Probabilistic Shape Modelling
- Foundational principles -

16. April 2019

Marcel Lüthi

Graphics and Vision Research Group
Department of Mathematics and Computer Science
University of Basel

Probabilistic Shape Modelling

Online Course / Futurelearn

Shape Modelling

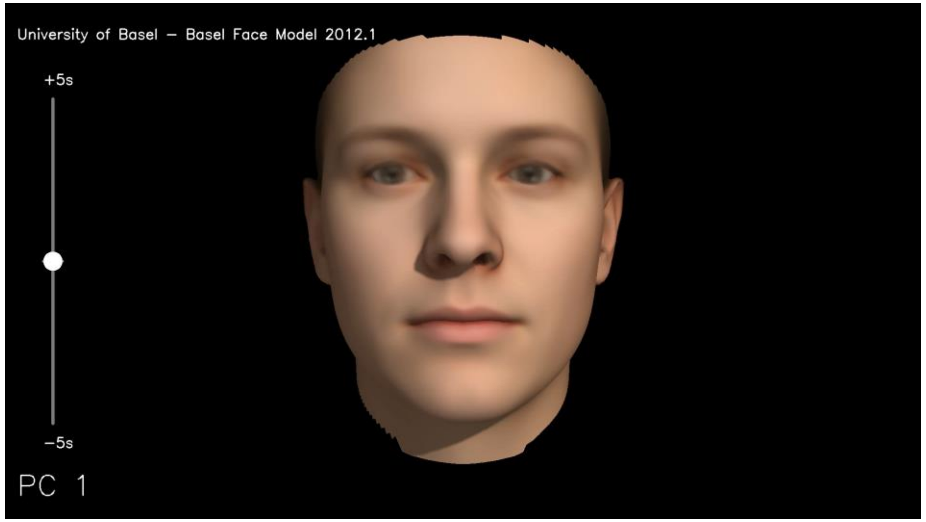
Next lectures

Model fitting

Scalismo

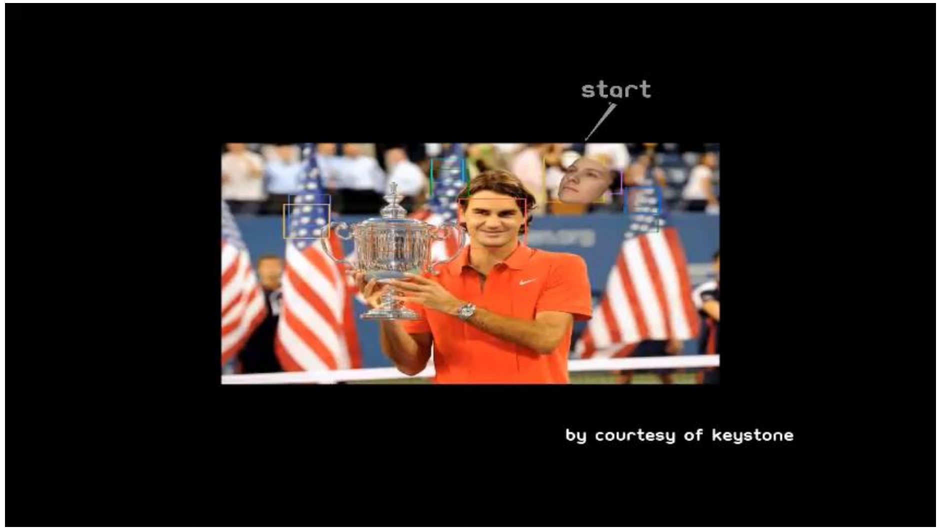
Probabilistic Shape Modelling

Online Course / Futurelearn



Shape Modelling

Next lectures



Model fitting

Scalismo

Programme

	Lecture (14.15 – 16.00)	Exercises (16.15 - 18.00)
16. April	• Analysis by Synthesis • Introduction to Bayesian modelling	
23. April	• Non-rigid registration A probabilistic interpretation	• Introducing Project 2 • Working on exercise sheet
30. April	• Markov Chain Monte Carlo for model fitting (I) • Feedback - Project 1	• Discussion: Exercise sheet 3
7. Mai	• Markov Chain Monte Carlo for model fitting (II)	• Working on Project 2
14. Mai	• Face Image Analysis	• Progress discussion: Project 2
21. Mai	• Gaussian processes More insights / connections to other methods	• Working on Project 2
28. Mai	• Summary • Q & A (Exam, …)	

Outline

Analysis by synthesis

- The conceptual framework we follow in this course

Intermezzo: Bayesian inference

- How we reason in this course

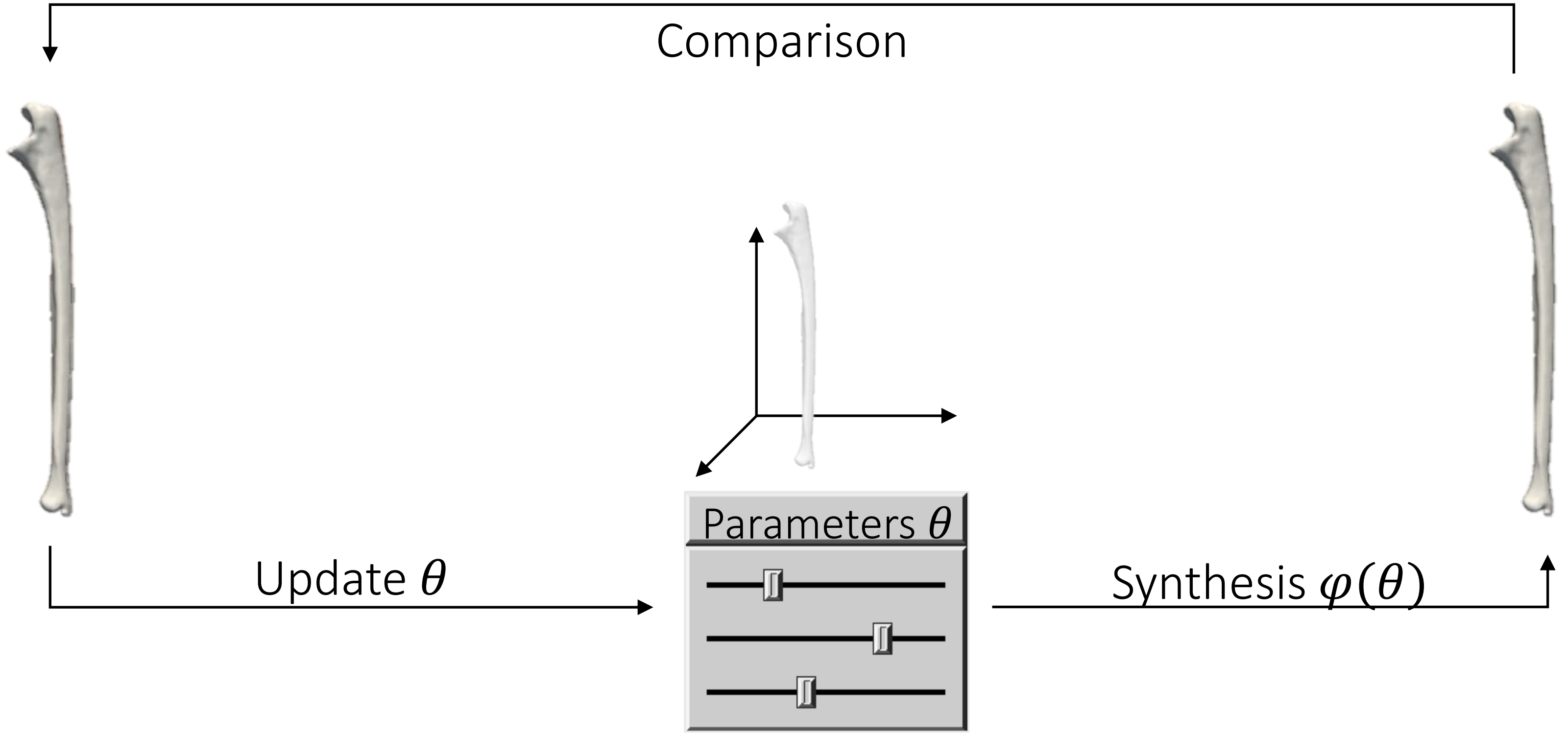
Analysis by Synthesis in 5 (simple) steps

