Statistical learning in the nervous system/ Intelligent systems
Probabilistic models

Gergő Orbán
golab.wigner.mta.hu

Course layout

Introduction of the mathematical framework

Probabilistic models

Probabilistic models in cognition: Bayesian behaviour

Approximate methods

Approximate methods in cognition: Sampling

Discovering the probabilistic models used by humans

Using learned statistical regularities: coding and compression

Theory of coding and compression in memory processes

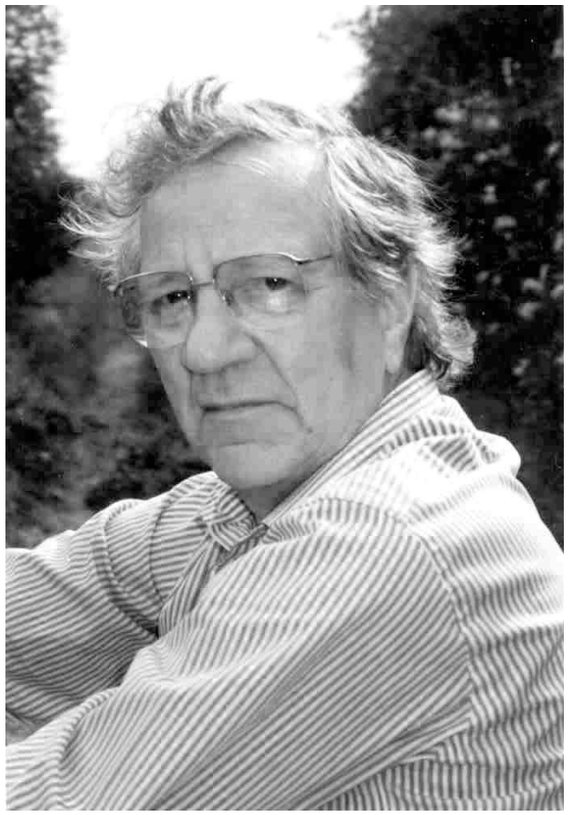
Effective probabilistic models: nonparametric models

Nonparametric models in cognition: memory dynamics

Learning the structure of probabilistic models

Structure learning in cognition: causal learning

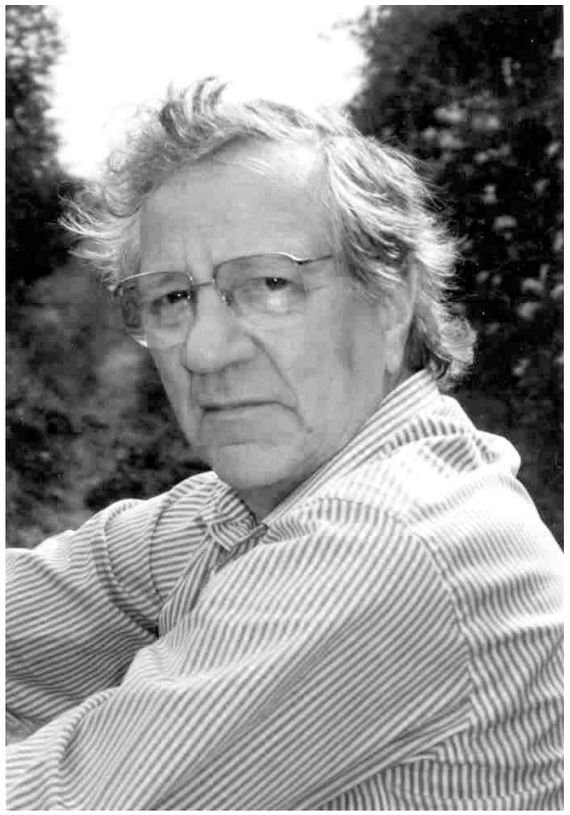
RECAP: rational analysis



Horace Barlow
Trinity College,
Cambridge

A wing would be the most mystifying structure if one did not know that birds flew. One might observe that it could be extended a considerable distance, that it had smooth covering of feathers with conspicuous markings, that it was operated by powerful muscles, and that strength and lightness were prominent features of its constructions. These are important facts, but in themselves they do not tell us that birds fly. Yet, without knowing this and without understanding something of the principles of flight, a more detailed examination of the wing itself would probably be unrewarding.

RECAP: rational analysis



Horace Barlow
Trinity College,
Cambridge

A wing would be the most mystifying structure if one did not know that birds flew. One might observe that it could be extended a considerable distance, that it had smooth covering of feathers with conspicuous markings, that it was operated by powerful muscles, and that strength and lightness were prominent features of its constructions. These are important facts, but in themselves they do not tell us that birds fly. Yet, without knowing this and without understanding something of the principles of flight, a more detailed examination of the wing itself would probably be unrewarding.

Sensory communication (1962) MIT Press

RECAP: deductive reasoning

- propositional logic: reasoning can be systematically performed

logic function: new propositions using basic operations (*conjunction, disjunction, negation, implication*)

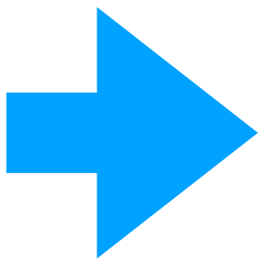
$$C = f(A, B) = (A + \overline{B})(\overline{A} + A\overline{B}) + \overline{A}B(A + B)$$

RECAP: deductive reasoning

- propositional logic: reasoning can be systematically performed
- a finite set of operations is sufficient to cover the full proposition space

disjunctive normal form

A, B	TT	TF	FT	FF
$f_1(A, B)$	T	F	F	F
$f_2(A, B)$	F	T	F	F
$f_3(A, B)$	F	F	T	F
$f_4(A, B)$	F	F	F	T

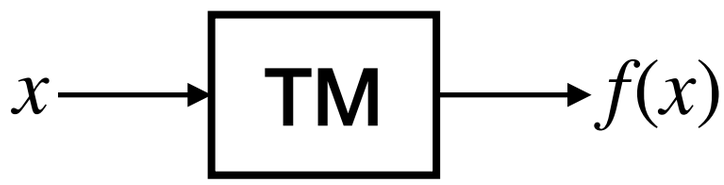


arbitrary f (A, B) can be composed

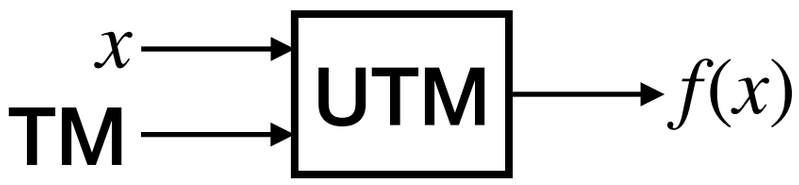
RECAP: deductive reasoning

- propositional logic: reasoning can be systematically performed
- a finite set of operations is sufficient to cover the full proposition space
- rational analysis can be performed in a world where truth values of statements can be established

Turing machine
an agent can be constructed that can perform an arbitrary function



Universal Turing machine
irrespective of the language used one can construct such computing machines



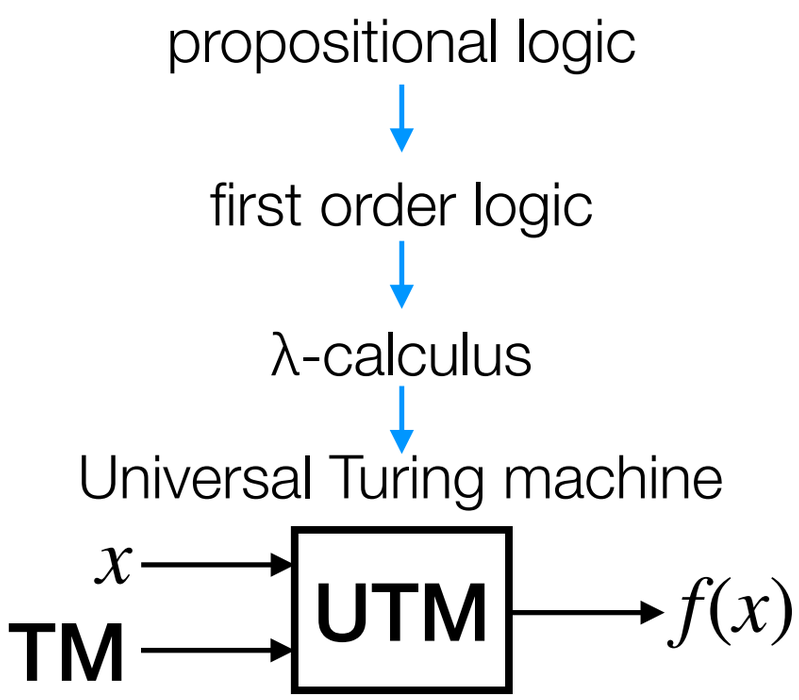
RECAP: deductive reasoning

- propositional logic: reasoning can be systematically performed
- a finite set of operations is sufficient to cover the full proposition space
- rational analysis can be performed in a world where truth values of statements can be established

propositional logic

RECAP: deductive reasoning

- propositional logic: reasoning can be systematically performed
- a finite set of operations is sufficient to cover the full proposition space
- rational analysis can be performed in a world where truth values of statements can be established



RECAP: deductive reasoning

- propositional logic: reasoning can be systematically performed
- a finite set of operations is sufficient to cover the full proposition space
- rational analysis can be performed in a world where truth values of statements can be established

extension:
predicates
logical connectives
quantifiers

propositional logic

↓

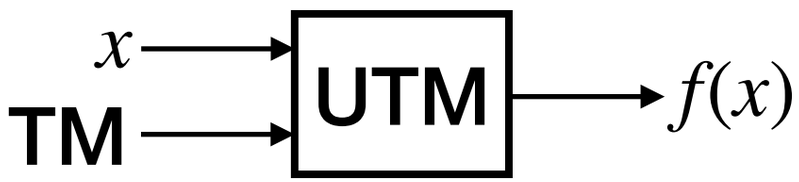
first order logic

↓

λ-calculus

↓

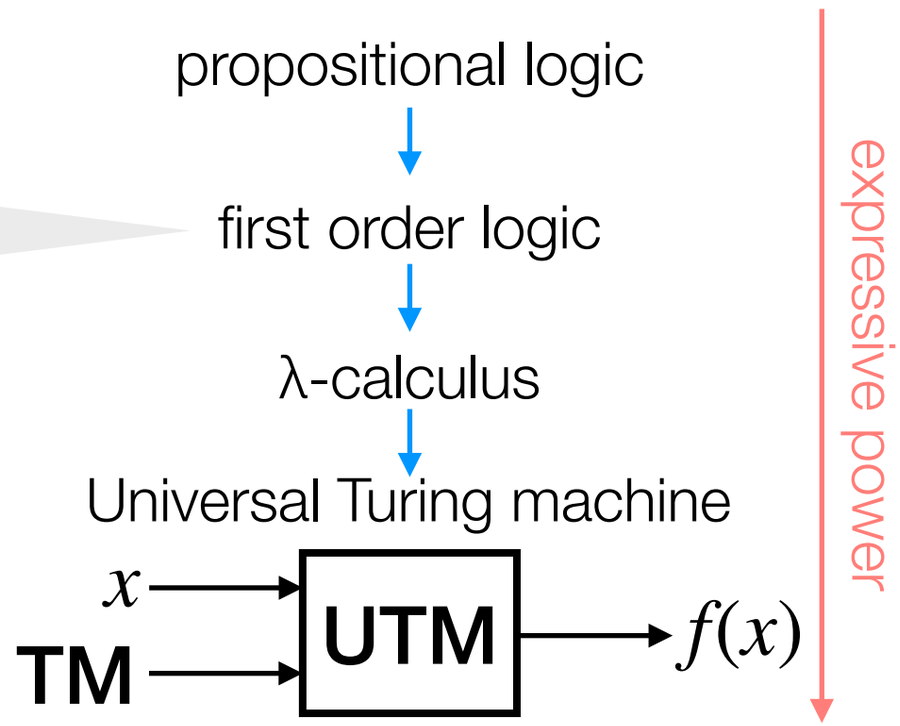
Universal Turing machine



RECAP: deductive reasoning

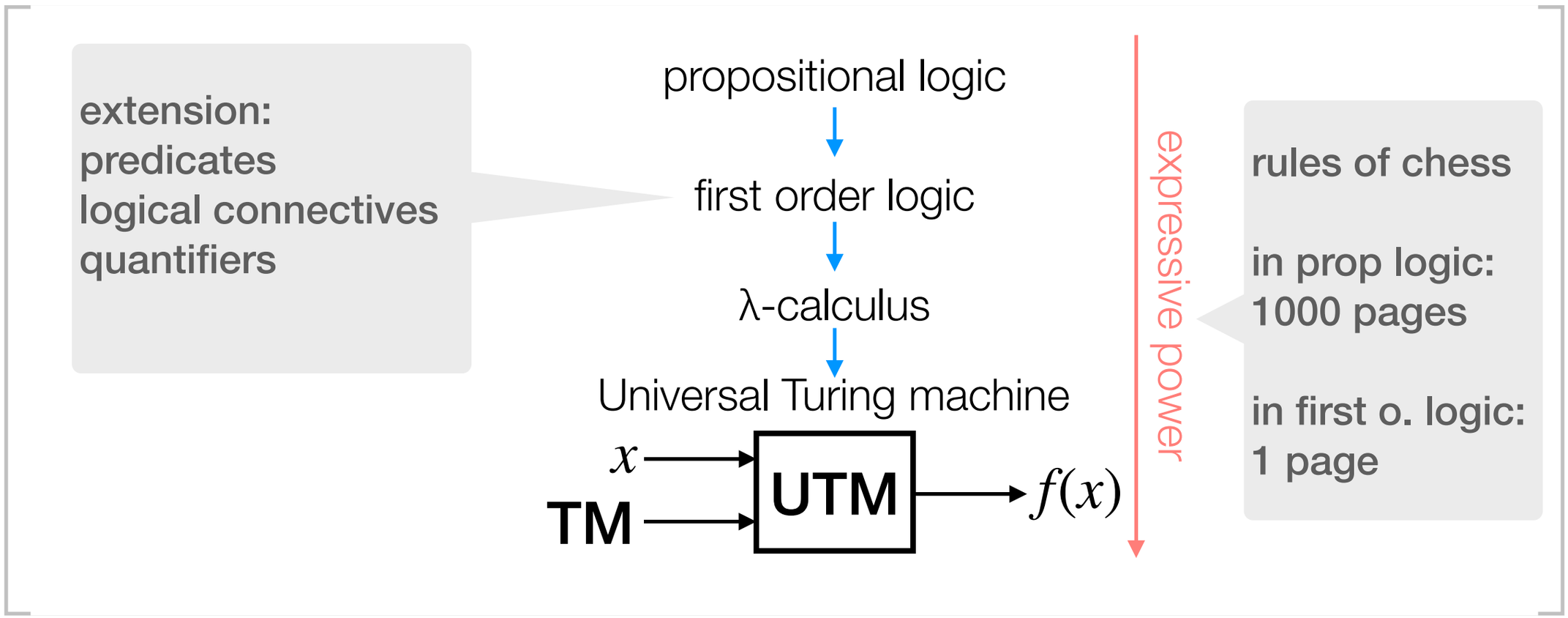
- propositional logic: reasoning can be systematically performed
- a finite set of operations is sufficient to cover the full proposition space
- rational analysis can be performed in a world where truth values of statements can be established

extension:
predicates
logical connectives
quantifiers



RECAP: deductive reasoning

- propositional logic: reasoning can be systematically performed
- a finite set of operations is sufficient to cover the full proposition space
- rational analysis can be performed in a world where truth values of statements can be established



