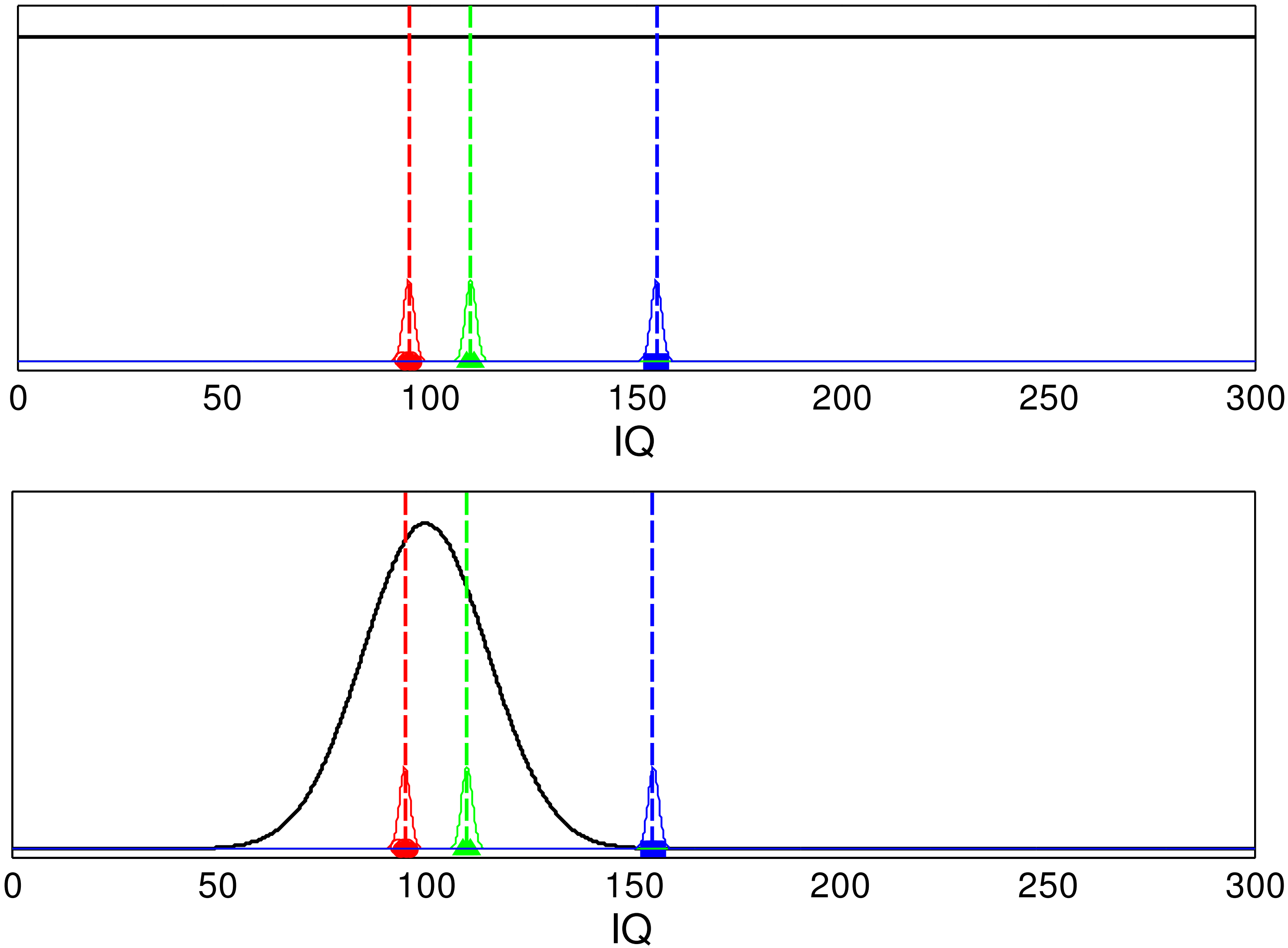
- The informed prior changes the estimate of the extreme case





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 sometimes 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

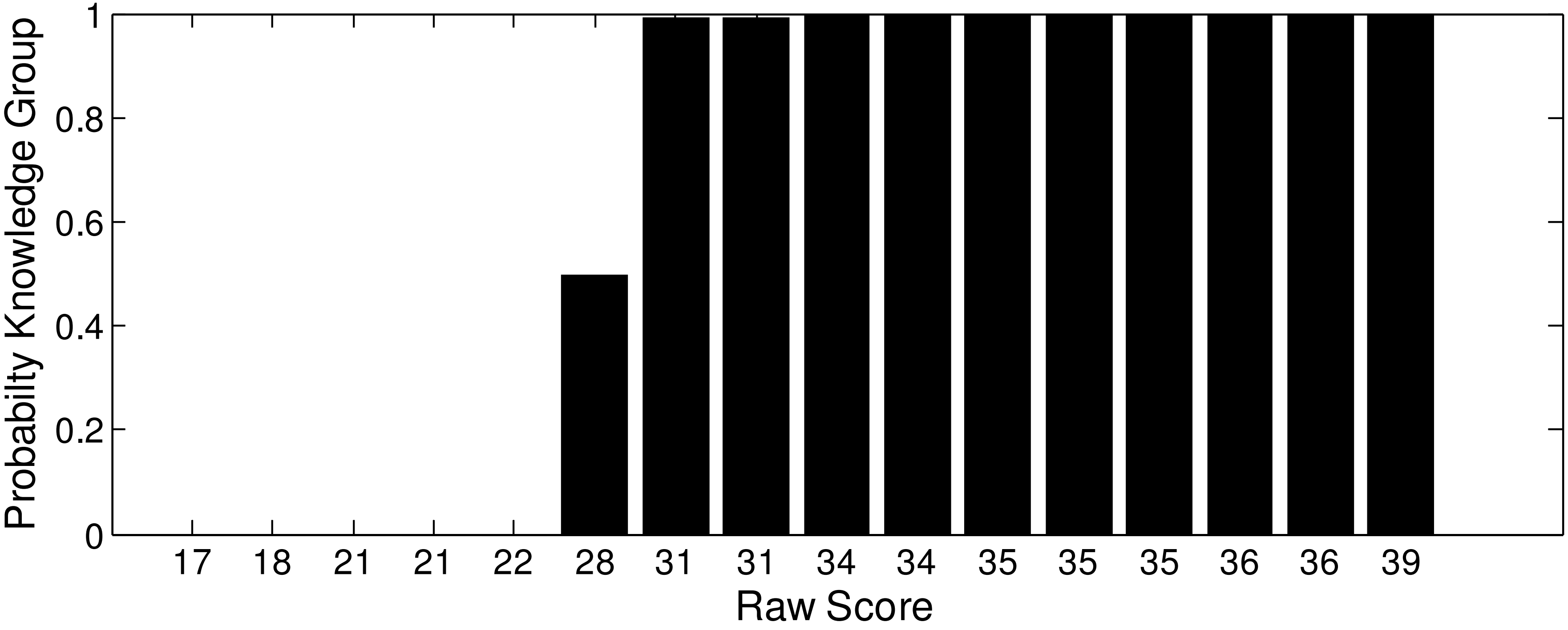
Another example: Exams and Quizzes (Lee & Wagenmakers to appear, ch. 6)

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

- 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
- A participant is equally likely to be part of the guessing or knowledge group (base rate for guessers vs knowers is the the same)

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 (and data analysis)
  - Account for data as hierarchical mixtures of quantitatively and qualitatively different processes