- A step by step guide to image analysis

Computer vision verse medical image analysis

- Some commonalities and differences of the two fields

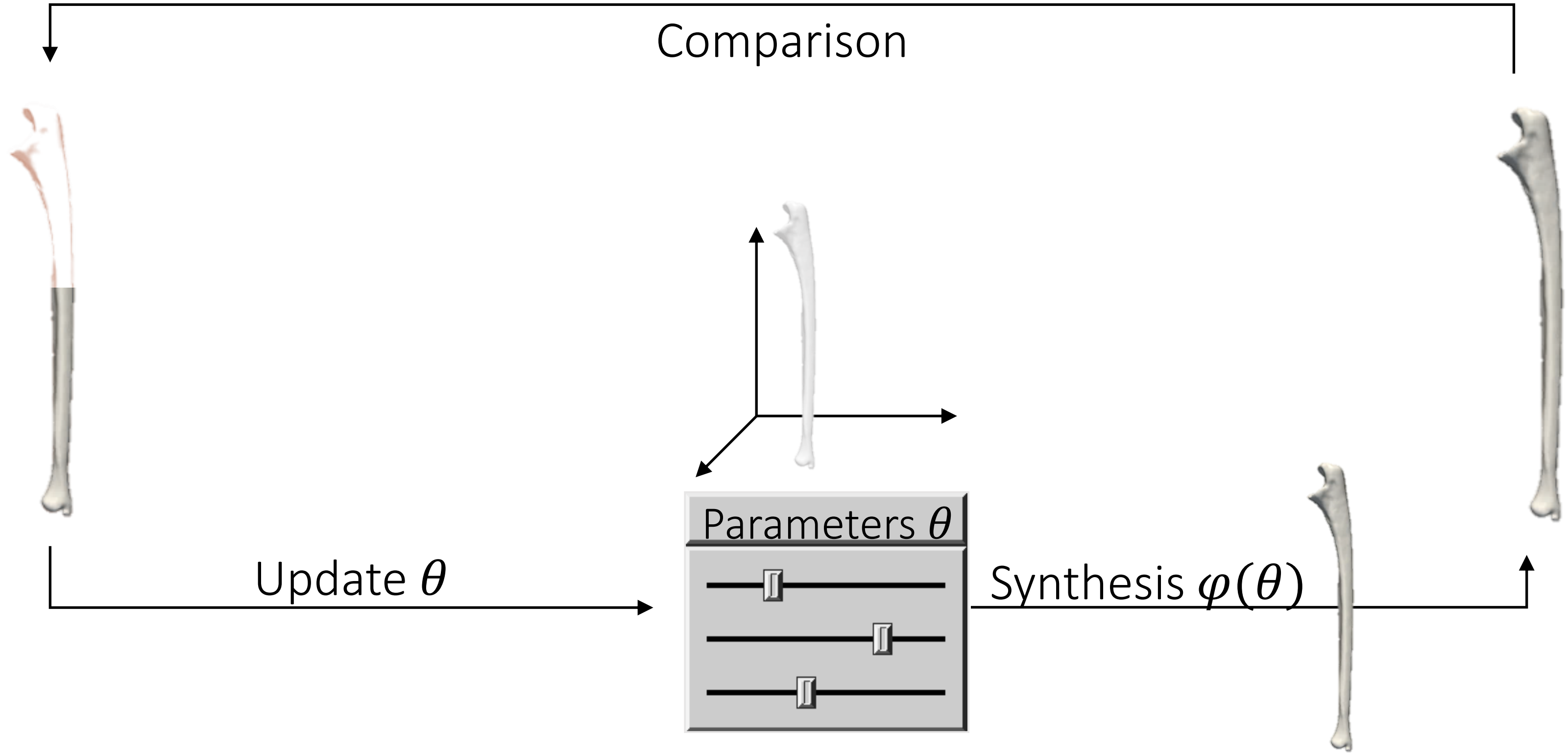
Conceptual Basis: Analysis by synthesis



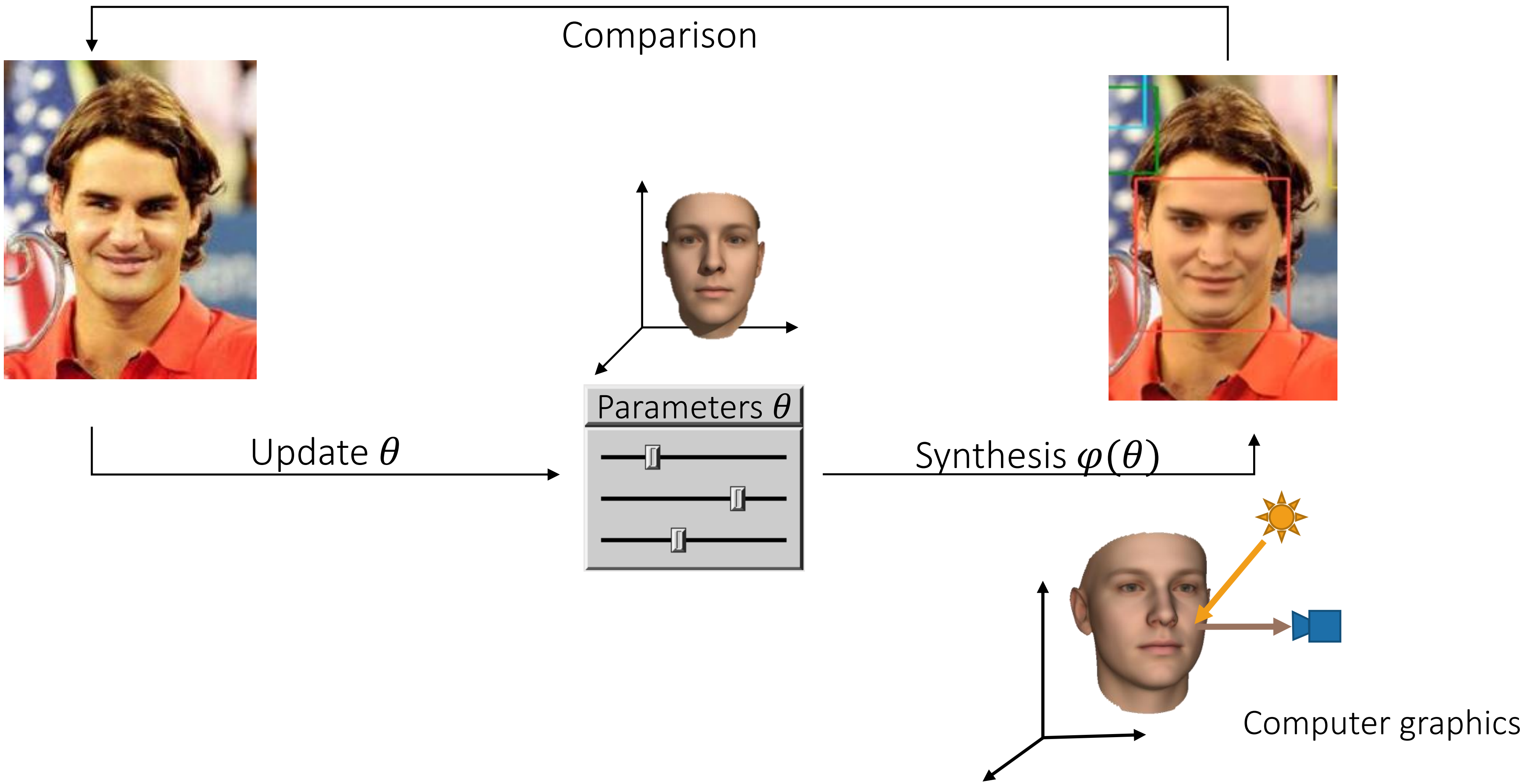
Being able to synthesize data means we can understand how it was formed.

– Allows reasoning about unseen parts.