Is this the whole story? — Challenges

- Scope of deductive reasoning is limited:
  we need to make inferences in uncertain situations

Is this the whole story? — Challenges

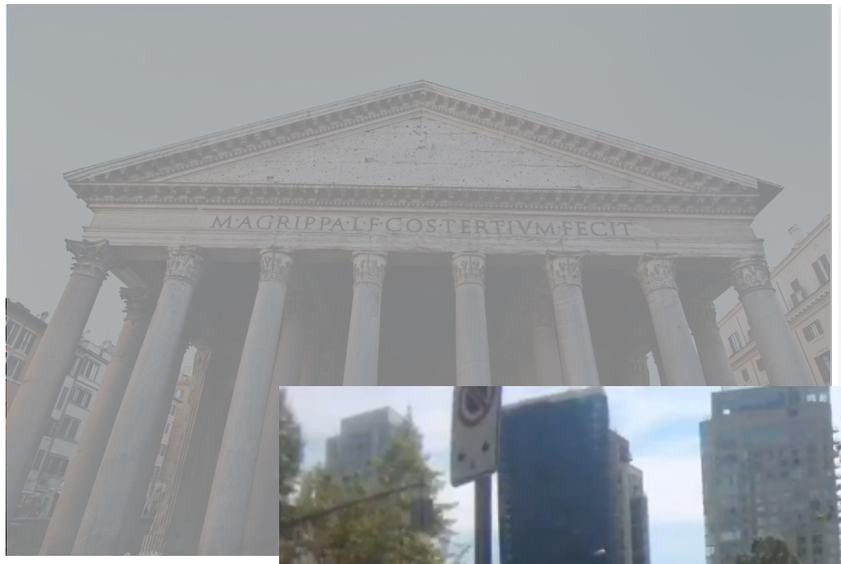
- Scope of deductive reasoning is limited:
  we need to make inferences in uncertain situations

Is this the whole story? — Challenges

- Scope of deductive reasoning is limited:
  we need to make inferences in uncertain situations

Is this the whole story? — Challenges

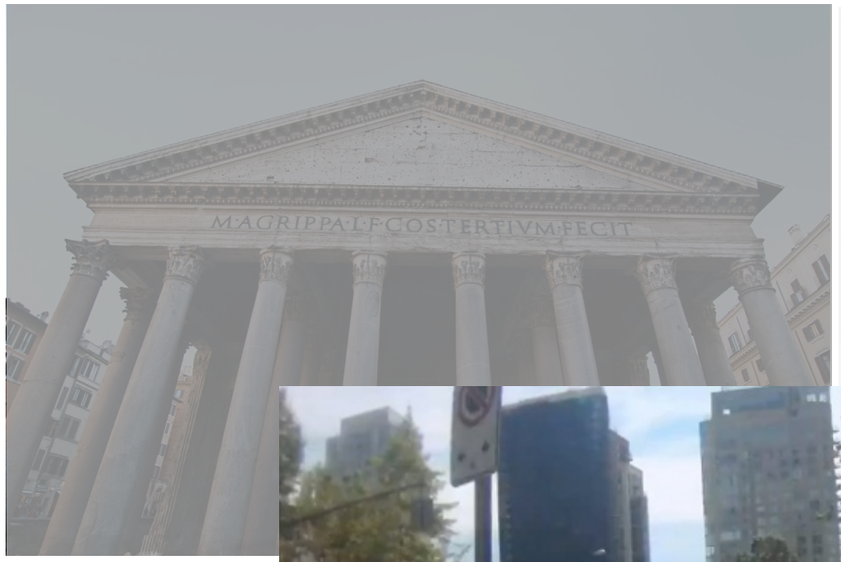
- Scope of deductive reasoning is limited:
  we need to make inferences in uncertain situations



Is this the whole story? — Challenges

- Scope of deductive reasoning is limited:
  we need to make inferences in uncertain situations

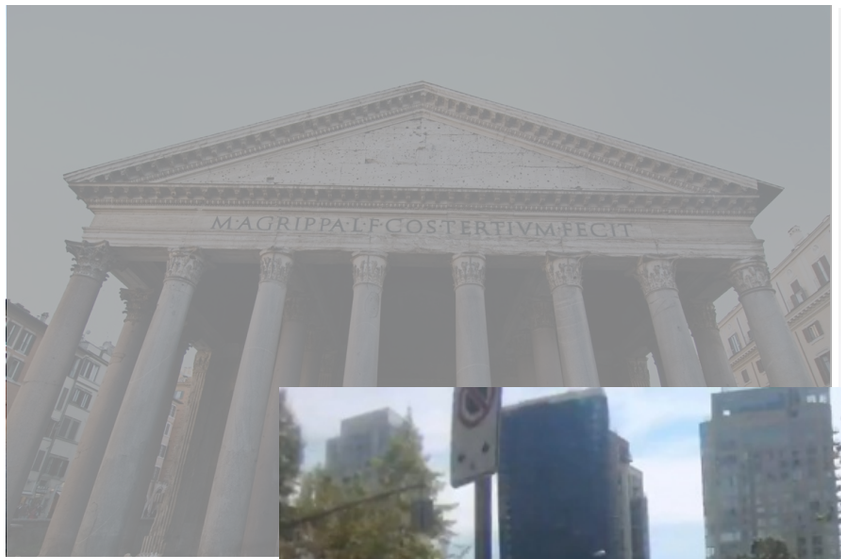
"The girl saw the boy with the telescope"

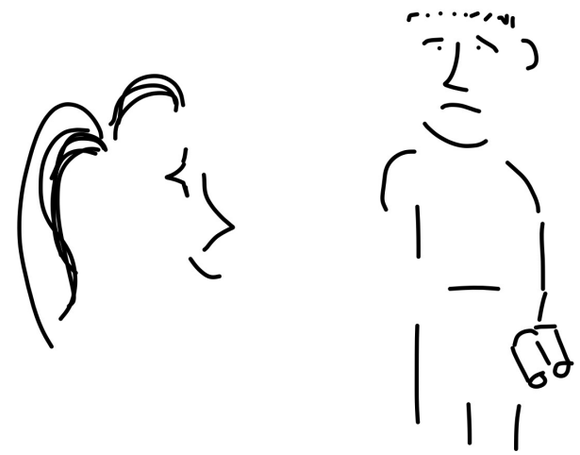


Is this the whole story? — Challenges

- Scope of deductive reasoning is limited:
  we need to make inferences in uncertain situations

"The girl saw the boy with the telescope"

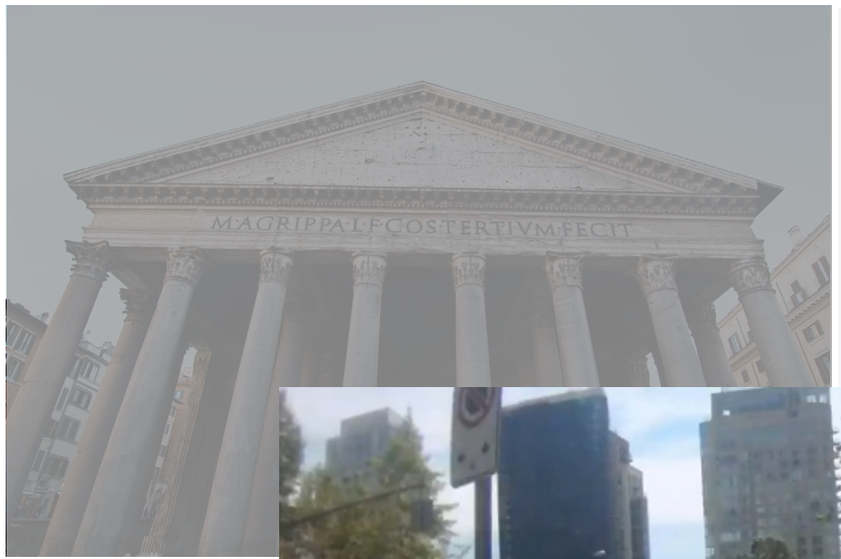


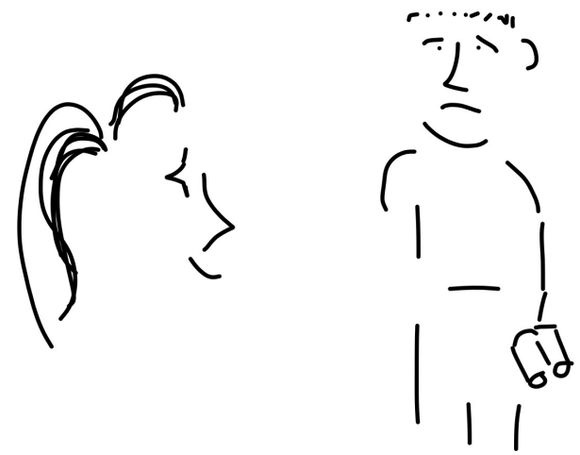


Is this the whole story? — Challenges

- Scope of deductive reasoning is limited:
  we need to make inferences in uncertain situations

"The girl saw the boy with the telescope"





Is this the whole story? — Challenges

- Scope of deductive reasoning is limited:
  we need to make inferences in uncertain situations
- It is unclear how to construct adaptive forms of reasoning:
  training is hard, hand engineering is required

Rational analysis revisited

Rational analysis revisited

- Are there ways to formalise plausible reasoning?

Rational analysis revisited

- Are there ways to formalise plausible reasoning?
- Is there a rational way to approach this problem?

Rational analysis revisited

- Are there ways to formalise plausible reasoning?
- Is there a rational way to approach this problem?
- What exactly determines how much the observation of Aristotle's reading skills alters the plausibility of him being a philosopher?

Rational analysis revisited

- Are there ways to formalise plausible reasoning?
- Is there a rational way to approach this problem?
- What exactly determines how much the observation of Aristotle's reading skills alters the plausibility of him being a philosopher?
- In deductive reasoning, the veracity of statements can be identified after exhaustive chains of inferences. In plausible reasoning uncertainty grows with subsequent iterations. Can this be characterised?

Rational analysis revisited

- Are there ways to formalise plausible reasoning?
- Is there a rational way to approach this problem?
- What exactly determines how much the observation of Aristotle's reading skills alters the plausibility of him being a philosopher?
- In deductive reasoning, the veracity of statements can be identified after exhaustive chains of inferences. In plausible reasoning uncertainty grows with subsequent iterations. Can this be characterised?
- Discovering an efficient math behind plausible reasoning can provide a useful tool to model cognitive processes

Rational analysis revisited

- Are there ways to formalise plausible reasoning?
- Is there a rational way to approach this problem?
- What exactly determines how much the observation of Aristotle's reading skills alters the plausibility of him being a philosopher?
- In deductive reasoning, the veracity of statements can be identified after exhaustive chains of inferences. In plausible reasoning uncertainty grows with subsequent iterations. Can this be characterised?
- Discovering an efficient math behind plausible reasoning can provide a useful tool to model cognitive processes
- What does this do with 'weakly' rational behaviour at Wason's card selection task?

Common sense: Plausible reasoning

weak syllogisms

if A is true then B is true

B is true

---

therefore, A becomes more plausible

if A is true then B is true

A is false

---

therefore, B becomes less plausible

Common sense: Plausible reasoning

weak syllogisms

if A is true then B is true	if A is true then B is true
B is true	A is false
therefore, A becomes more plausible	therefore, B becomes less plausible

- The truth value of many statements cannot be fully determined

Common sense: Plausible reasoning

weak syllogisms

if A is true then B is true	if A is true then B is true
B is true	A is false
therefore, A becomes more plausible	therefore, B becomes less plausible

- The truth value of many statements cannot be fully determined
- Reasoning in uncertain situations is both useful and possible:
  we cannot rely on deductive reasoning

Common sense: Plausible reasoning

weak syllogisms

if A is true then B is true

B is true

---

therefore, A becomes more plausible

if A is true then B is true

A is false

---

therefore, B becomes less plausible

- The truth value of many statements cannot be fully determined
- Reasoning in uncertain situations is both useful and possible:
  we cannot rely on deductive reasoning
- We do it automatically while making inferences,
  terming it *common sense*

Common sense: Plausible reasoning

weak syllogisms

if A is true then B is true

B is true

---

therefore, A becomes more plausible

if A is true then B is true

A is false

---

therefore, B becomes less plausible

- The truth value of many statements cannot be fully determined
- Reasoning in uncertain situations is both useful and possible:
  we cannot rely on deductive reasoning
- We do it automatically while making inferences,
  terming it *common sense*
- Importantly, common sense is not an arbitrary relaxation of rules:
  George Pólya formalised a compact set of desiderata for reasoning

Artificial agent revisited

- We are pursuing common sense reasoning to understand thinking

- This provides better tools to build ‘machines that think’, i.e. better AI

- The rational agent developed provides a normative approach to human behaviour, in other words:

- ‘rules will be deduced from simple desiderata which, it appears to us, would be desirable in human brains; i.e. we think that a rational person, on discovering that they were violating one of these desiderata, would wish to revise their thinking’, *Jaynes*

Plausibilities

- Aristotle wants to learn about the effectiveness of his teaching

what is the truth value of $S \rightarrow L$?

possible experiment outcomes

scribbling	excited look	learning	possible?
1	1	1	1
1	1	0	1
0	1	1	1
1	0	0	0
0	1	1	0
0	1	0	0
0	0	1	0
0	0	0	1

Plausibilities

- Aristotle wants to learn about the effectiveness of his teaching

what is the truth value of $S \rightarrow L$?

possible experiment outcomes

scribbling	excited look	learning	possible?
1	1	1	1
1	1	0	1
0	1	1	1
1	0	0	0
0	1	1	0
0	1	0	0
0	0	1	0
0	0	0	1

How to establish numerical values for plausibilities?

Rational desiderata

- The proposition needs to be assigned a plausibility based on available evidence: the conditional plausibility that A is true, given that B is true, $A \mid B$

- The richness of real numbers is required by theory
- Convention: larger plausibility corresponds to higher numbers

combining predicaments:

the plausibility that A is true, given that both B and C are true:

$$A \mid BC$$

the plausibility that at least one of the propositions A and B is true, given that both C and D are true:

$$A + B \mid CD$$

1. Level of certainty is expressed with real numbers

Rational desiderata

- Qualitative match of the formalism with common sense is required
- If $(A|C') > (A|C)$
  and $(B|AC') = (B|AC)$
  then $(AB|C') \geq (AB|C) \quad \& \quad (\overline{A}|C') < (\overline{A}|C)$
- Provides '*sense of direction*'

1. Level of certainty is expressed with real numbers

Rational desiderata

- Qualitative match of the formalism with common sense is required
- If $(A|C') > (A|C)$
  and $(B|AC') = (B|AC)$
  then $(AB|C') \geq (AB|C)$ & $(\overline{A}|C') < (\overline{A}|C)$
- Provides '*sense of direction*'

1. Level of certainty is expressed with real numbers
2. Qualitative correspondence with common sense

Rational desiderata

- 3.a, Alternative ways to achieve conclusion lead to the same result
- 3.b, All evidence is taken into account by the agent
- 3.c, Equivalent state of knowledge is represented by equal level of plausibility

1. Level of certainty is expressed with real numbers
2. Qualitative correspondence with common sense
3. Consistency

Cox's theorem

Fundamental laws of probability (Kolmogorov's axioms) directly come from the three desiderata

- The three desiderata uniquely determine the mathematical form of representation of plausibility

- Plausibility has to behave as probability

Kolmogorov axioms

1. Probability is a real number between 0 and 1
2. Certainty is represented by $P(A \mid B) = 1$
   Certain falsehood: $P(A \mid B) = 0$
3. $P(A \mid B) + P(\overline{A} \mid B) = 1$

+1 Conjunction: $P(AB \mid C) = P(A \mid BC)P(B \mid C)$

Cox's theorem

Fundamental laws of probability (Kolmogorov's axioms) directly come from the three desiderata

- The three desiderata uniquely determine the mathematical form of representation of plausibility

- Plausibility has to behave as probability

Kolmogorov axioms

1. Probability is a real number between 0 and 1
2. Certainty is represented by $P(A \mid B) = 1$
   Certain falsehood: $P(A \mid B) = 0$
3. $P(A \mid B) + P(\overline{A} \mid B) = 1$

+1 Conjunction: $P(AB \mid C) = P(A \mid BC)P(B \mid C)$

We can use these to express operations in plausible reasoning

Cox's theorem

Fundamental laws of probability (Kolmogorov's axioms) directly come from the three desiderata

- The three desiderata uniquely determine the mathematical form of representation of plausibility

- Plausibility has to behave as probability

Kolmogorov axioms

1. Probability is a real number between 0 and 1
2. Certainty is represented by $P(A \mid B) = 1$
   Certain falsehood: $P(A \mid B) = 0$
3. $P(A \mid B) + P(\overline{A} \mid B) = 1$

+1 Conjunction: $P(AB \mid C) = P(A \mid BC)P(B \mid C)$

We can use these to express operations in plausible reasoning

It will have similar power than deductive reasoning, but extends it

Dutch book argument

- Assume that we are willing to take bets with odds proportional to the probability of the occurrence of events
- If the plausibilities are not behaving according to the rules of probabilities then we will certainly loose money against someone using that strategy

Dutch book argument

- Assume that we are willing to take bets with odds proportional to the probability of the occurrence of events
- If the plausibilities are not behaving according to the rules of probabilities then we will certainly loose money against someone using that strategy

In an environment where truth values of statements cannot fully determined, a probabilistic description is lucrative

Probability table

possible experiment outcomes

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
…	…	…	
0	0	0	0.43

$$\sum P(A, E, L) = 1$$

Probability table

possible experiment outcomes

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
...	...	...	
0	0	0	0.43

$$\sum P(A, E, L) = 1$$

The probability table fully characterises the system

Probability table

possible experiment outcomes

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
...	...	...	
0	0	0	0.43

$$\sum P(A, E, L) = 1$$

The probability table fully characterises the system

Probability table

possible experiment outcomes

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
...	...	...	
0	0	0	0.43

$$\sum P(A, E, L) = 1$$

The probability table fully characterises the system

Instead of truth tables, we extended the scope of propositions

Inference & conditional probability

Goal: infer the probability of learning given the disciple wrote feverishly

$$P(L \mid s, \bar{e}) = ?$$

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
…	…	…	
0	0	0	0.43

Inference & conditional probability

Stochastic variable S taking value s

Goal: infer the probability of l … erishly

$$P(L \mid s, \bar{e}) = ?$$

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
…	…	…	
0	0	0	0.43

Inference & conditional probability

Goal: infer the probability of learning given the disciple wrote feverishly

$$P(L \mid s, \bar{e}) = ?$$

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
…	…	…	
0	0	0	0.43

Inference & conditional probability

Goal: infer the probability of learning given the disciple wrote feverishly

$$P(L \mid s, \bar{e}) = ?$$

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
…	…	…	
0	0	0	0.43

Inference & conditional probability

Goal: infer the probability of learning given the disciple wrote feverishly

$$P(L \mid s, \bar{e}) = ?$$

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
…	…	…	
0	0	0	0.43

Inference & conditional probability

Goal: infer the probability of learning given the disciple wrote feverishly

$$P(L \mid s, \bar{e}) = ?$$

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
…	…	…	
0	0	0	0.43

$\sum P_i = 1$

Algebra on probabilities?

George Boole, 1815-1864

Propositions: A, B, C

A B — *logical product* or *conjunction*

'both of the propositions, A, B are true'

A + B — *logical sum* or *disjunction*

'at least one of the propositions, A, B is true'

Sum rule

scribbling	learning	P(S,L)
1	1	0.45
1	0	0.05
0	1	0.15
0	0	0.35

$$\mathrm{P}(l) = \mathrm{P}(s, l) + \mathrm{P}(\bar{s}, l)$$

Sum rule

scribbling	learning	P(S,L)
1	1	0.45
1	0	0.05
0	1	0.15
0	0	0.35

Stochastic variable L taking value l

$$\mathrm{P}(l) = \mathrm{P}(s, l) + \mathrm{P}(\bar{s}, l)$$

Sum rule

scribbling	learning	P(S,L)
1	1	0.45
1	0	0.05
0	1	0.15
0	0	0.35

Stochastic variable L taking value l

$$\mathrm{P}(l) = \mathrm{P}(s, l) + \mathrm{P}(\bar{s}, l)$$

P the disciple has learned the course material

P she learned *and* was scribbling

OR

P she learned *and* was not scribbling

Sum rule

scribbling	learning	P(S,L)
1	1	0.45
1	0	0.05
0	1	0.15
0	0	0.35

+

$P(l) = 0.6$

Stochastic variable L taking value l

$$P(l) = P(s, l) + P(\bar{s}, l)$$

P the disciple has learned the course material

P she learned *and* was scribbling

OR

P she learned *and* was not scribbling

Sum rule

scribbling	learning	P(S,L)
1	1	0.45
1	0	0.05
0	1	0.15
0	0	0.35

+

Stochastic variable L taking value l

$$\mathrm{P}(l) = \mathrm{P}(s, l) + \mathrm{P}(\bar{s}, l)$$

P the disciple has learned the course material

$\mathrm{P}(l) = 0.6$

$\mathrm{P}(\bar{l}) = 0.4$

P she learned *and* was scribbling

OR

P she learned *and* was not scribbling

Sum rule

scribbling	learning	P(S,L)
1	1	0.45
1	0	0.05
0	1	0.15
0	0	0.35

+

Stochastic variable L taking value l

$$\mathrm{P}(l) = \mathrm{P}(s, l) + \mathrm{P}(\bar{s}, l)$$

P the disciple has learned the course material

$\mathrm{P}(l) = 0.6$

$\mathrm{P}(\bar{l}) = 0.4$

P she learned *and* was scribbling

OR

P she learned *and* was not scribbling

in general cases when stochastic variables can take more values:

$$\mathrm{P}(x) = \sum_{y \in \mathrm{Y}} \mathrm{P}(x, y)$$

also known as marginalisation

Product rule

$$\mathrm{P}(s, l) = \mathrm{P}(l \mid s)\mathrm{P}(s)$$

Product rule

This turned up at the Cox theorem,
being a consequence of the *Consistency* desideratum

$$\mathsf{P}(s, l) = \mathsf{P}(l \mid s)\mathsf{P}(s)$$

Product rule

This turned up at the Cox theorem,
being a consequence of the *Consistency* desideratum

$$\mathsf{P}(s, l) = \mathsf{P}(l \mid s)\mathsf{P}(s)$$

P the disciple has scribbled and she learned the course material

P the disciple has learned the course material if she was scribbling AND
P she was scribbling

Product rule

This turned up at the Cox theorem,
being a consequence of the *Consistency* desideratum

$$\mathsf{P}(s, l) = \mathsf{P}(l \mid s)\mathsf{P}(s)$$

P the disciple has scribbled and she learned the course material

P the disciple has learned the course material if she was scribbling AND
P she was scribbling

chain rule

Inference & conditional probability

Goal: infer the probability of learning given the disciple wrote feverishly

$$P(L \mid s, \bar{e}) = ?$$

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
…	…	…	
0	0	0	0.43

Inference & conditional probability

Goal: infer the probability of learning given the disciple wrote feverishly

$$P(L \mid s, \bar{e}) = ?$$

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
…	…	…	
0	0	0	0.43

Inference & conditional probability

Goal: infer the probability of learning given the disciple wrote feverishly

$$P(L \mid s, \bar{e}) = ?$$

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
…	…	…	
0	0	0	0.43

Inference & conditional probability

Goal: infer the probability of learning given the disciple wrote feverishly

$$P(L \mid s, \bar{e}) = ?$$

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
…	…	…	
0	0	0	0.43

$\sum P_i = 1$

Inference & conditional probability

Goal: infer the probability of learning given the disciple wrote feverishly

$$P(L \mid s, \bar{e}) = ?$$

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
…	…	…	
0	0	0	0.43

$\sum P_i = 1$

- Use the chain rule to construct the conditional probability:

$$P(s, e, l) = P(l \mid s, \bar{e})P(s, \bar{e})$$

Inference & conditional probability

Goal: infer the probability of learning given the disciple wrote feverishly

$$P(L \mid s, \bar{e}) = ?$$

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
…	…	…	
0	0	0	0.43

$\sum P_i = 1$

- Use the chain rule to construct the conditional probability:

$$P(s, e, l) = P(l \mid s, \bar{e})P(s, \bar{e})$$

$$P(l \mid s, \bar{e}) = \frac{P(s, e, l)}{P(s, \bar{e})}$$

Inference & conditional probability

Goal: infer the probability of learning given the disciple wrote feverishly

$$P(L \mid s, \bar{e}) = ?$$

scribbling	excited look	learning	P(S,E,L)
1	1	1	0.2
1	1	0	0.1
0	1	1	0.06
1	0	0	0.17
…	…	…	
0	0	0	0.43

$\sum P_i = 1$

- Use the chain rule to construct the conditional probability:

$$P(s, e, l) = P(l \mid s, \bar{e})P(s, \bar{e})$$

- Probability of A assuming B is known ('given B'):

$$P(A \mid B) = \frac{P(A, B)}{P(B)}$$

$$P(l \mid s, \bar{e}) = \frac{P(s, e, l)}{P(s, \bar{e})}$$

Example

scribbling	learning	P(S,L)
1	1	0.45
1	0	0.05
0	1	0.15
0	0	0.35

$$P(l \mid s) = \frac{\mathrm{P}(s, l)}{\mathrm{P}(s)}$$

Example

scribbling	learning	P(S,L)
1	1	0.45
1	0	0.05
0	1	0.15
0	0	0.35

+

$$P(l \mid s) = \frac{\mathrm{P}(s, l)}{\mathrm{P}(s)}$$

Example

scribbling	learning	P(S,L)
1	1	0.45
1	0	0.05
0	1	0.15
0	0	0.35

+

$$P(l \mid s) = \frac{\mathrm{P}(s, l)}{\mathrm{P}(s)}$$

Example

scribbling	learning	P(S,L)
1	1	0.45
1	0	0.05
0	1	0.15
0	0	0.35

+

$$P(l \mid s) = \frac{\mathrm{P}(s, l)}{\mathrm{P}(s)} = 0.75$$

Example

scribbling	learning	P(S,L)
1	1	0.45
1	0	0.05
0	1	0.15
0	0	0.35

+

$$P(l \mid s) = \frac{\mathrm{P}(s, l)}{\mathrm{P}(s)} = 0.75$$

By observing that the disciple was writing, the reasoning that she learned the lesson became more plausible

Example

scribbling	learning	P(S,L)
1	1	0.45
1	0	0.05
0	1	0.15
0	0	0.35

+

$$P(l \mid s) = \frac{\mathrm{P}(s, l)}{\mathrm{P}(s)} = 0.75$$

By observing that the disciple was writing, the reasoning that she learned the lesson became more plausible

$$P(\bar{s} \mid l) = ?$$

Example

scribbling	learning	P(S,L)
1	1	0.45
1	0	0.05
0	1	0.15
0	0	0.35

+

$$P(l \mid s) = \frac{\mathrm{P}(s, l)}{\mathrm{P}(s)} = 0.75$$

By observing that the disciple was writing, the reasoning that she learned the lesson became more plausible

$$P(\bar{s} \mid l) = ?$$

Knowing the probability table rich reasoning can be performed

Interim summary

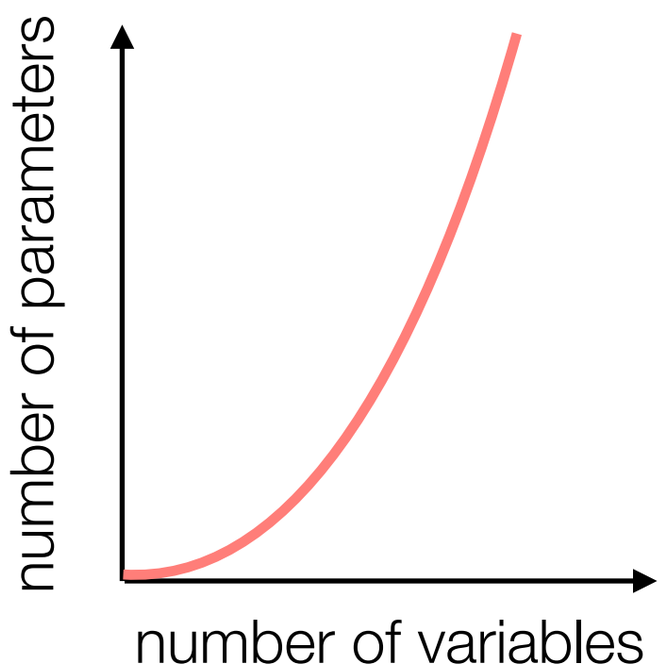
- Probabilities can be used as a basis of plausible reasoning
- Instead of the truth table, the probability table describes the system
- Rich inferences can be made
- Using the two fundamental rules of probability theory, the sum and the product rules we can express these inferences

Challenge

- Complete characterisation requires filling the probability table
- Number of lines in the probability table increases drastically with the number of variables
- How many parameters?
  (how many numbers need to be given to specify the probabilistic model?)

Challenge

- Complete characterisation requires filling the probability table
- Number of lines in the probability table increases drastically with the number of variables
- How many parameters?
  (how many numbers need to be given to specify the probabilistic model?)