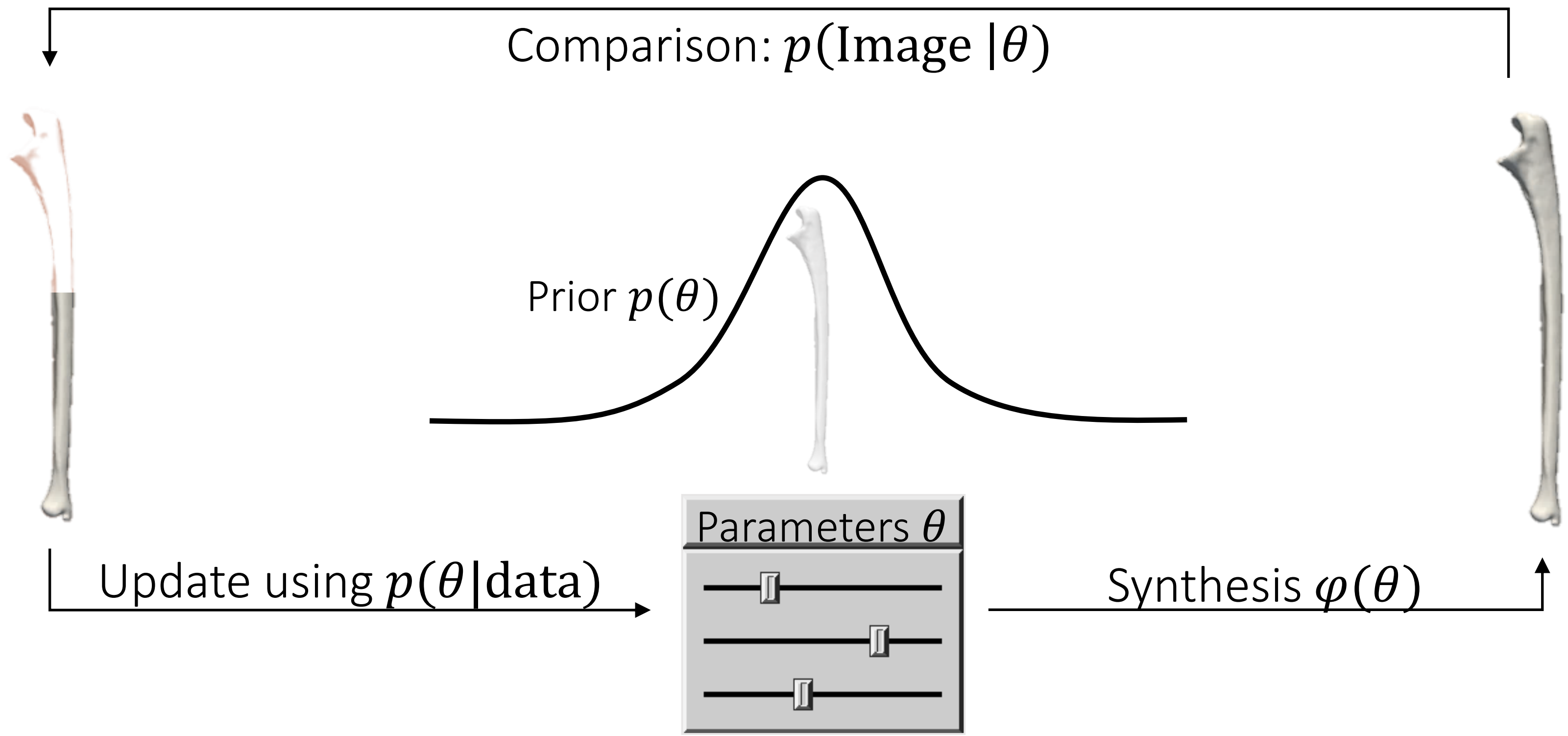
Conceptual Basis: Analysis by synthesis



Conceptual Basis: Analysis by synthesis

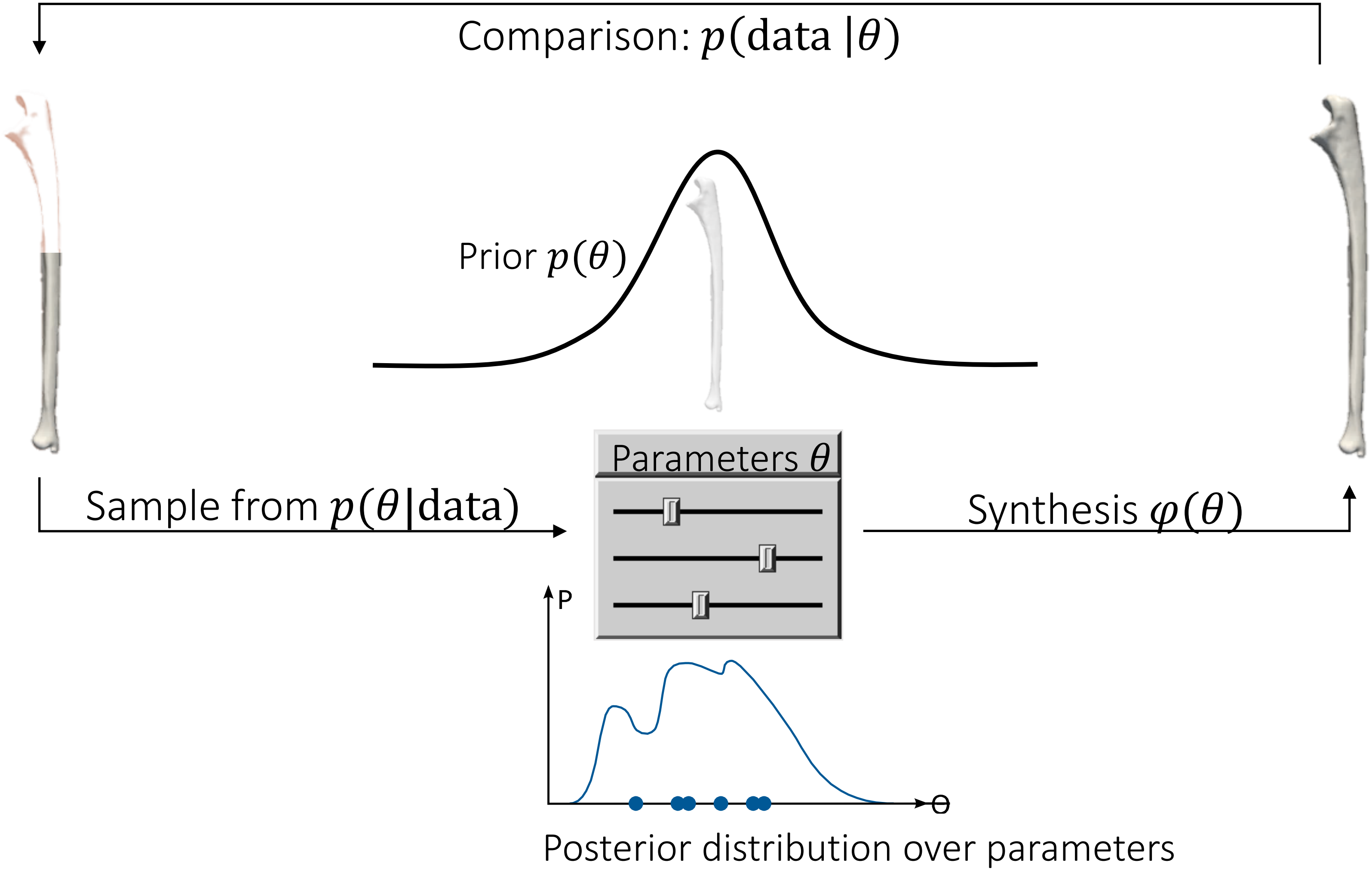


Mathematical Framework: Bayesian inference

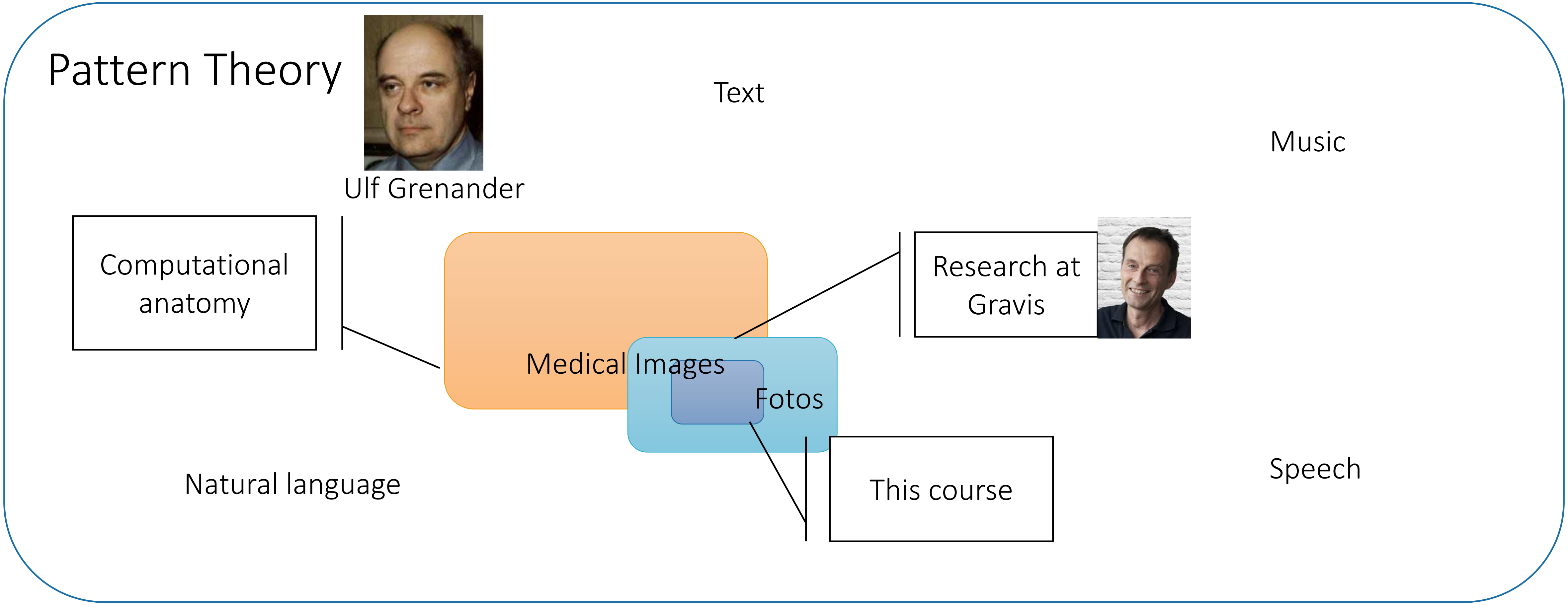


- Principled way of dealing with uncertainty.

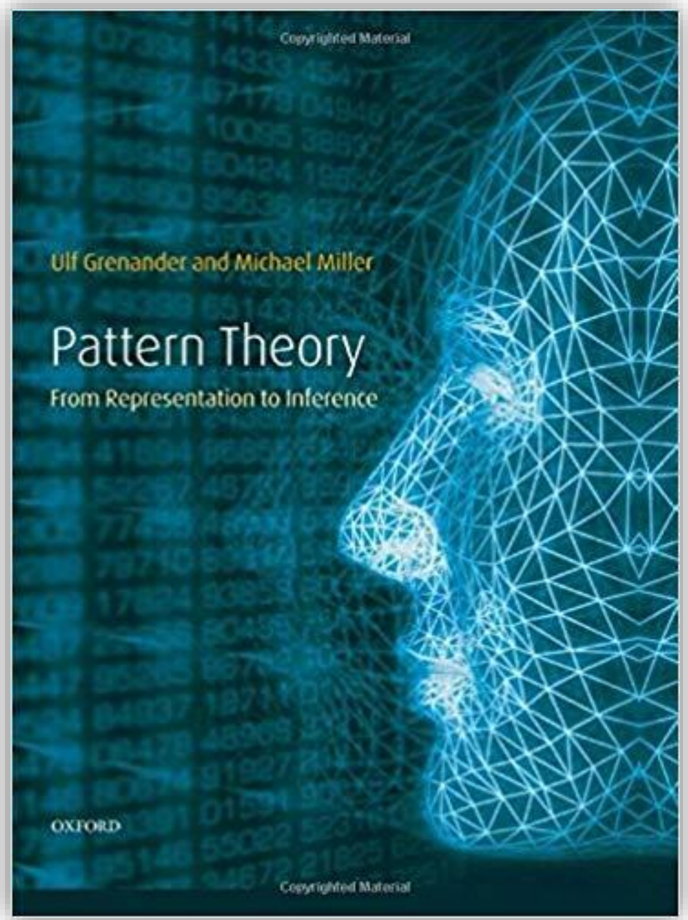
Algorithmic implementation: MCMC

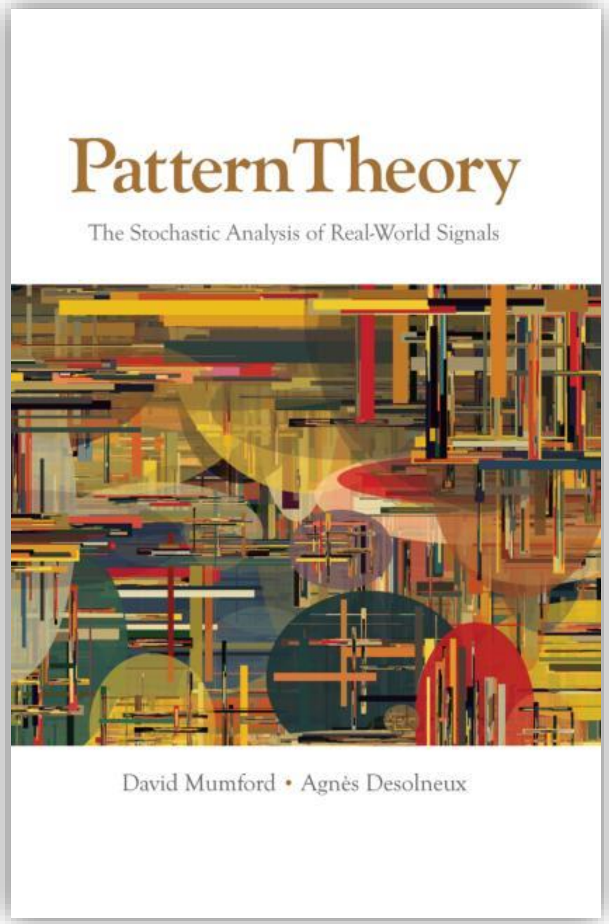


The course in context



Pattern theory – The mathematics





Intermezzo: Bayesian inference

Probabilities: What are they?

Four possible interpretations:

1. Long-term frequencies
   - Relative frequency of an event over time
2. Physical tendencies (propensities)
   - Arguments about a physical situation (causes of relative frequencies)
3. Degree of belief (Bayesian probabilities)
   - Subjective beliefs about events/hypothesis/facts
4. Logic
   - Degree of logical support for a particular hypothesis

Bayesian probabilities for image analysis

Bayesian probabilities make sense where frequentists interpretations are not applicable!

- No amount of repetition makes image sharp.
  - Uncertainty is not due to random effect, but because of bad telescope.
- Still possible to use Bayesian inference.
  - Uncertainty summarizes our ignorance.