$2^M - 1$

Building on knowledge about dependencies

P(S, E, R, T, F, L)

Building on knowledge about dependencies

P(S, E, R, T, F, L)

Scribbling
Excited look
Room where lecture was held
Textbook usage
Fast talking
Learning

Building on knowledge about dependencies

$$P(S, E, R, T, F, L)$$

Scribbling
Excited look
Room where lecture was held
Textbook usage
Fast talking
Learning

$$= P(L \mid S, E, R, T, F) P(S, E, R, T, F)$$

Building on knowledge about dependencies

$$P(S, E, R, T, F, L)$$

Scribbling
Excited look
Room where lecture was held
Textbook usage
Fast talking
Learning

Chain rule

$$= P(L \mid S, E, R, T, F) P(S, E, R, T, F)$$

Building on knowledge about dependencies

$$P(S, E, R, T, F, L)$$

Scribbling
Excited look
Room where lecture was held
Textbook usage
Fast talking
Learning

Chain rule

$$= P(L \mid S, E, R, T, F)P(S, E, R, T, F)$$

Textbook usage is independent of other factors

$$P(A, B) = P(A \mid B)P(B) = P(A)P(B)$$

Building on knowledge about dependencies

$P(S, E, R, T, F, L)$

Scribbling
Excited look
Room where lecture was held
Textbook usage
Fast talking
Learning

Chain rule

$$= P(L \mid S, E, R, T, F)P(S, E, R, T, F) = P(L \mid S, E, R, T, F)P(S, E, R, F)P(T)$$

Textbook usage is independent of other factors

$$P(A, B) = P(A \mid B)P(B) = P(A)P(B)$$

Building on knowledge about dependencies

$P(S, E, R, T, F, L)$

Scribbling
Excited look
Room where lecture was held
Textbook usage
Fast talking
Learning

Chain rule

$2^5 - 1 = 31$ parameters

$$= P(L \mid S, E, R, T, F)P(S, E, R, T, F) = P(L \mid S, E, R, T, F)P(S, E, R, F)P(T)$$

Textbook usage is independent of other factors

$$P(A, B) = P(A \mid B)P(B) = P(A)P(B)$$

Building on knowledge about dependencies

$P(S, E, R, T, F, L)$

Scribbling
Excited look
Room where lecture was held
Textbook usage
Fast talking
Learning

Chain rule

$2^5 - 1 = 31$ parameters

$(2^4 - 1) + (2^1 - 1) = 16$ parameters

$$= P(L \mid S, E, R, T, F)P(S, E, R, T, F) = P(L \mid S, E, R, T, F)P(S, E, R, F)P(T)$$

Textbook usage is independent of other factors

$$P(A, B) = P(A \mid B)P(B) = P(A)P(B)$$

Building on knowledge about dependencies

$P(S, E, R, T, F, L)$

Scribbling
Excited look
Room where lecture was held
Textbook usage
Fast talking
Learning

Chain rule

$2^5 - 1 = 31$ parameters

$(2^4 - 1) + (2^1 - 1) = 16$ parameters

$$= P(L \mid S, E, R, T, F)P(S, E, R, T, F) = P(L \mid S, E, R, T, F)P(S, E, R, F)P(T)$$

Textbook usage is independent of other factors

$$P(A, B) = P(A \mid B)P(B) = P(A)P(B)$$

discovering structure in the problem reduces complexity

Building on knowledge about dependencies

$P(S, E, R, T, F, L)$

Scribbling
Excited look
Room where lecture was held
Textbook usage
Fast talking
Learning

Chain rule

$2^5 - 1 = 31$ parameters

$(2^4 - 1) + (2^1 - 1) = 16$ parameters

$$= P(L \mid S, E, R, T, F)P(S, E, R, T, F) = P(L \mid S, E, R, T, F)P(S, E, R, F)P(T)$$

Textbook usage is independent of other factors

$$P(A, B) = P(A \mid B)P(B) = P(A)P(B)$$

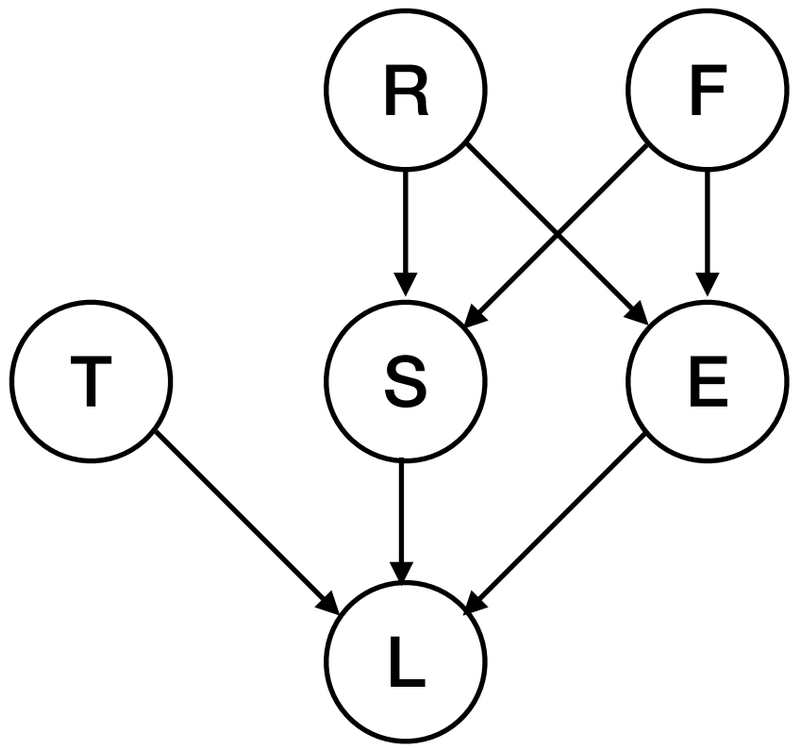
discovering structure in the problem reduces complexity

finding independencies splits up the probability table into smaller subtables

Graphical models

P(S, E, R, T, F, L)

- Stochastic variables are represented by nodes
- Dependencies (conditional probabilities) are represented by links

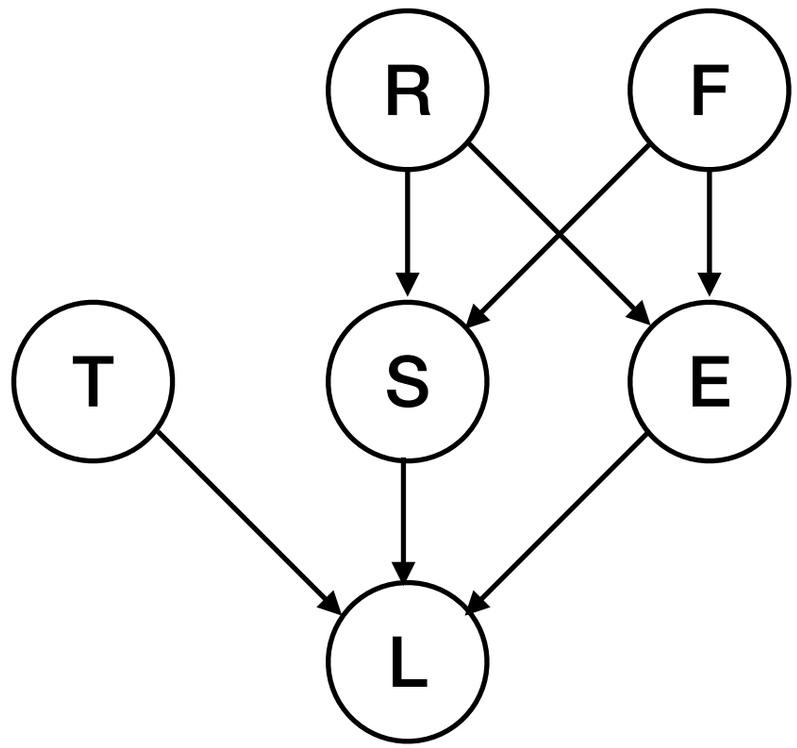


Graphical models

P(S, E, R, T, F, L)

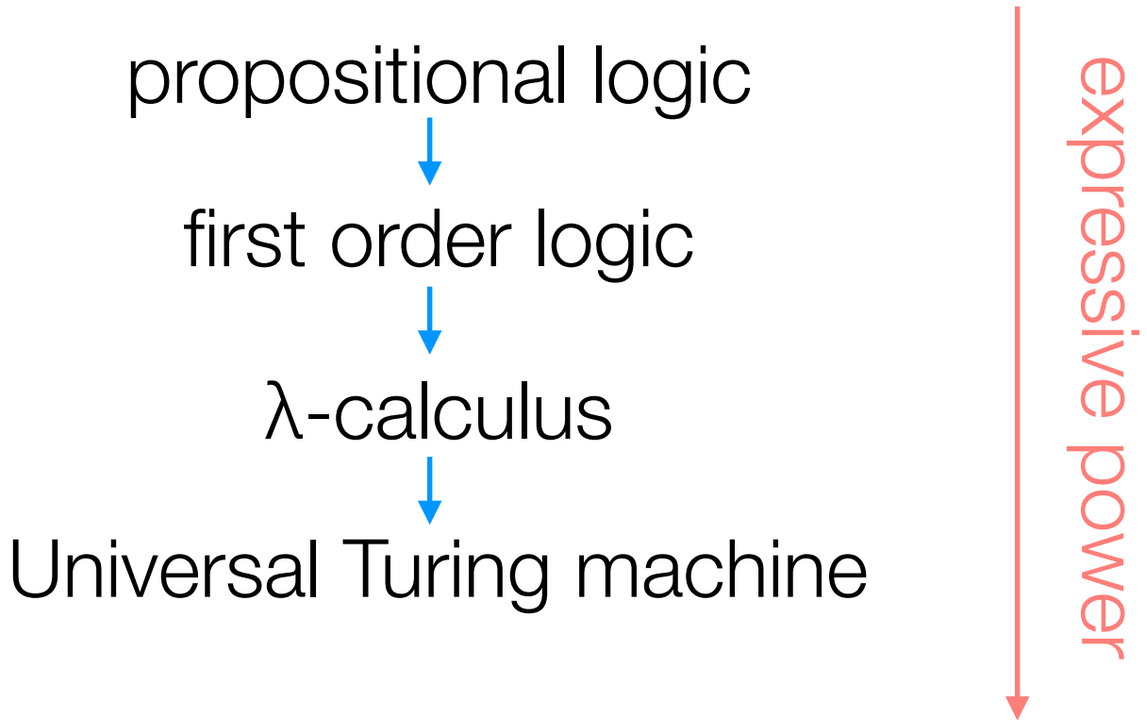
Scribbling
Excited look
Room where lecture was held
Textbook usage
Fast talking
Learning

- Stochastic variables are represented by nodes
- Dependencies (conditional probabilities) are represented by links



Graphical models vs. Probability tables

deductive reasoning



Graphical models vs. Probability tables

deductive reasoning **plausible reasoning**

propositional logic

↓

first order logic

↓

λ-calculus

↓

Universal Turing machine

expressive power

Graphical models vs. Probability tables

deductive reasoning

propositional logic

↓

first order logic

↓

λ-calculus

↓

Universal Turing machine

expressive power

plausible reasoning

probability table

Graphical models vs. Probability tables

deductive reasoning

propositional logic

↓

first order logic

↓

λ-calculus

↓

Universal Turing machine

expressive power

plausible reasoning

probability table

↓

graphical models

Graphical models vs. Probability tables

deductive reasoning

propositional logic

↓

first order logic

↓

λ-calculus

↓

Universal Turing machine

expressive power

plausible reasoning

probability table

↓

graphical models

?

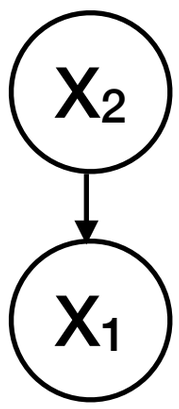
Benefits of a graphical model representation

- helps in visualising the structure of the probabilistic model
- provides insights into the properties of the model (e.g. conditional independence)
- complex calculations for inference and learning can be expressed in graphical terms

Computational insight

Independence:

$$X \perp Y \quad \Longleftrightarrow \quad P(X) = P(X \mid Y)$$
$$P(X, Y) = P(X)P(Y)$$



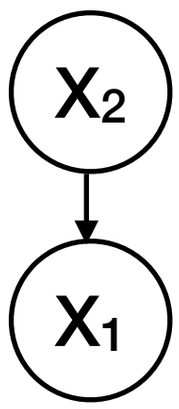
Computational insight

Independence:

$$X \perp Y \iff P(X) = P(X \mid Y)$$
$$P(X, Y) = P(X)P(Y)$$

Conditional independence:

$$X \perp Y \mid Z \iff P(X \mid Z) = P(X \mid Y, Z)$$
$$P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$$



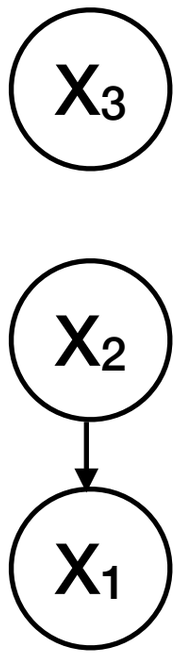
Computational insight

Independence:

$$X \perp Y \iff P(X) = P(X \mid Y)$$
$$P(X, Y) = P(X)P(Y)$$

Conditional independence:

$$X \perp Y \mid Z \iff P(X \mid Z) = P(X \mid Y, Z)$$
$$P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$$



Computational insight

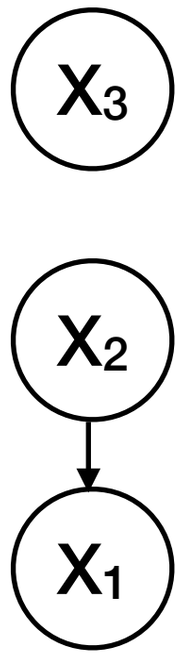
Independence:

$$X \perp Y \iff P(X) = P(X \mid Y)$$
$$P(X, Y) = P(X)P(Y)$$

Conditional independence:

$$X \perp Y \mid Z \iff P(X \mid Z) = P(X \mid Y, Z)$$
$$P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$$

Example:



Computational insight

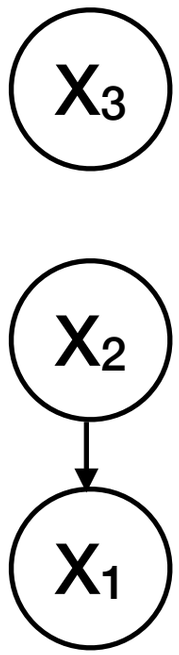
Independence:

$$X \perp Y \iff P(X) = P(X \mid Y)$$
$$P(X, Y) = P(X)P(Y)$$

Conditional independence:

$$X \perp Y \mid Z \iff P(X \mid Z) = P(X \mid Y, Z)$$
$$P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$$

Example:



$$P(X_1, X_2, X_3) = P(X_1 \mid X_2, X_3)P(X_2 \mid X_3)P(X_3)$$

Computational insight

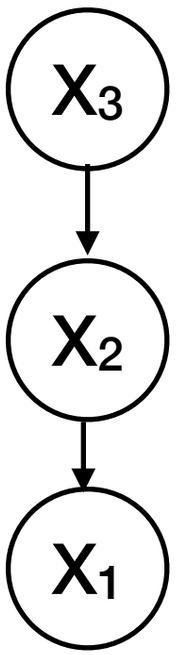
Independence:

$$X \perp Y \iff P(X) = P(X \mid Y)$$
$$P(X, Y) = P(X)P(Y)$$

Conditional independence:

$$X \perp Y \mid Z \iff P(X \mid Z) = P(X \mid Y, Z)$$
$$P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$$