Gallileo’s view on Saturn

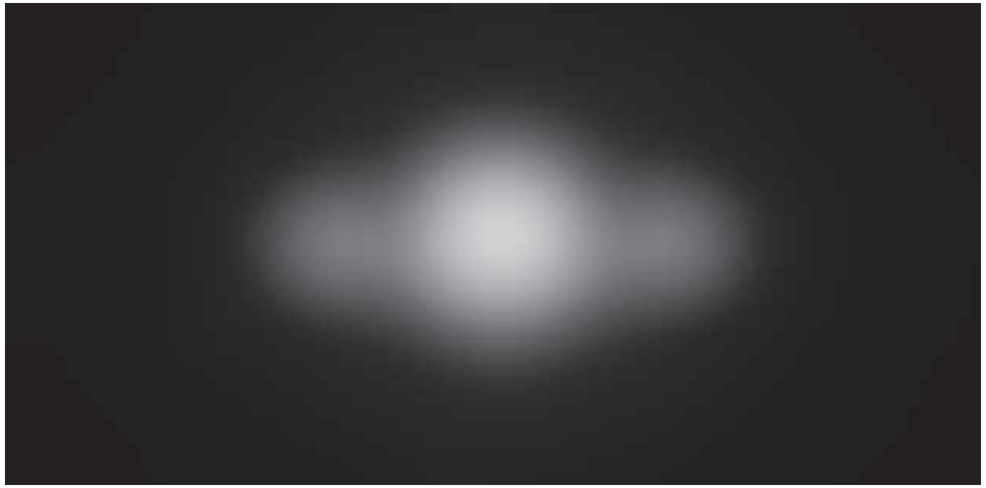


Image credit: McElrath, Statistical Rethinking: Figure 1.12

Degree of belief: An example

- Dentist example: *Does the patient have a cavity?*

$$P(\text{cavity}) = 0.1$$

$$P(\text{cavity}|\text{toothache}) = 0.8$$

$$P(\text{cavity}|\text{toothache}, \text{gum problems}) = 0.4$$

But the patient either has a cavity or does not

- *There is no 80% cavity!*
- *Having a cavity should not depend on whether the patient has a toothache or gum problems*

Statements do not contradict each other - they summarize ***the dentist's knowledge*** about the patient

AIMA: Russell & Norvig, *Artificial Intelligence. A Modern Approach,* 3rd edition,

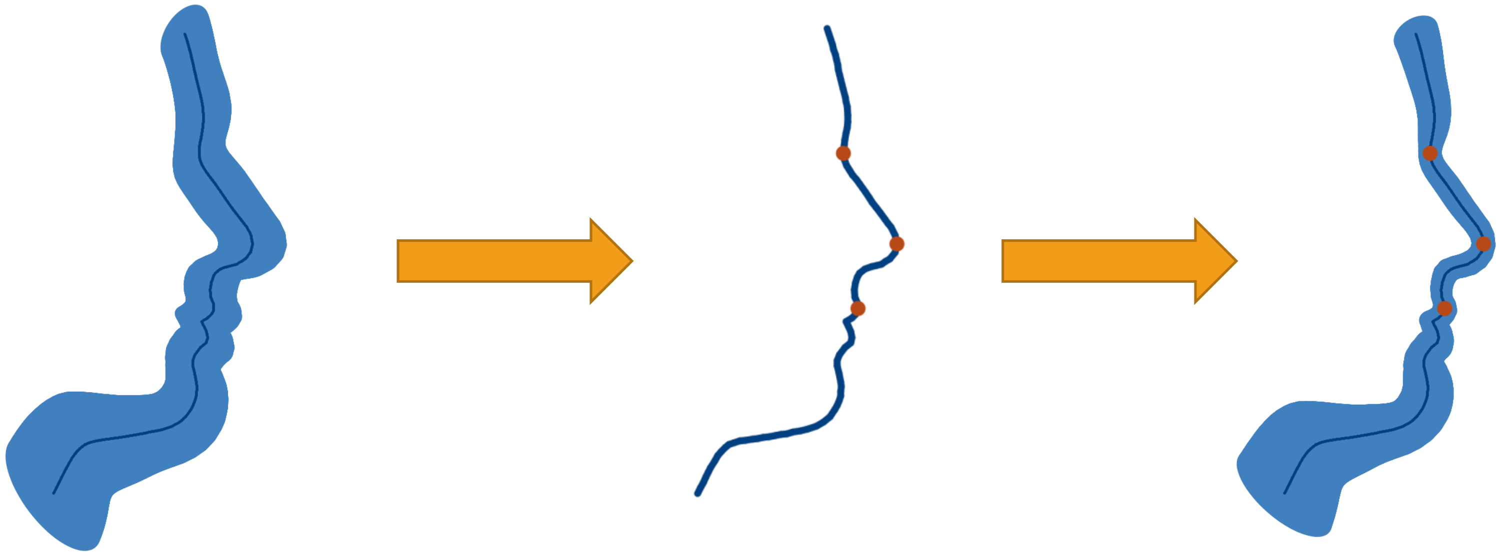
Uncertainty: Bayesian probability

- Bayesian probabilities rely on a *subjective* perspective:
  - Probabilities express our *current knowledge*.
  - Can *change* when we learn or see more
  - More data -> more *certain* about our result.

Subjectivity: There is no single, real underlying distribution. A probability distribution expresses our knowledge – It is different in different situations and for different observers since they have different knowledge.

  - Subjective != Arbitrary
  - Given belief, conclusions follow by laws of probability calculus

Belief Updates



Model
Face distribution

Observation
Concrete points
Possibly uncertain

Posterior
Face distribution
consistent with observation

Prior belief

More knowledge

Posterior belief

Two important rules

Probabilistic model: joint distribution of points

$$P(x_1, x_2)$$

Marginal

Distribution of certain points only

$$P(x_1) = \sum_{x_2} P(x_1, x_2)$$

Conditional

Distribution of points conditioned on *known* values of others

$$P(x_1|x_2) = \frac{P(x_1, x_2)}{P(x_2)}$$

Product rule: $P(x_1, x_2) = p(x_1|x_2)p(x_2)$

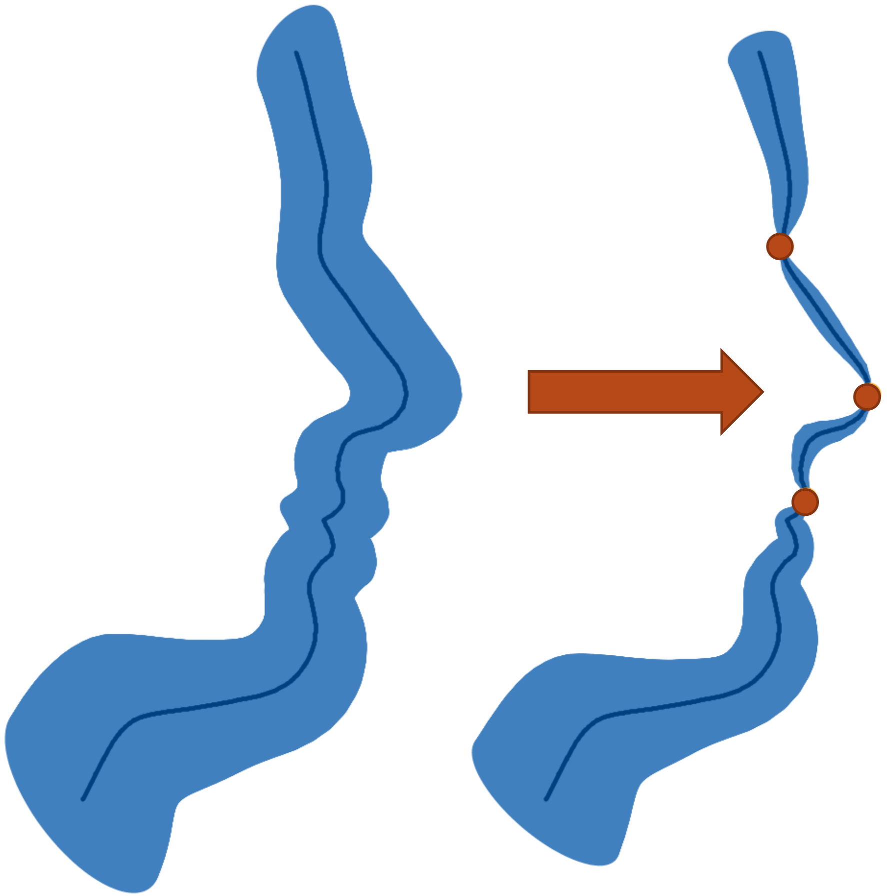
Simplest case: Known observations

- Observations are *known* values
- Distribution of X after *observing* $x_1, \ldots, x_N$:

$$P(X|x_1 \ldots x_N)$$

- Conditional probability

$$P(X|x_1 \ldots x_N) = \frac{P(X, x_1, \ldots, x_N)}{P(x_1, \ldots, x_N)}$$



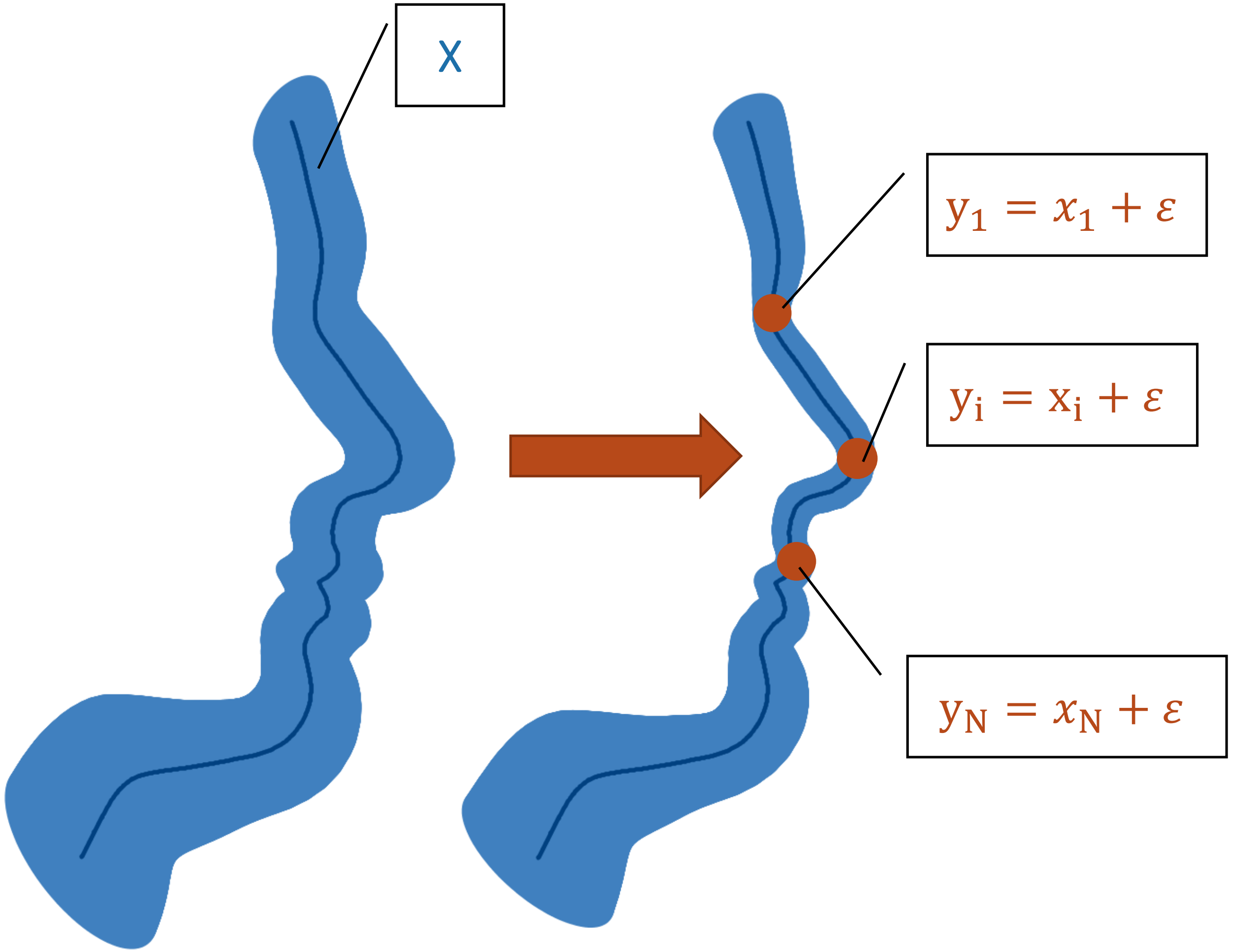
Noisy observations

- Observations are noisy measurements
- Distribution of X after *observing* $y_1, \ldots, y_N$:

$$P(X|y_1 \ldots y_N)$$

- Conditional probability

$$P(X|y_1 \ldots y_N) = \frac{P(X, y, \ldots, y_N)}{P(y_1, \ldots, y_N)}$$



Towards Bayesian Inference

- Update belief about X by *observing* $y_1, \dots, y_N$

$$P(X) \rightarrow P(X|y_1, \dots, y_N)$$

- Factorize joint distribution

$$P(X, y_1, \dots, y_N) = P(y_1, \dots, y_N|X)P(X)$$

- Rewrite conditional distribution

$$P(X|y_1, \dots, y_N) = \frac{P(X, y_1, \dots, y_N)}{P(y_1, \dots, y_N)} = \frac{P(y_1, \dots, y_N|X)P(X)}{P(y_1, \dots, y_N)}$$

More generally: distribution of model points X given data Y:

$$P(X|Y) = \frac{P(X, Y)}{P(Y)} = \frac{P(Y|X)P(X)}{P(Y)}$$

Likelihood

Joint Likelihood Prior

$$P(X,Y) = P(Y|X)P(X)$$

- *Likelihood x prior:* factorization is more flexible than full joint
  - Prior: distribution of core model *without observation*
  - Likelihood: describes how observations are distributed

Bayesian Inference

- Conditional/Bayes rule: method to update *beliefs*

$$\underbrace{P(X|Y)}_{\text{Posterior}} = \frac{\overbrace{P(Y|X)}^{\text{Likelihood}}\overbrace{P(X)}^{\text{Prior}}}{\underbrace{P(Y)}_{\text{Marginal Likelihood}}}$$

- Each observation updates our belief (changes knowledge!)

$$P(X) \rightarrow P(X|Y) \rightarrow P(X|Y,Z) \rightarrow P(X|Y,Z,W) \rightarrow \cdots$$

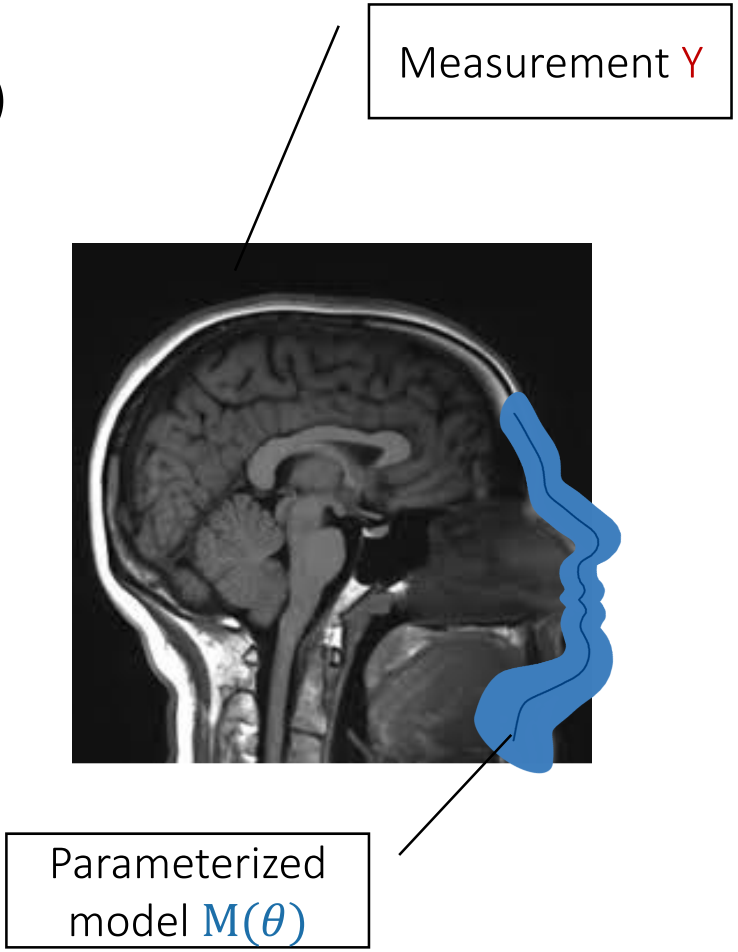
- Bayesian Inference: How beliefs *evolve* with observation
- Recursive: Posterior becomes prior of next inference step

General Bayesian Inference

- Observation of *additional* variables
  - Common case, e.g. image intensities, surrogate measures (size, sex, …)
  - Coupled to core model via likelihood factorization
- General Bayesian inference case:
  - Distribution of data Y
  - Parameters θ

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)} = \frac{P(Y|\theta)P(\theta)}{\int P(Y|\theta)P(\theta)d\theta}$$

$$P(\theta|Y) \propto P(Y|\theta)P(\theta)$$



Summary: Bayesian Inference

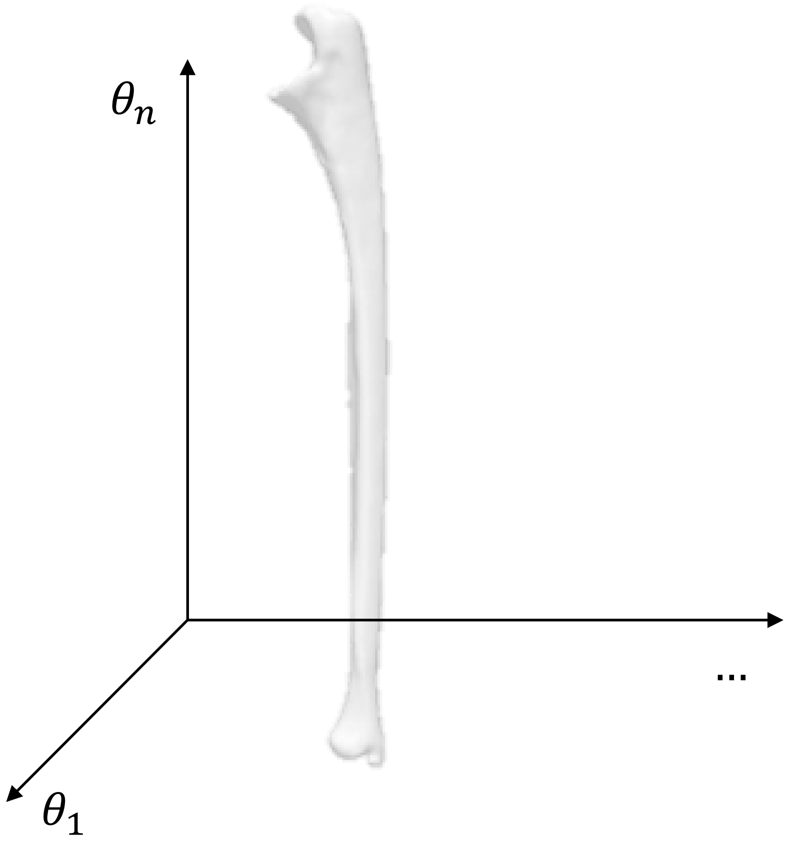
- *Belief*: formal expression of an *observer's knowledge*
  - Subjective state of knowledge about the world
- Beliefs are expressed as *probability* distributions
  - Formally not arbitrary: Consistency requires laws of probability
- *Observations* change knowledge and thus beliefs
- Bayesian inference formally updates *prior beliefs* to *posteriors*
  - Conditional Probability
  - Integration of observation via *likelihood x prior* factorization

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)}$$

Analysis by Synthesis in 5 (simple) steps

Analysis by synthesis in 5 simple steps

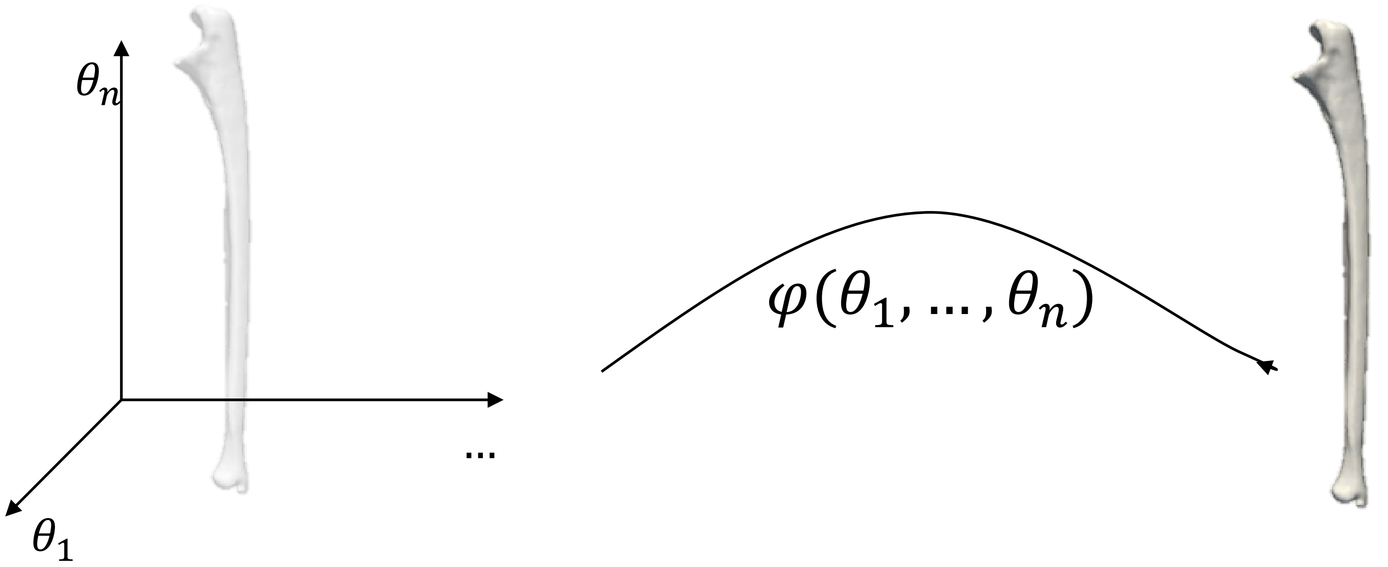
1. Define a parametric model
    - a representation of the world
    - State of the world is determined by parameters $\theta = (\theta_1, \dots, \theta_n)$