Example:



$$P(X_1, X_2, X_3) = P(X_1 \mid X_2, X_3)P(X_2 \mid X_3)P(X_3)$$

Computational insight

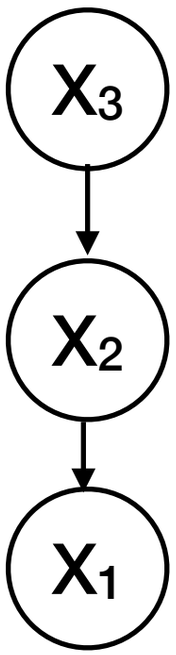
Independence:

$$X \perp Y \iff P(X) = P(X|Y)$$
$$P(X, Y) = P(X)P(Y)$$

Conditional independence:

$$X \perp Y|Z \iff P(X|Z) = P(X|Y, Z)$$
$$P(X, Y|Z) = P(X|Z)P(Y|Z)$$

Example:



$$P(X_1, X_2, X_3) = P(X_1|X_2, X_3)P(X_2|X_3)P(X_3)$$

$$X_1 \perp X_3|X_2$$

Computational insight

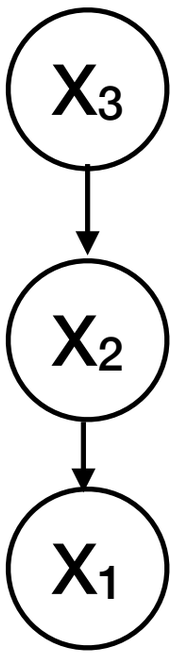
Independence:

$$X \perp Y \iff P(X) = P(X|Y)$$
$$P(X, Y) = P(X)P(Y)$$

Conditional independence:

$$X \perp Y|Z \iff P(X|Z) = P(X|Y, Z)$$
$$P(X, Y|Z) = P(X|Z)P(Y|Z)$$

Example:



$$P(X_1, X_2, X_3) = P(X_1|X_2, X_3)P(X_2|X_3)P(X_3)$$

$$X_1 \perp X_3|X_2$$

$$= P(X_1|X_2)P(X_2|X_3)P(X_3)$$

Computational insight

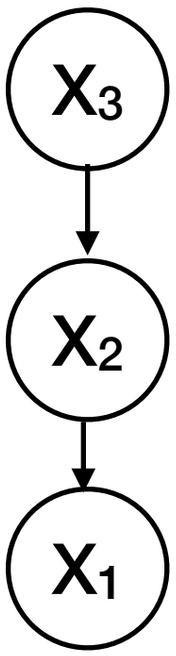
Independence:

$$X \perp Y \iff P(X) = P(X|Y)$$
$$P(X, Y) = P(X)P(Y)$$

Conditional independence:

$$X \perp Y|Z \iff P(X|Z) = P(X|Y, Z)$$
$$P(X, Y|Z) = P(X|Z)P(Y|Z)$$

Example:



$$P(X_1, X_2, X_3) = P(X_1|X_2, X_3)P(X_2|X_3)P(X_3)$$

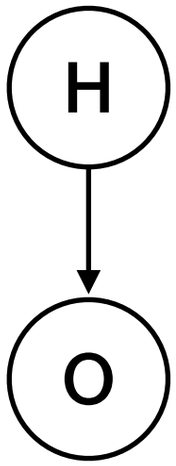
$$X_1 \perp X_3|X_2$$

$$= P(X_1|X_2)P(X_2|X_3)P(X_3)$$

in general: $$P(X_1, X_2, \ldots X_N) = \prod_i^N P(X_i|\text{Parent}(X_i))$$

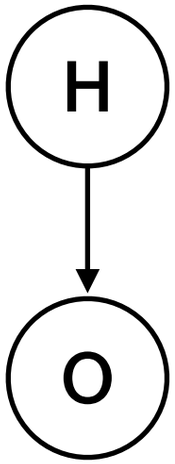
Bayes rule

Take the simplest example:



Bayes rule

Take the simplest example:

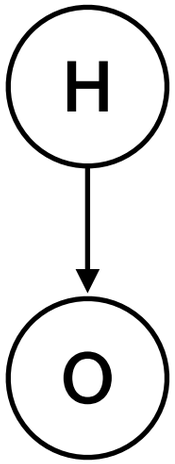


hidden variable: headache

observed variable: Aristotle's mood

Bayes rule

Take the simplest example:



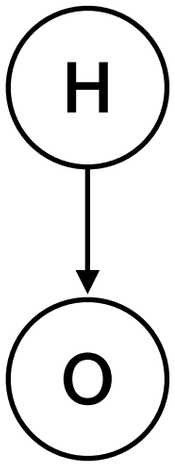
hidden variable: headache

observed variable: Aristotle's mood

- The conditional dependency describes our knowledge on how variables depend on each other
- Based on observation we might want to perform inferences (reason about latent variables)

Bayes rule

Take the simplest example:



hidden variable: headache

observed variable: Aristotle's mood

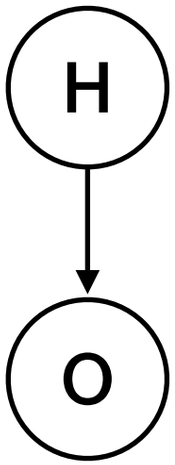
- The conditional dependency describes our knowledge on how variables depend on each other
- Based on observation we might want to perform inferences (reason about latent variables)

The full model is specified by the joint distribution:

$$P(O, H) = P(O \mid H)P(H)$$

Bayes rule

Take the simplest example:



hidden variable: headache

observed variable: Aristotle's mood

- The conditional dependency describes our knowledge on how variables depend on each other
- Based on observation we might want to perform inferences (reason about latent variables)

The full model is specified by the joint distribution:

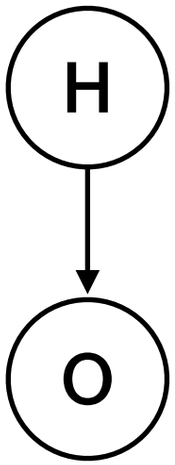
$$P(O, H) = P(O \mid H)P(H)$$

The full model is specified by the joint distribution:

$$P(O, H) = P(O \mid H)P(H)$$

Bayes rule

Take the simplest example:



hidden variable: headache

observed variable: Aristotle's mood

- The conditional dependency describes our knowledge on how variables depend on each other
- Based on observation we might want to perform inferences (reason about latent variables)

The full model is specified by the joint distribution:

$$P(O, H) = P(O \mid H)P(H)$$

however, we want to reason about the hidden state given the observation:

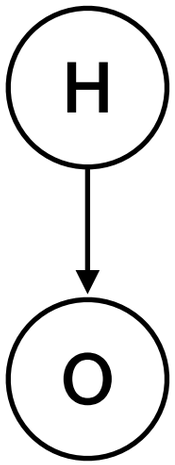
$$P(H \mid O)$$

The full model is specified by the joint distribution:

$$P(O, H) = P(O \mid H)P(H)$$

Bayes rule

Take the simplest example:



hidden variable: headache

observed variable: Aristotle's mood

- The conditional dependency describes our knowledge on how variables depend on each other
- Based on observation we might want to perform inferences (reason about latent variables)

The full model is specified by the joint distribution:

$$P(O, H) = P(O \mid H)P(H)$$

however, we want to reason about the hidden state given the observation:

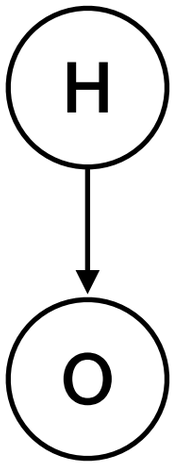
$$P(H \mid O)$$

The full model is specified by the joint distribution:

$$P(O, H) = P(O \mid H)P(H) \quad = P(H \mid O)P(O)$$

Bayes rule

Take the simplest example:



hidden variable: headache

observed variable: Aristotle's mood

- The conditional dependency describes our knowledge on how variables depend on each other
- Based on observation we might want to perform inferences (reason about latent variables)

The full model is specified by the joint distribution:

$$P(O, H) = P(O \mid H)P(H)$$

however, we want to reason about the hidden state given the observation:

$$P(H \mid O)$$

The full model is specified by the joint distribution:

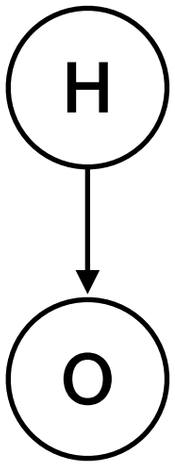
$$P(O, H) = P(O \mid H)P(H) \quad = P(H \mid O)P(O)$$

Bayes rule:

$$P(H \mid O) = \frac{P(O \mid H)P(H)}{P(O)}$$

Bayes rule

Take the simplest example:



hidden variable: headache

observed variable: Aristotle's mood

- The conditional dependency describes our knowledge on how variables depend on each other
- Based on observation we might want to perform inferences (reason about latent variables)

The full model is specified by the joint distribution:

$$P(O, H) = P(O \mid H)P(H)$$

however, we want to reason about the hidden state given the observation:

$$P(H \mid O)$$

The full model is specified by the joint distribution:

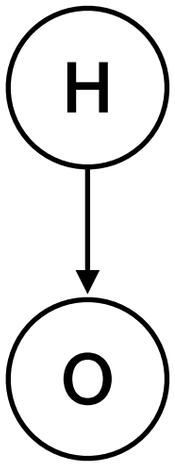
$$P(O, H) = P(O \mid H)P(H) \quad = P(H \mid O)P(O)$$

posterior probability

Bayes rule: $$P(H \mid O) = \frac{P(O \mid H)P(H)}{P(O)}$$

Bayes rule

Take the simplest example:



hidden variable: headache

observed variable: Aristotle's mood

- The conditional dependency describes our knowledge on how variables depend on each other
- Based on observation we might want to perform inferences (reason about latent variables)

The full model is specified by the joint distribution:

$$P(O, H) = P(O \mid H)P(H)$$

however, we want to reason about the hidden state given the observation:

$$P(H \mid O)$$

The full model is specified by the joint distribution:

$$P(O, H) = P(O \mid H)P$$

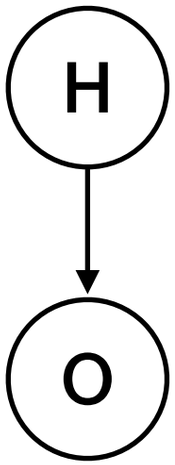
posterior probability

likelihood function

Bayes rule: $$P(H \mid O) = \frac{P(O \mid H)P(H)}{P(O)}$$

Bayes rule

Take the simplest example:



hidden variable: headache

observed variable: Aristotle's mood

- The conditional dependency describes our knowledge on how variables depend on each other
- Based on observation we might want to perform inferences (reason about latent variables)

The full model is specified by the joint distribution:

$$P(O, H) = P(O \mid H)P(H)$$

however, we want to reason about the hidden state given the observation:

$$P(H \mid O)$$

The full model is specified by the joint distribution:

$$P(O, H) = P(O \mid H)P$$

posterior probability

likelihood function

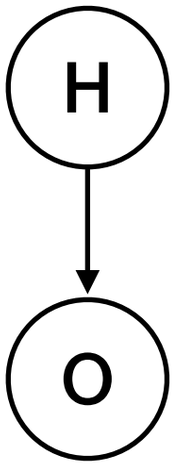
Bayes rule:

$$P(H \mid O) = \frac{P(O \mid H)P(H)}{P(O)}$$

prior distribution

Bayes rule

Take the simplest example:



hidden variable: headache

observed variable: Aristotle's mood

- The conditional dependency describes our knowledge on how variables depend on each other
- Based on observation we might want to perform inferences (reason about latent variables)

The full model is specified by the joint distribution:

$$P(O, H) = P(O \mid H)P(H)$$

however, we want to reason about the hidden state given the observation:

$$P(H \mid O)$$

The full model is specified by the joint distribution:

$$P(O, H) = P(O \mid H)P$$

posterior probability

likelihood function

Bayes rule:

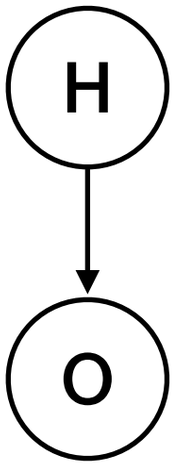
$$P(H \mid O) = \frac{P(O \mid H)P(H)}{P(O)}$$

prior distribution

marginal distribution

Bayes rule

Take the simplest example:



hidden variable: headache

observed variable: Aristotle's mood

- The conditional dependency describes our knowledge on how variables depend on each other
- Based on observation we might want to perform inferences (reason about latent variables)

The full model is specified by the joint distribution:

$$P(O, H) = P(O \mid H)P(H)$$

however, we want to reason about the hidden state given the observation:

$$P(H \mid O)$$

The full model is specified by the joint distribution:

$$P(O, H) = P(O \mid H)P$$

posterior probability **likelihood function**

Bayes rule:

$$P(H \mid O) = \frac{P(O \mid H)P(H)}{P(O)}$$

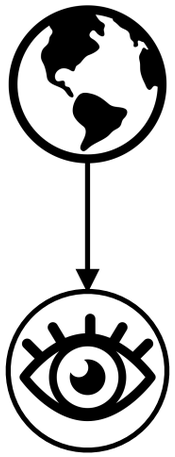
prior distribution

marginal distribution

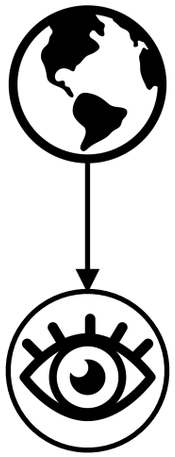
alternative phrasing:

$$P(H \mid O) \propto P(O \mid H)P(H)$$

Generative models

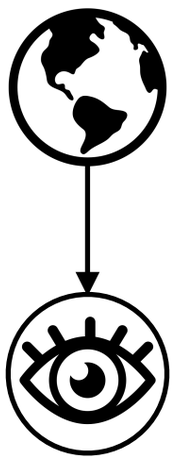


Generative models



$P(H \mid O)$

Generative models

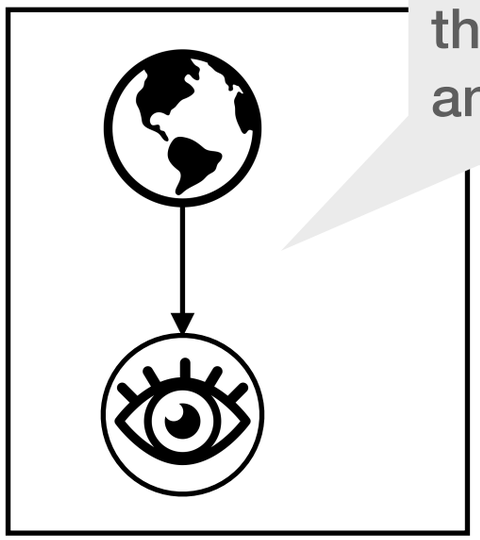


the model how the world works:
how observations are determined by
the state of the environment
and possibly noise as well

$$P(H \mid O)$$

Generative models

the model how the world works:
how observations are determined by
the state of the environment
and possibly noise as well

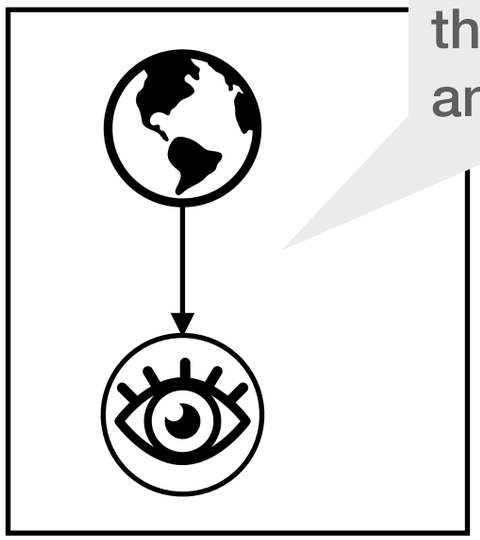


generative process

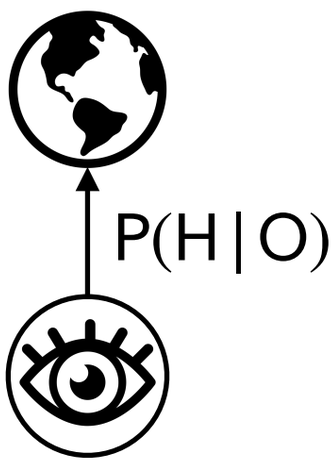
$$P(H \mid O)$$

Generative models

the model how the world works:
how observations are determined by
the state of the environment
and possibly noise as well

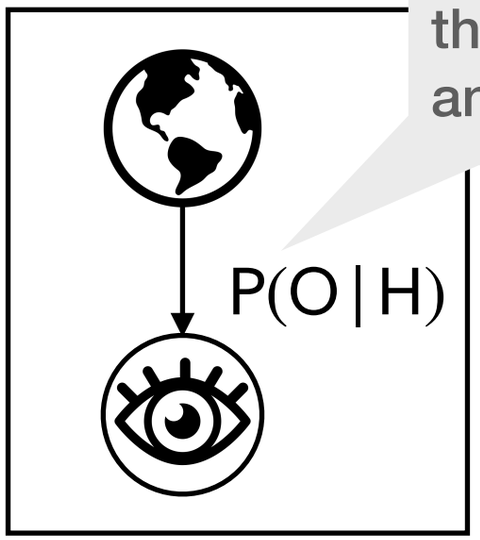


generative process

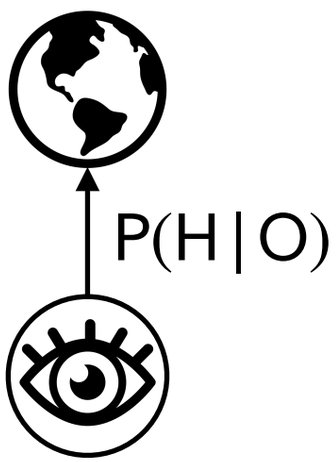


Generative models

the model how the world works:
how observations are determined by
the state of the environment
and possibly noise as well

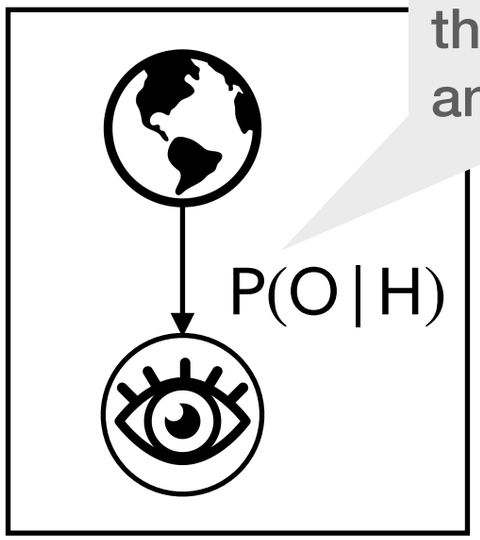


generative process

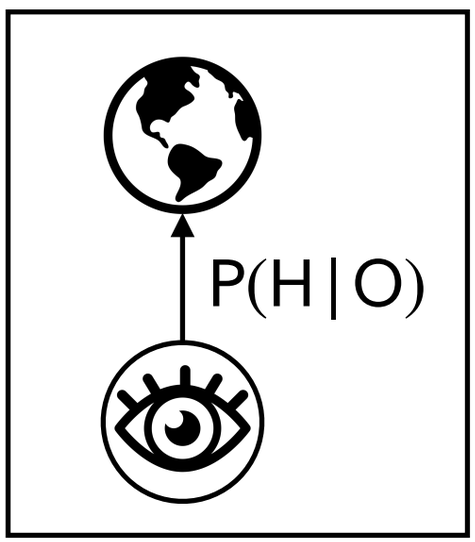


Generative models

the model how the world works:
how observations are determined by
the state of the environment
and possibly noise as well



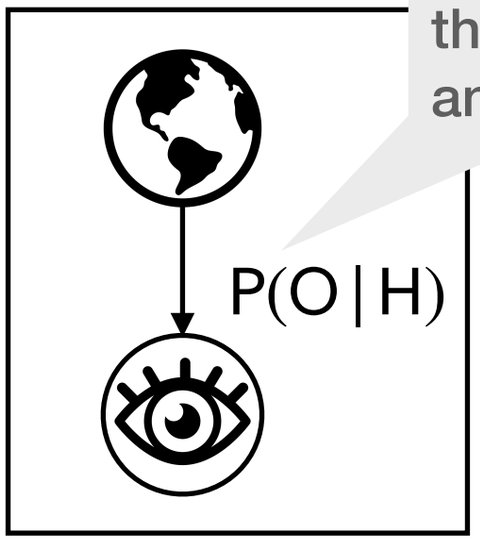
generative process



recognition model

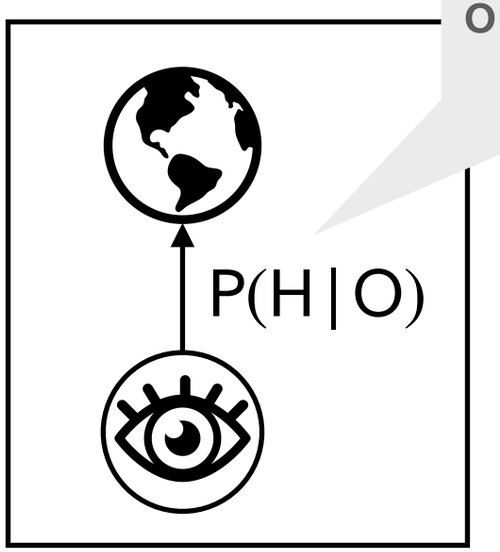
Generative models

the model how the world works:
how observations are determined by
the state of the environment
and possibly noise as well



generative process

inverting the generative model:
inferences / reasoning about
observed data



recognition model

Bayes rule

measurement and **inference**

Bayes rule

measurement and **inference**

measurement:
measuring the probability of coughing when having a flu or not

Bayes rule

measurement and **inference**

measurement:
measuring the probability of coughing when having a flu or not

$$P(\text{coughing} = 1 \,|\, \text{flu} = 1)$$

Bayes rule

measurement and **inference**

measurement:
measuring the probability of coughing when having a flu or not

$$P(\text{coughing} = 1 \,|\, \text{flu} = 1)$$

inference:
infer the probability of a hypothesis under different conditions

Bayes rule

measurement and **inference**

measurement:
measuring the probability of coughing when having a flu or not

$$P(\text{coughing} = 1 \,|\, \text{flu} = 1)$$

inference:
infer the probability of a hypothesis under different conditions

$$P(\text{flu} = 1 \,|\, \text{coughing} = 1)$$

Bayes rule

measurement and **inference**

measurement:
measuring the probability of coughing when having a flu or not

$$P(\text{coughing} = 1 \,|\, \text{flu} = 1)$$

inference:
infer the probability of a hypothesis under different conditions

$$P(\text{flu} = 1 \,|\, \text{coughing} = 1)$$

what is the connection between the two quantities?

Bayes rule

measurement and **inference**

measurement:
measuring the probability of coughing when having a flu or not

$$P(\text{coughing} = 1 \,|\, \text{flu} = 1)$$

inference:
infer the probability of a hypothesis under different conditions

$$P(\text{flu} = 1 \,|\, \text{coughing} = 1)$$

what is the connection between the two quantities?

remember: **multiplication rule:** $P(x, y) = P(x \,|\, y)P(y)$

Bayes rule

measurement and **inference**

measurement:
measuring the probability of coughing when having a flu or not

$$P(\text{coughing} = 1 \,|\, \text{flu} = 1)$$

inference:
infer the probability of a hypothesis under different conditions

$$P(\text{flu} = 1 \,|\, \text{coughing} = 1)$$

what is the connection between the two quantities?

remember: **multiplication rule:** $P(x, y) = P(x \,|\, y)P(y)$

or equivalently:

$$P(x, y) = P(y \,|\, x)P(x)$$

Bayes rule

measurement and **inference**

measurement:
measuring the probability of coughing when having a flu or not

$$P(\text{coughing} = 1 \,|\, \text{flu} = 1)$$

inference:
infer the probability of a hypothesis under different conditions

$$P(\text{flu} = 1 \,|\, \text{coughing} = 1)$$

what is the connection between the two quantities?

remember: **multiplication rule:** $P(x, y) = P(x \,|\, y)P(y)$

or equivalently:

$$P(x, y) = P(y \,|\, x)P(x)$$

$$P(\text{flu} = 1 \,|\, \text{coughing} = 1) = \frac{P(\text{coughing} = 1 \,|\, \text{flu} = 1)P(\text{flu} = 1)}{P(\text{coughing} = 1)}$$

Bayes rule

measurement and **inference**

measurement:
measuring the probability of coughing when having a flu or not

$$P(\text{coughing} = 1 \,|\, \text{flu} = 1)$$

inference:
infer the probability of a hypothesis under different conditions

$$P(\text{flu} = 1 \,|\, \text{coughing} = 1)$$

what is the connection between the two quantities?

remember: **multiplication rule:** $P(x, y) = P(x \,|\, y)P(y)$

or equivalently:

$$P(x, y) = P(y \,|\, x)P(x)$$

$$P(\text{flu} = 1 \,|\, \text{coughing} = 1) = \frac{P(\text{coughing} = 1 \,|\, \text{flu} = 1)P(\text{flu} = 1)}{P(\text{coughing} = 1)}$$

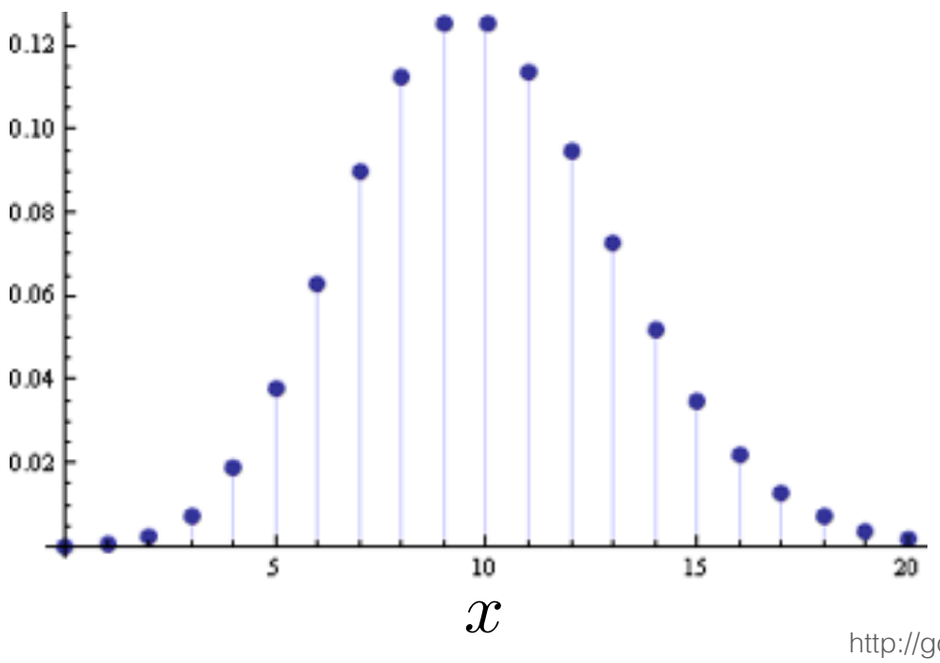
Bayes rule: ‘inverts’ a probabilistic relationship

Probability mass function / probability density

Probability mass function / probability density

$P(x)$ summarises all the possible values of a variable in a single function

$P(x)$

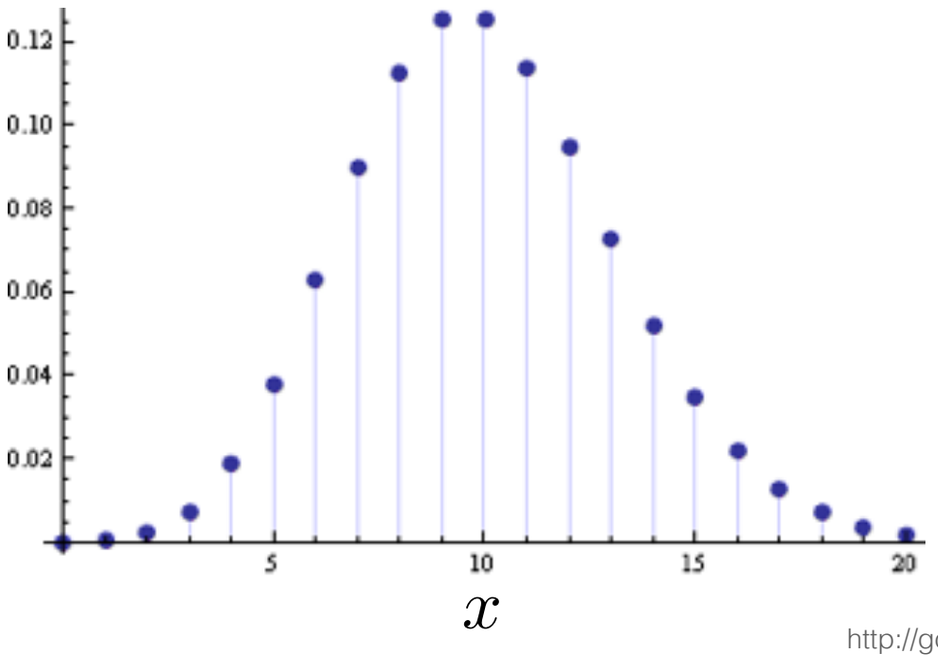


Probability mass function / probability density

$P(x)$ summarises all the possible values of a variable in a single function

- discrete variables
  - value of the function is listed at all possible values of *x*
  - the function is the probability mass function