Analysis by synthesis in 5 simple steps

2. Define a synthesis function $\varphi(\theta_1, \dots, \theta_n)$

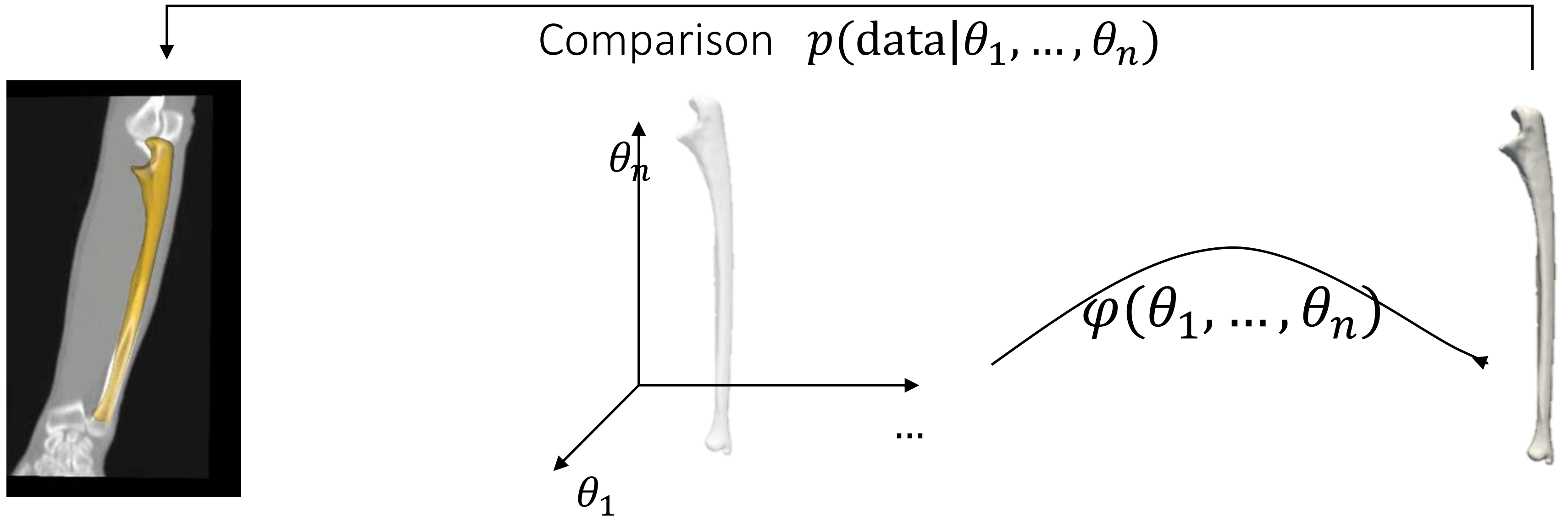
- generates/synthesize the data given the "state of the world"
- φ can be deterministic or stochastic



Analysis by synthesis in 5 simple steps

3. Define likelihood function:

- Define a probabilistic model $p(\text{data}|\theta_1, \ldots, \theta_n)$ that models how the synthesized data compares to the real data
- Includes stochastic factors on the data, such as noise



Bayesian inference

We have: $P(data|\theta_1, \dots, \theta_n)$
We want: $P(\theta_1, \dots, \theta_n|data)$

Bayes rule:

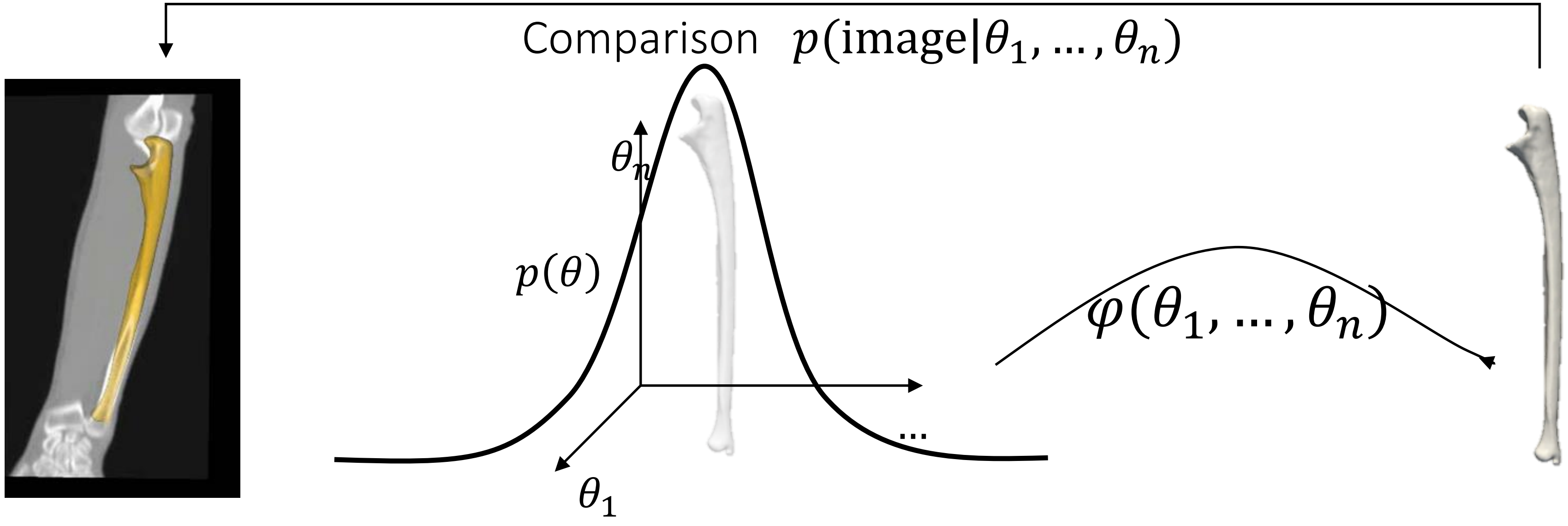
$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

Lets us compute from $p(D|\theta)$ *its "inverse"* $p(\theta|D)$

Analysis by synthesis in 5 simple steps

4. Define prior distribution: $p(\theta) = p(\theta_1, \dots, \theta_n)$
    - Our believe about the "state of the world"

- Makes it possible to invert mapping $p(\text{data}|\theta_1, \dots, \theta_n)$



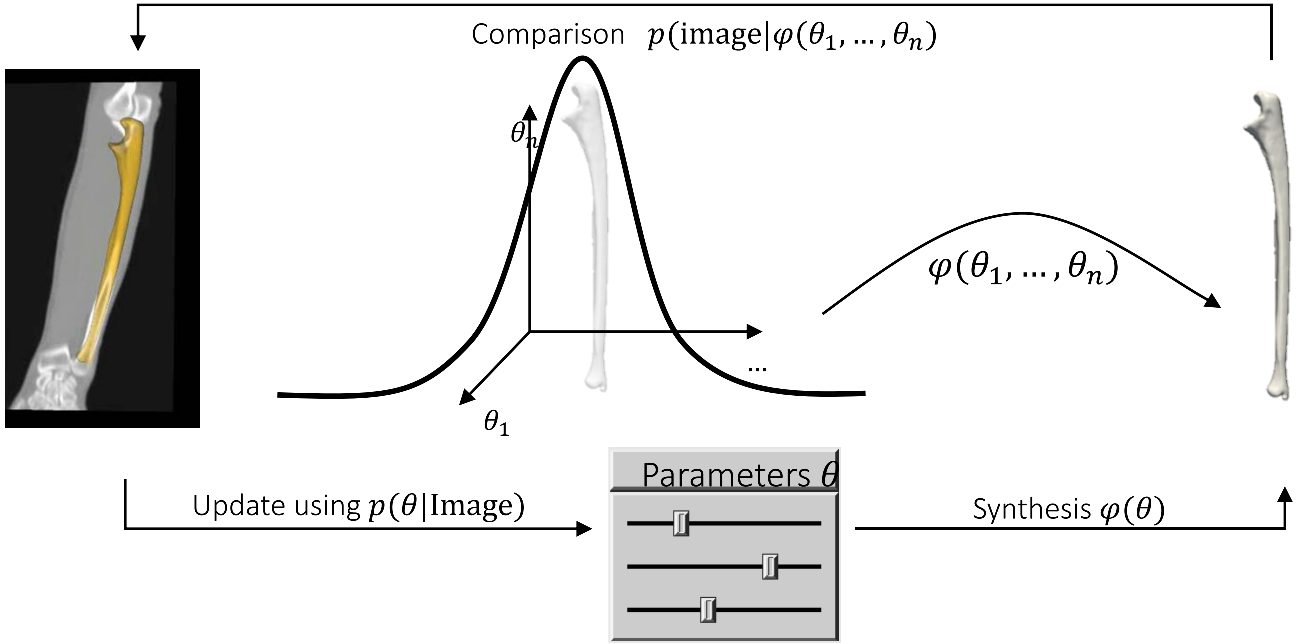
Analysis by synthesis in 5 simple steps

5. Do inference

$$p(\theta_1, \ldots, \theta_n | \text{data}) = \frac{p(\theta_1, \ldots, \theta_n) p(\text{data} | \theta_1, \ldots, \theta_n)}{p(\text{data})}$$