$P(x)$

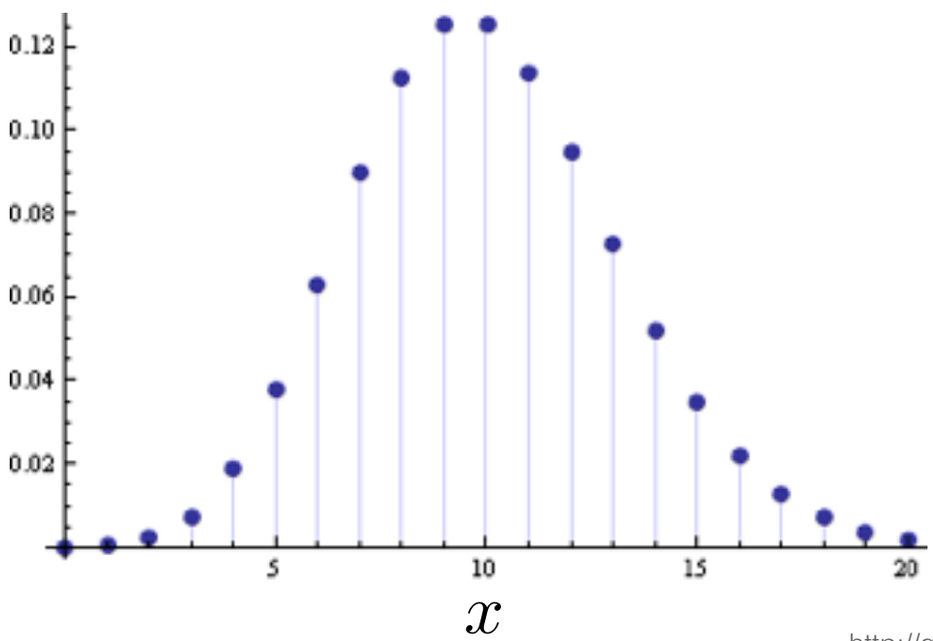


Probability mass function / probability density

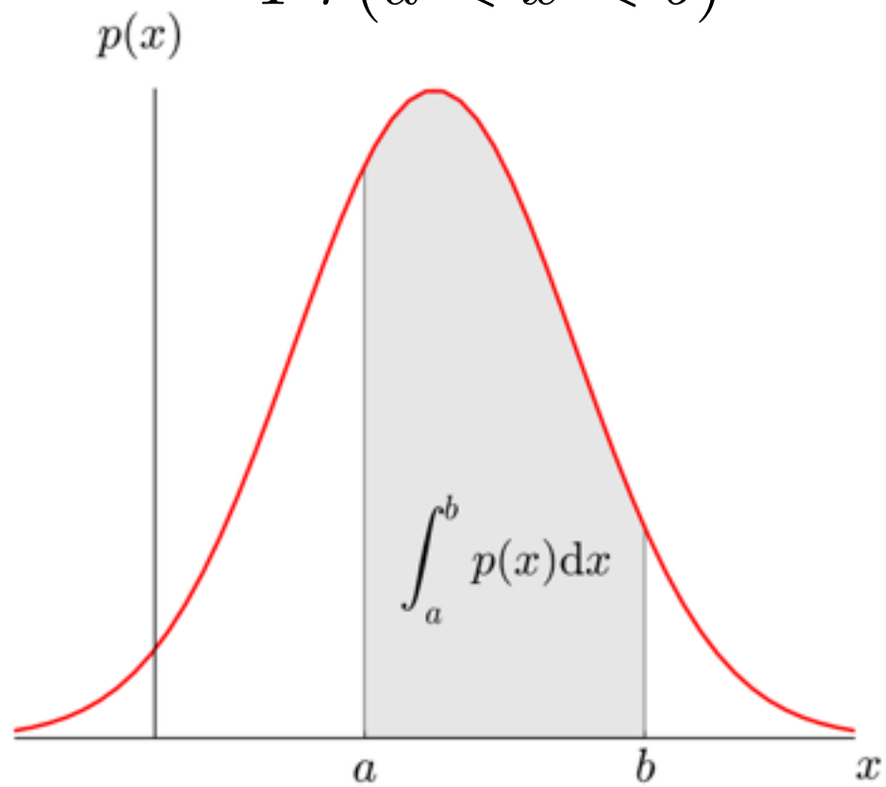
$P(x)$ summarises all the possible values of a variable in a single function

- discrete variables
  - value of the function is listed at all possible values of *x*
  - the function is the probability mass function
- continuous variables
  - the probability of any particular value is equal to 0
  - a finite interval on the variable can have a probability different from 0
  - the function is a probability density function

$P(x)$



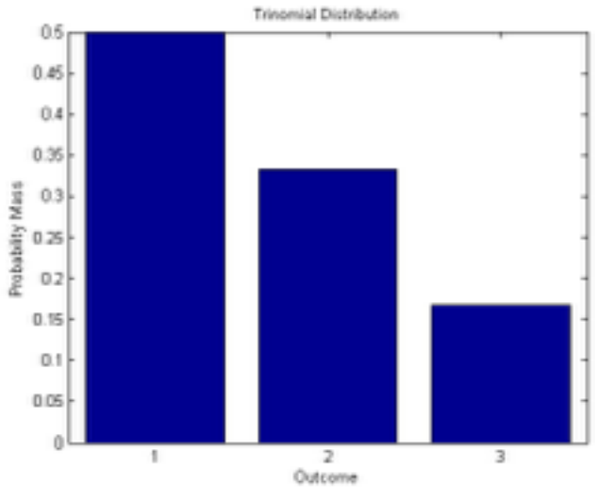
$$Pr(a < x < b) =$$



Examples of probability distributions

- Discrete valued
  - Bernoulli — coin toss

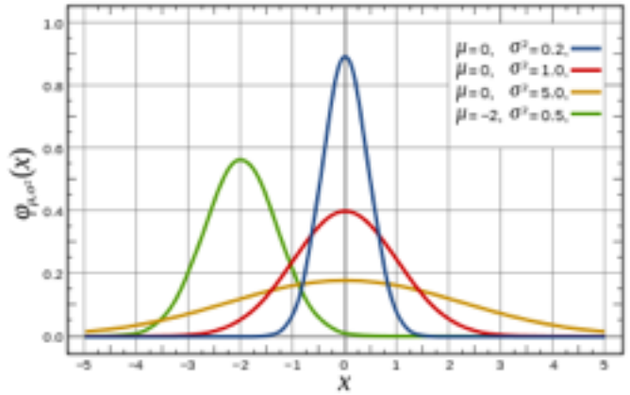
$$Ber\left(x;p\right) = p^x \cdot \left(1-p\right)^{(1-x)}$$



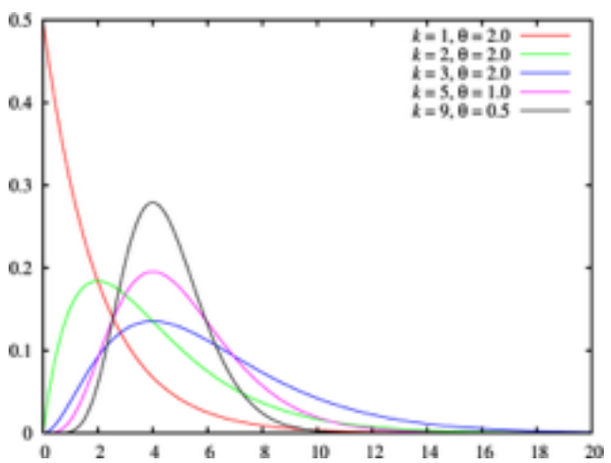
  - multinomial — dice throwing

- Continuous valued
  - uniform
  - Gaussian (normal)

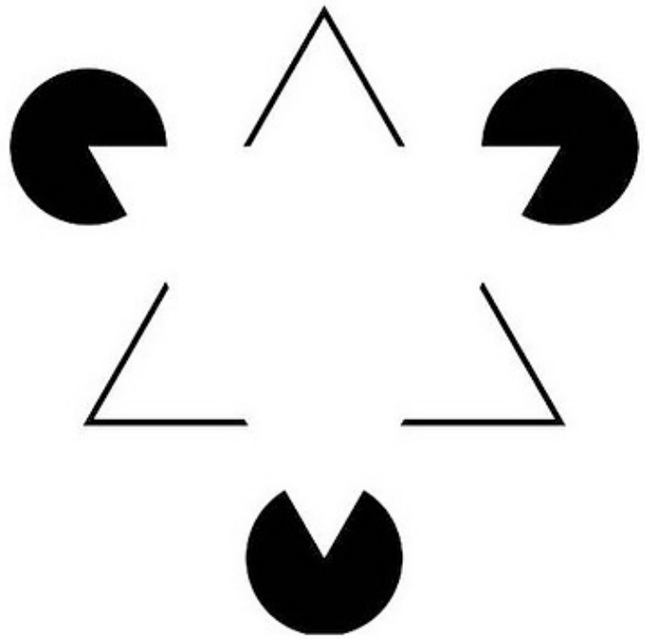
$$Normal(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



  - Gamma



Expectations and uncertainty



Reason #1

limited information

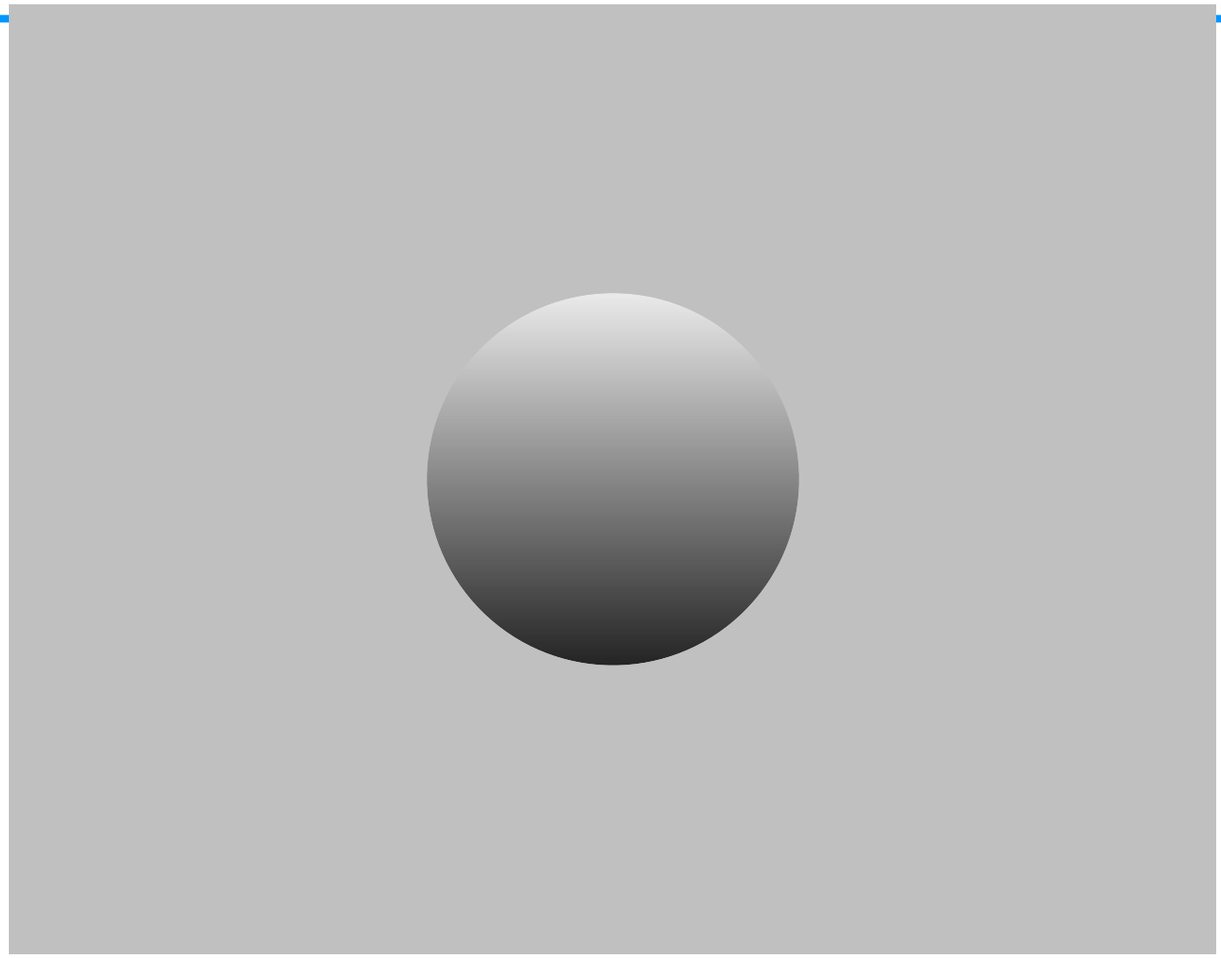
⇓

uncertainty in inferences

Formally:

- multiple hypotheses
- alternative hypotheses characterised by different probabilities

Bayesian inference



Bayesian inference

Where is the sun?

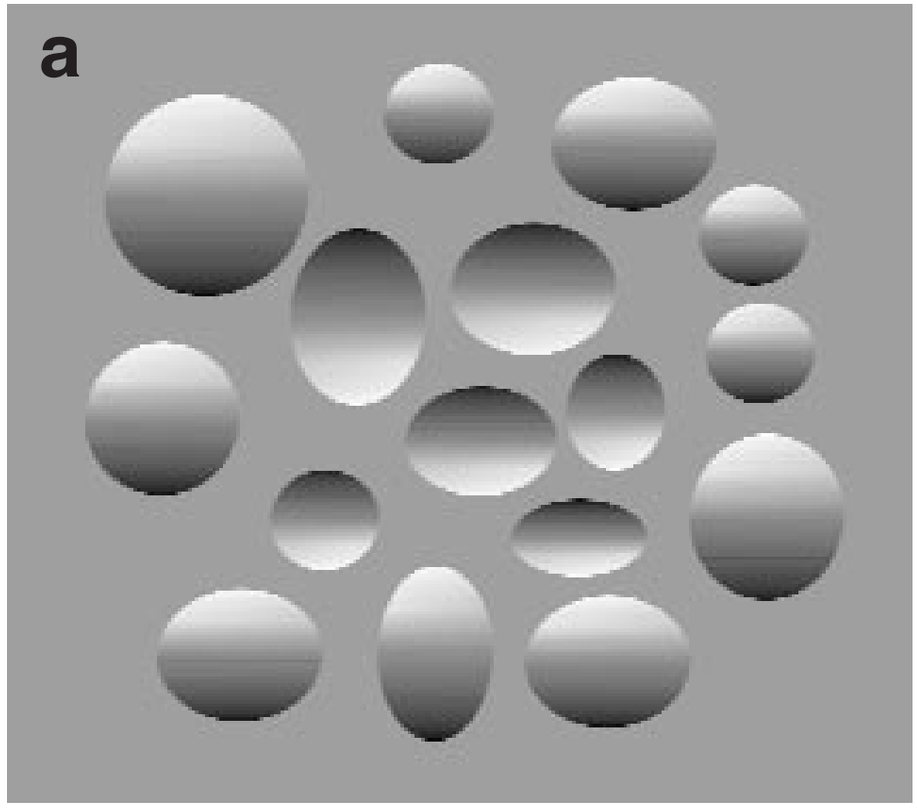
Jennifer Sun[1] and Pietro Perona[1,2]

[1] *California Institute of Technology 136-93, Pasadena, California 91125, USA*

[2] *Universita di Padova, Via Ognissanti 72, 35131 Padova, Italy*

Correspondence should be addressed to P.P. (perona@vision.caltech.edu)

nature *neuroscience* • volume 1 no 3 • july 1998



Bayesian inference

Where is the sun?