Purely conceptual formulation:

- Independent of algorithmic implementation
- But usually done iteratively



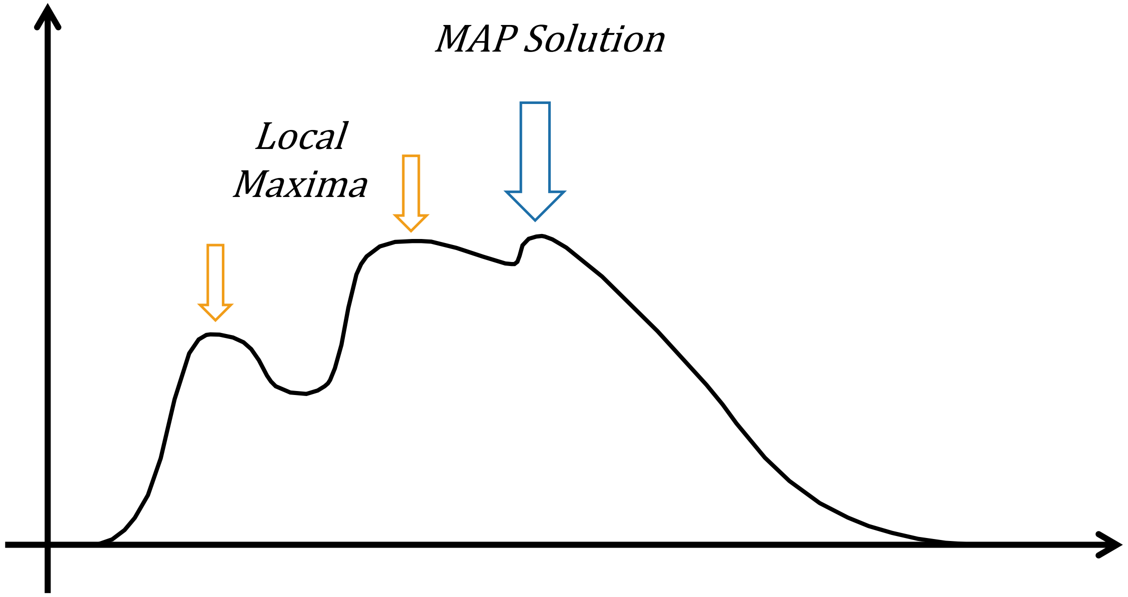
Analysis by synthesis in 5 simple steps

5. Possibility 1: Find best (most likely) solution:

$$\arg\max_{\theta_1,\dots,\theta_n} p(\theta_1, \dots, \theta_n | \text{data}) = \arg\max_{\theta_1,\dots,\theta_n} \frac{p(\theta_1, \dots, \theta_n) p(\text{data} | \theta_1, \dots, \theta_n)}{p(\text{data})}$$

Most popular approach

- Usually based on gradient-descent
- May miss good solutions



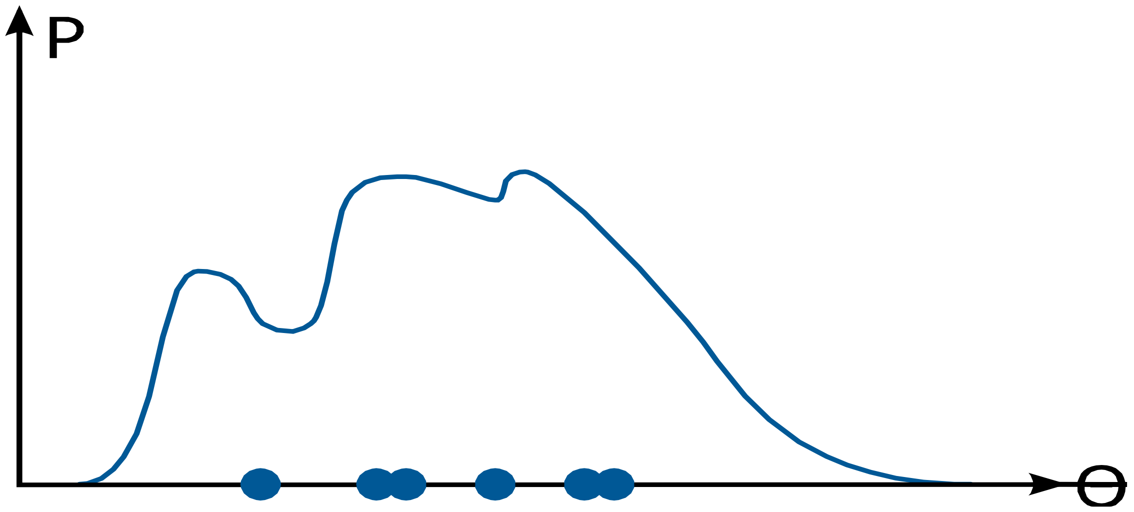
Analysis by synthesis in 5 simple steps

5. Possibility 2: Find posterior distribution:

$$p(\theta_1, \dots, \theta_n | \text{data}) = \frac{p(\theta_1, \dots, \theta_n) p(\text{data} | \theta_1, \dots, \theta_n)}{p(\text{data})}$$

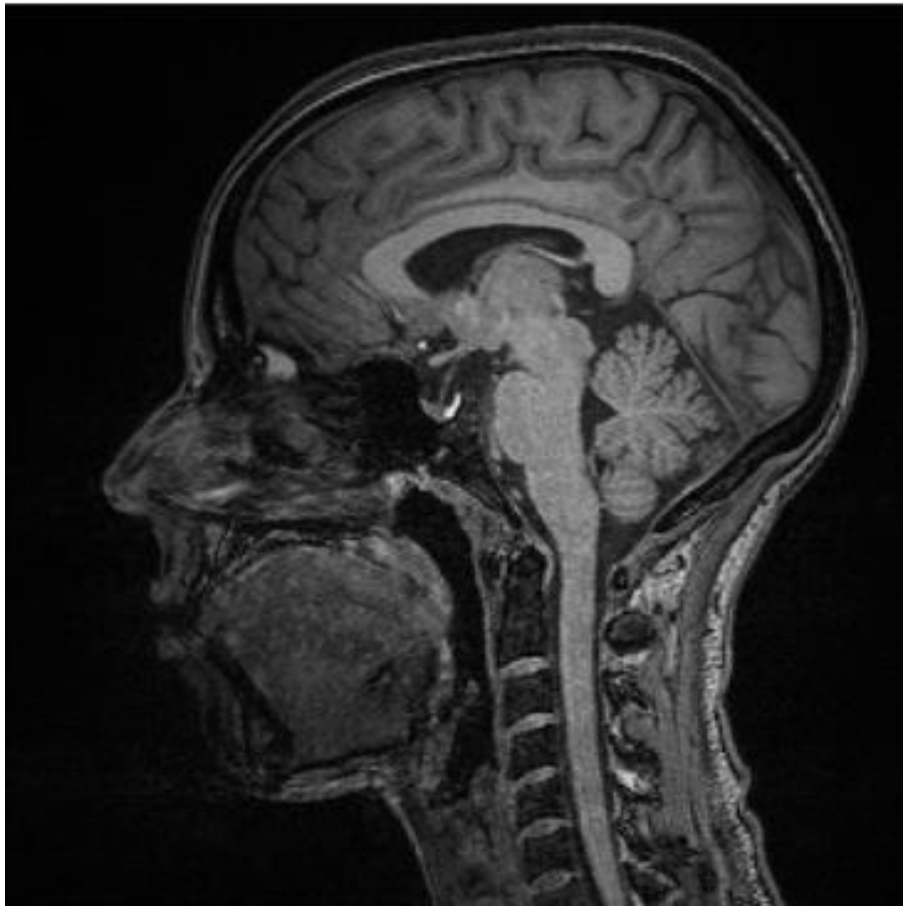
Core of this course

- Obtain samples from the distribution
- Based on Markov Chain Monte Carlo methods



Medical image analysis vs. Computer vision

Images: Medical Image Analysis vs Computer Vision



Source: OneYoungWorld.com

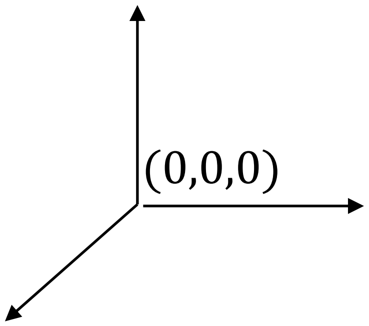
Images in medical image analysis

Goal: Measure and visualize the unseen

- Acquired with specific purpose
  - Controlled measurement
  - Done by experts
- Calibrated, specialized devices

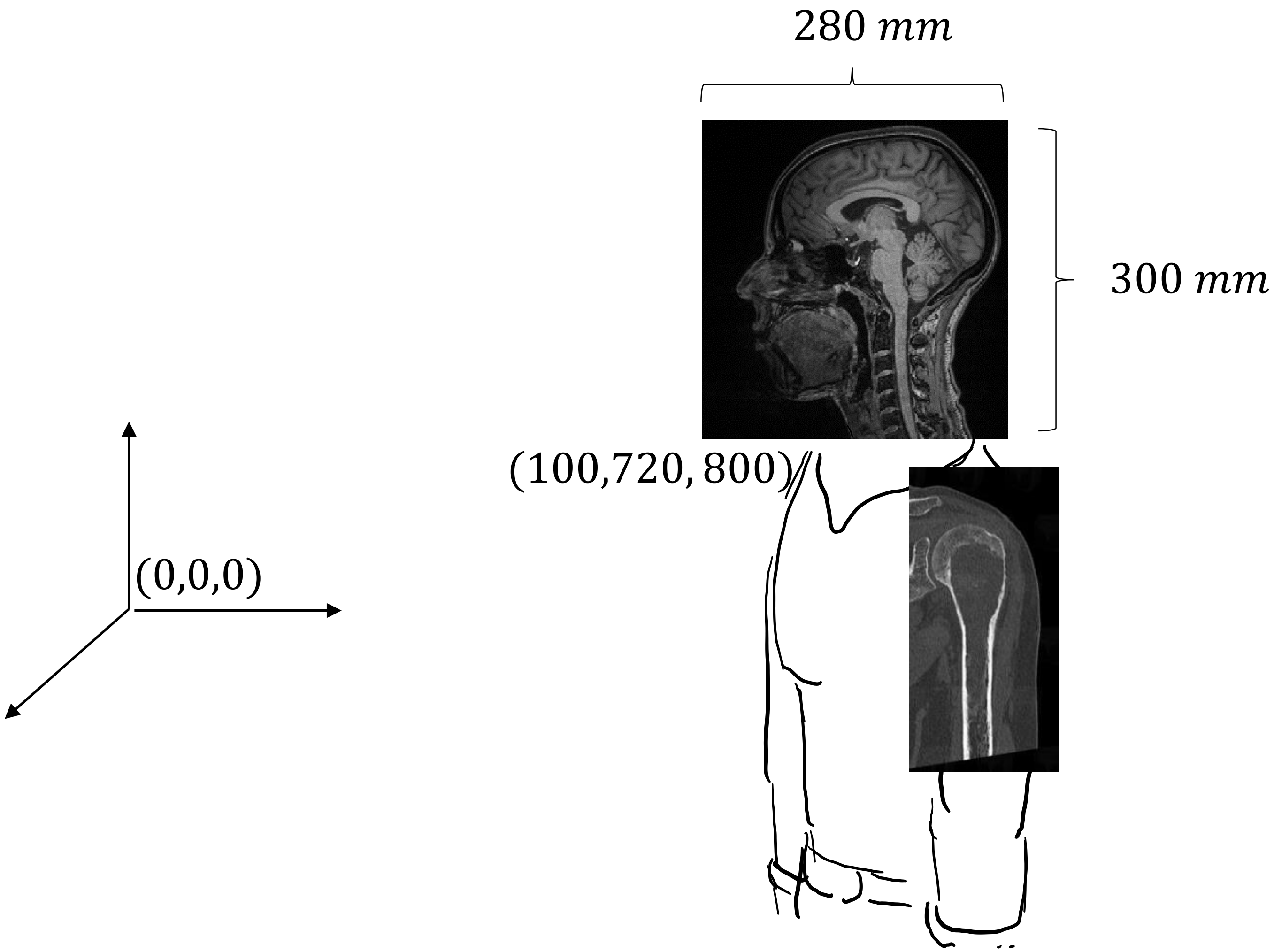
Source: www.siemens.com

Images in medical image analysis



- Images live in a coordinate system (units: mm)

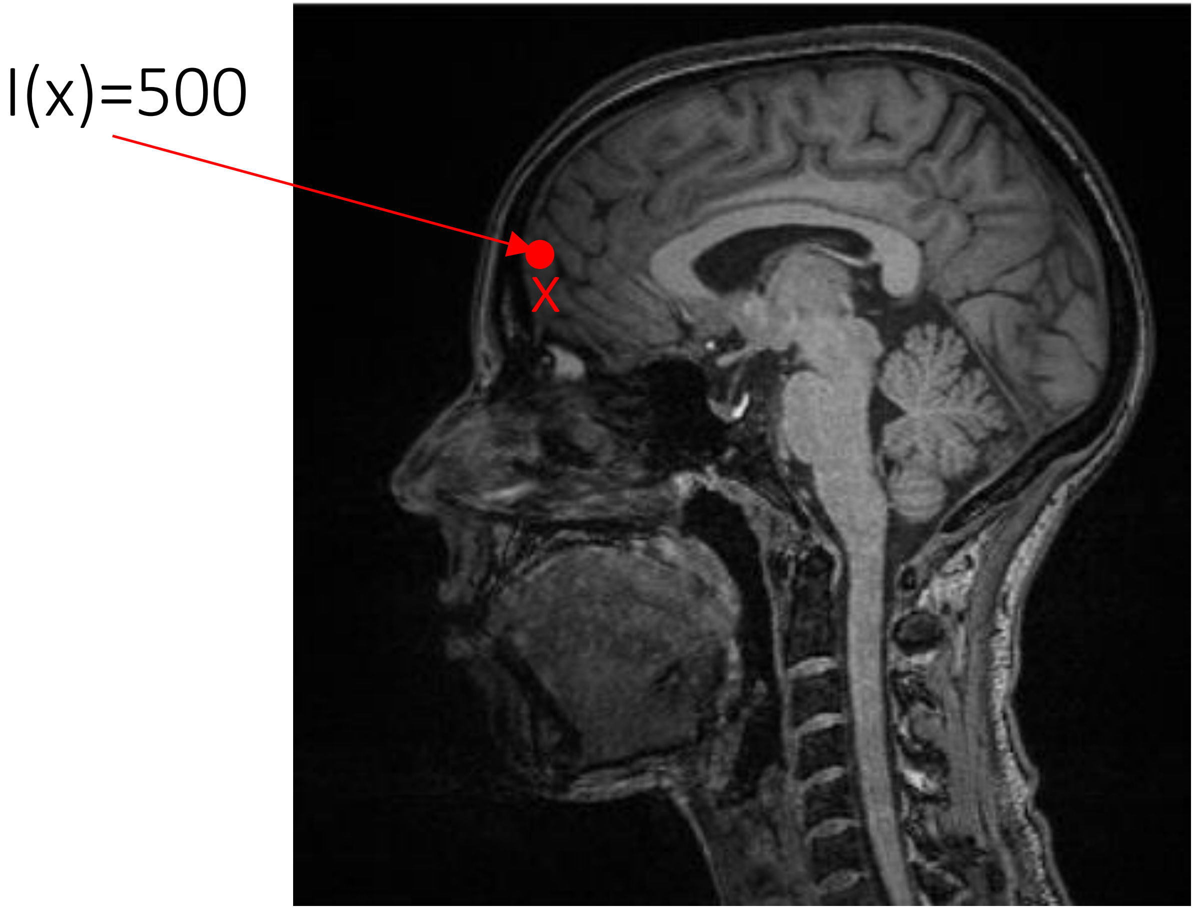
Images in medical image analysis



Images in medical image analysis

Values measure properties of the patient's tissue

- Usually scalar-valued
- Often calibrated
- CT Example:
  -1000 HU -> Air
  3000 HU -> cortical bone



Images in computer vision

Goal: Capture what we see in a realistic way

- Perspective projection from 3D object to 2D image
  - Many parts are occluded

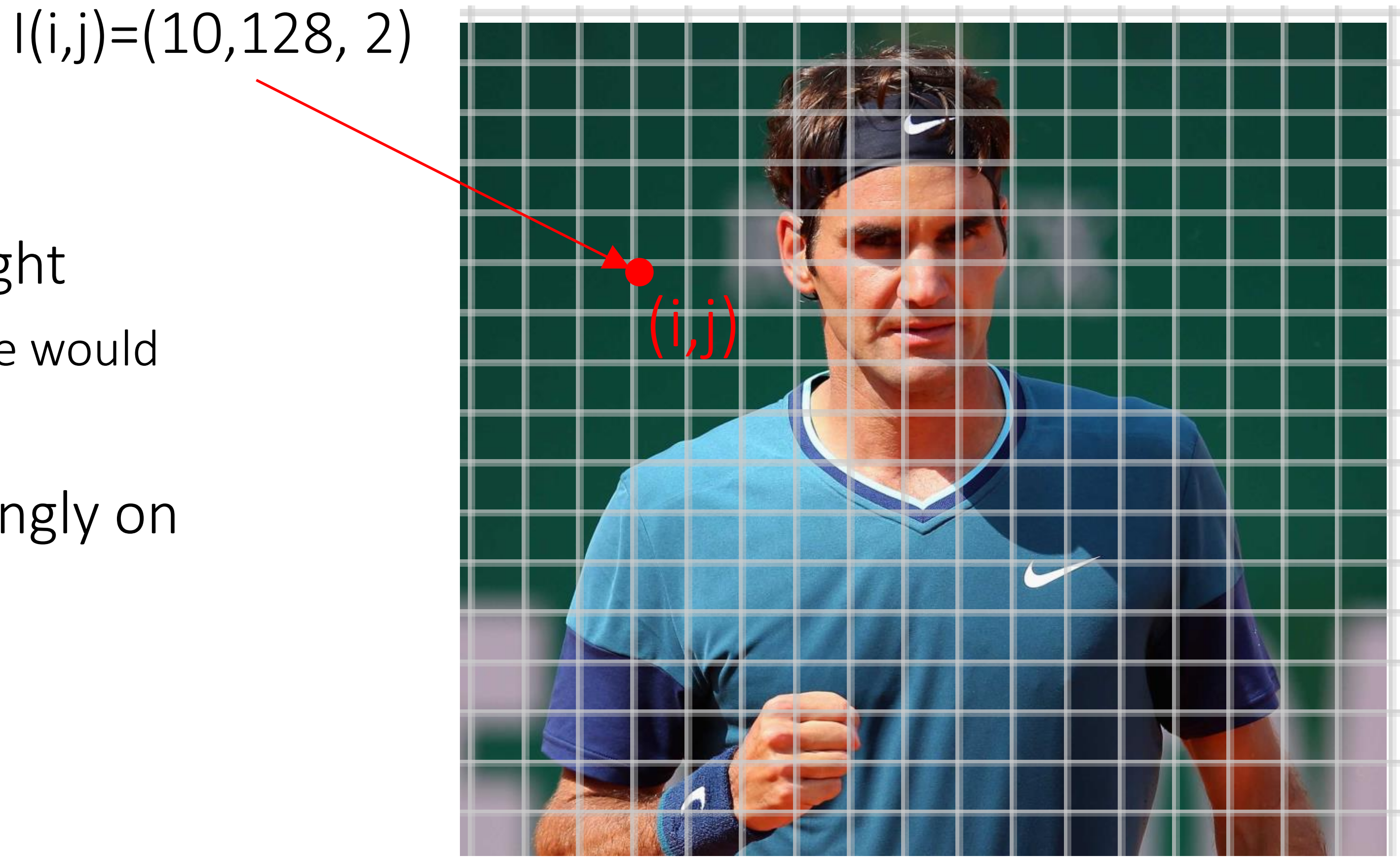
Images in computer vision

- Can be done by anybody
  - Acquisition device usually unknown
  - Uncontrolled background, lighting, …
- No clear scale
  - What is the camera distance?
- No natural coordinate system
  - Unit usually pixel

Source: twitter.com

Images in computer vision

- Pixels represent RGB values
- Values are measurement of light
  - Reproduce what the human eye would see
- Exact RGB value depends strongly on lighting conditions
  - Shadows
  - Ambient vs diffuse light



Images: Medical Image analysis vs Computer Vision

Medical image

- Controlled measurement
- Values have (often) clear interpretation
- Explicit setup to visualize unseen
- Coordinate system with clear scale

Computer vision

- Uncontrolled snapshot
- Values are mixture of different (unknown factors)
- Many occlusion due to perspective
- Scale unknown

Many complications of computer vision arise in different form also in a medical setting.

Structure in images

- Not t
  the ir
  - H
    n
- Imag
  worl
  - St
    p

Our mission:

- Model this structure
  - Needs only few parameters
- Explain image by finding appropriate parameters that reflect objects / laws / processes