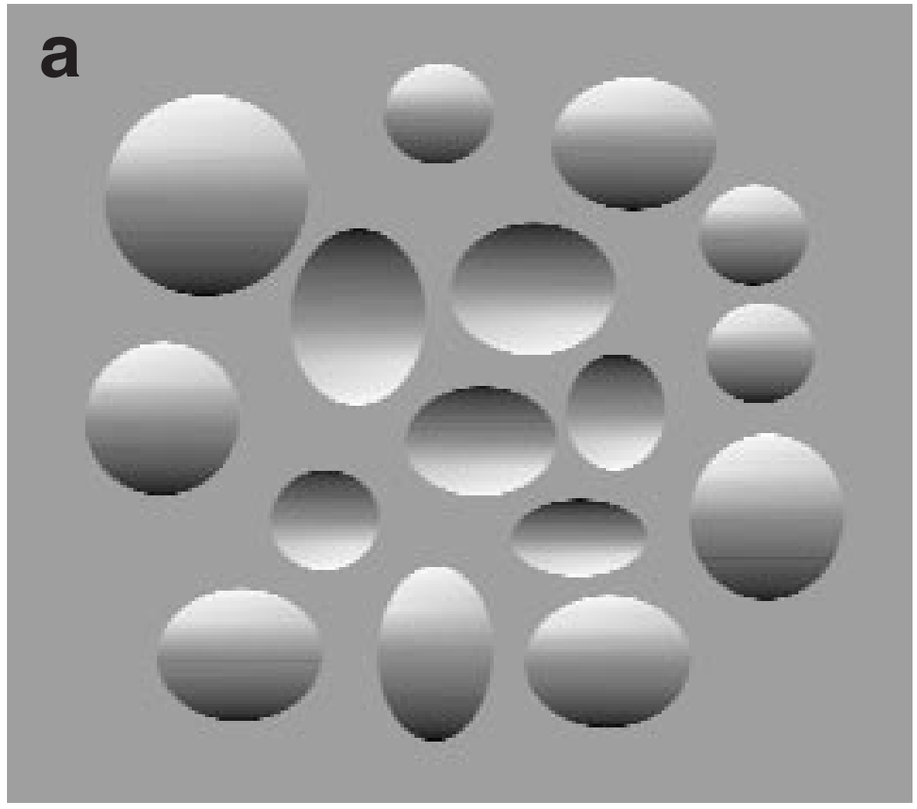
Jennifer Sun[1] and Pietro Perona[1,2]

[1] *California Institute of Technology 136-93, Pasadena, California 91125, USA*

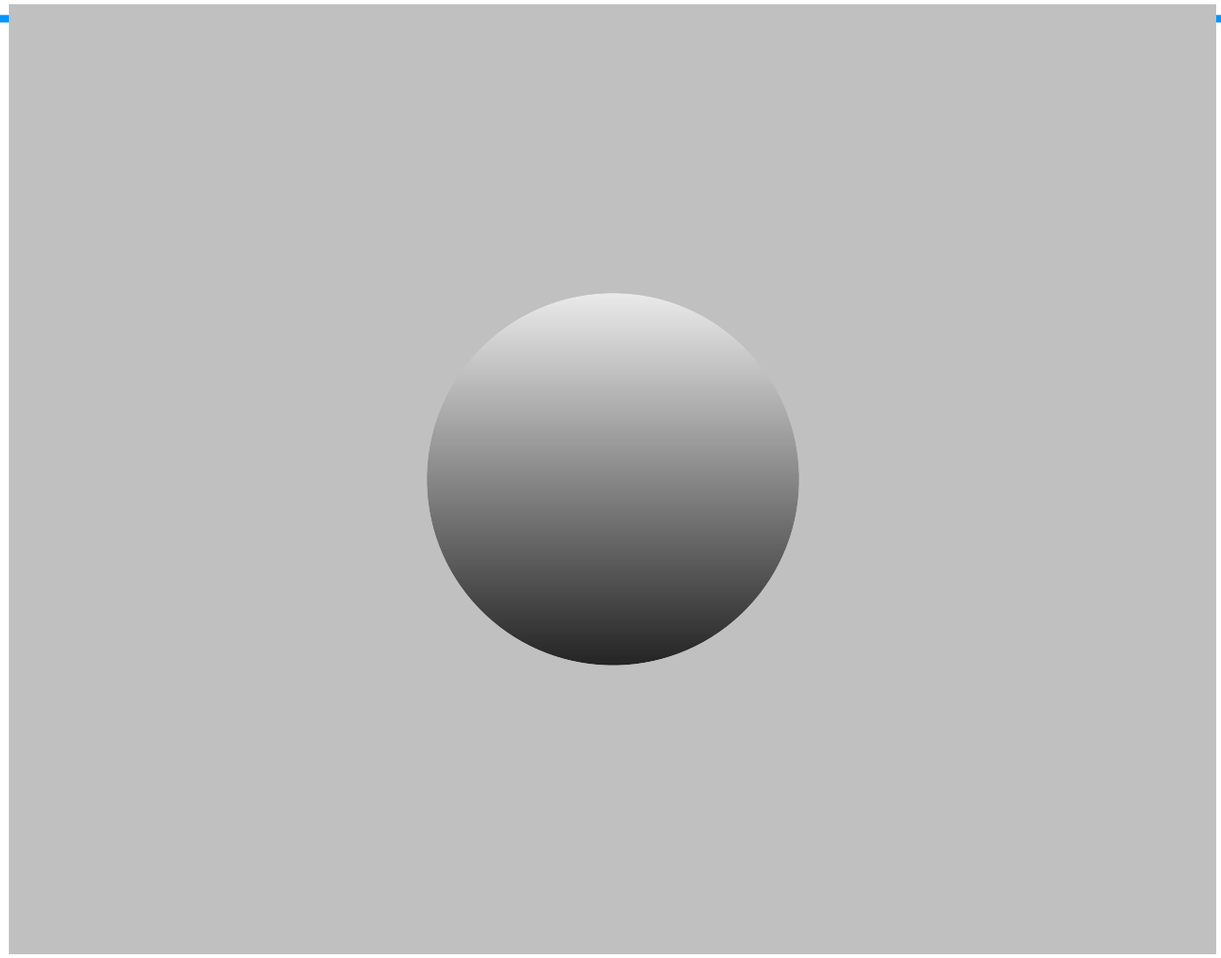
[2] *Universita di Padova, Via Ognissanti 72, 35131 Padova, Italy*

Correspondence should be addressed to P.P. (perona@vision.caltech.edu)

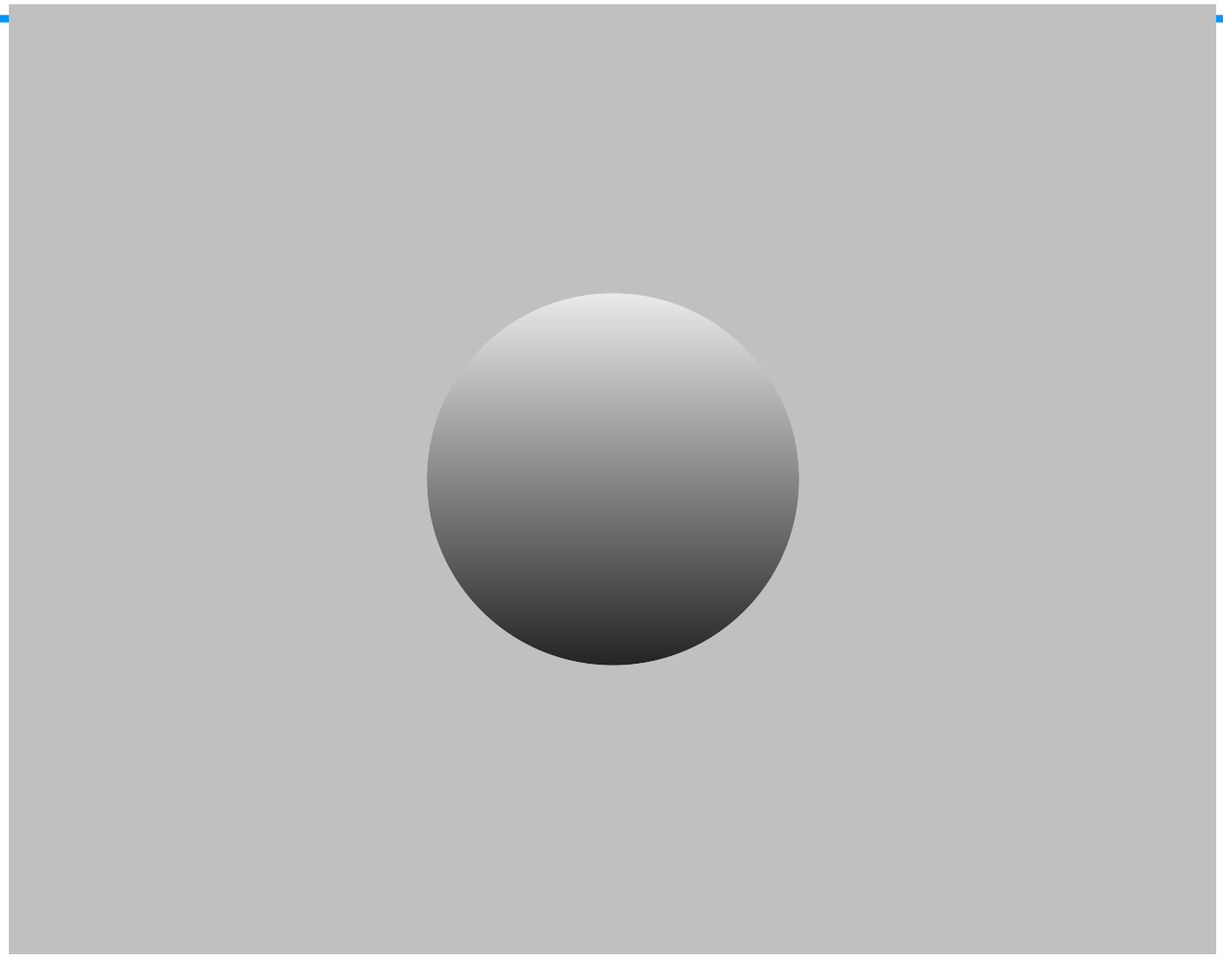
nature *neuroscience* • volume 1 no 3 • july 1998

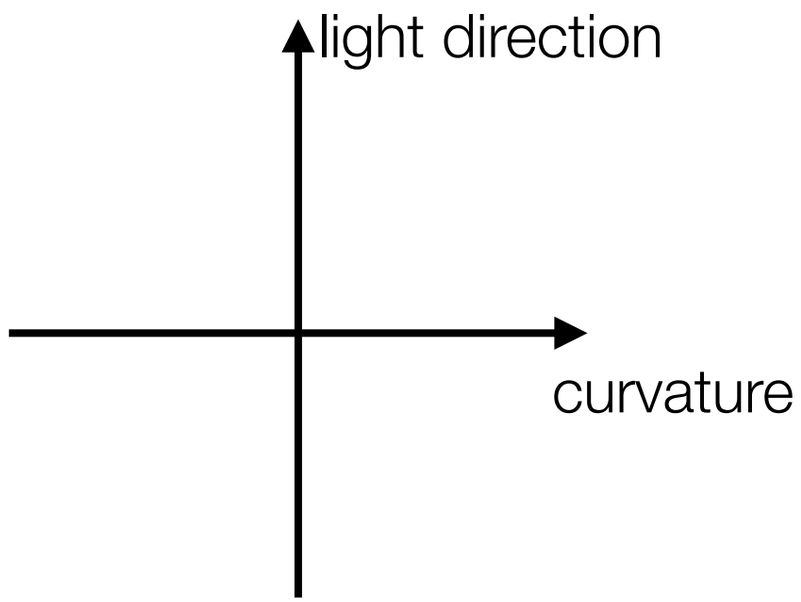


Bayesian inference

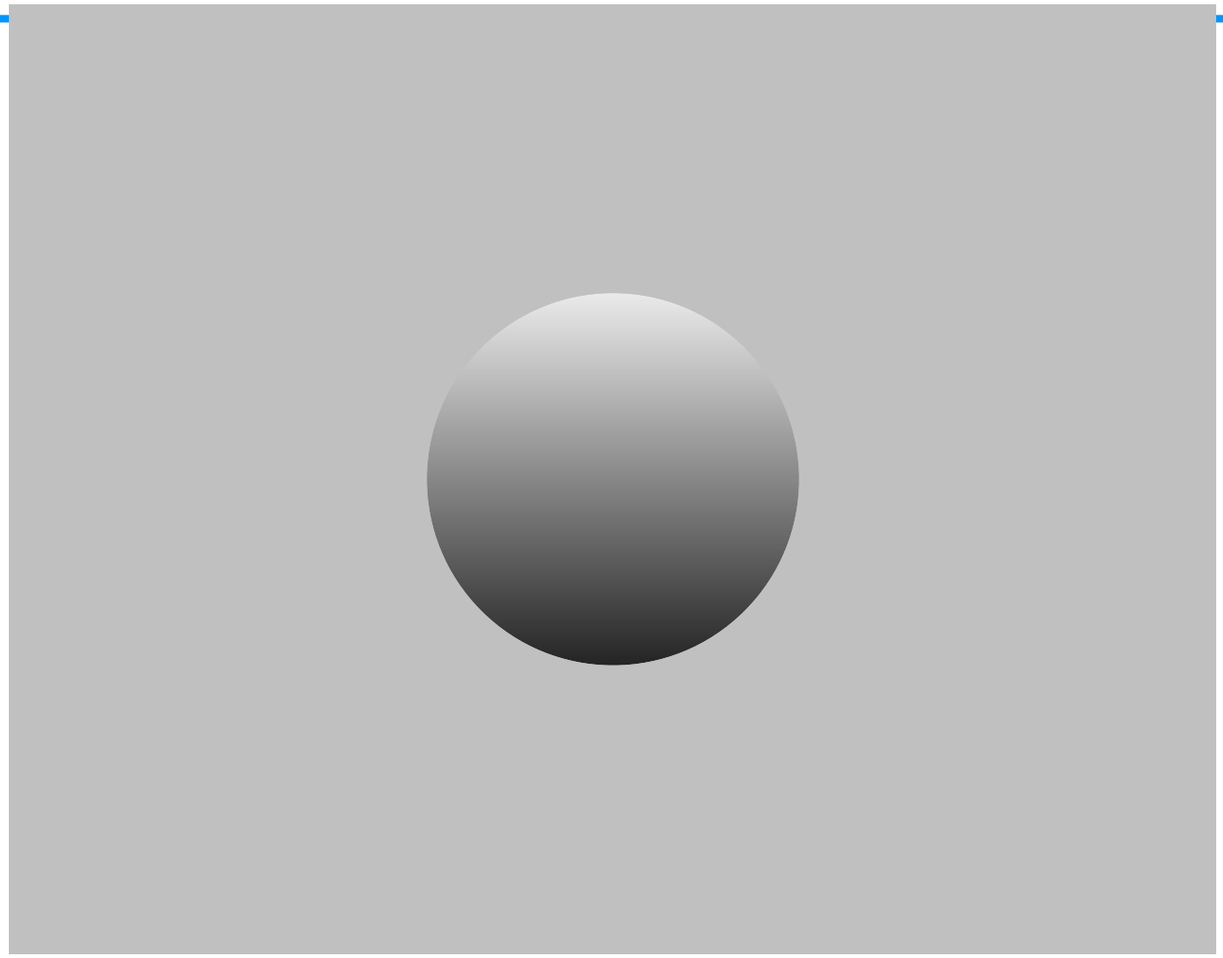


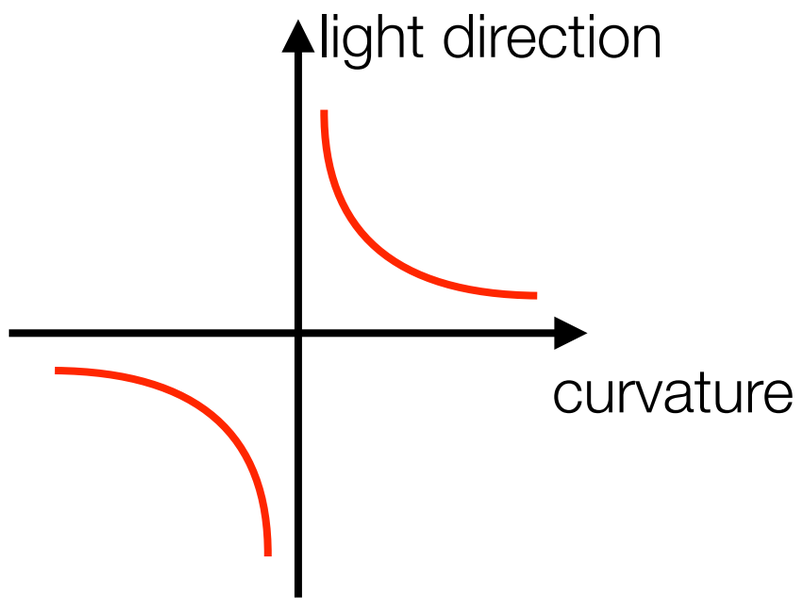
Bayesian inference



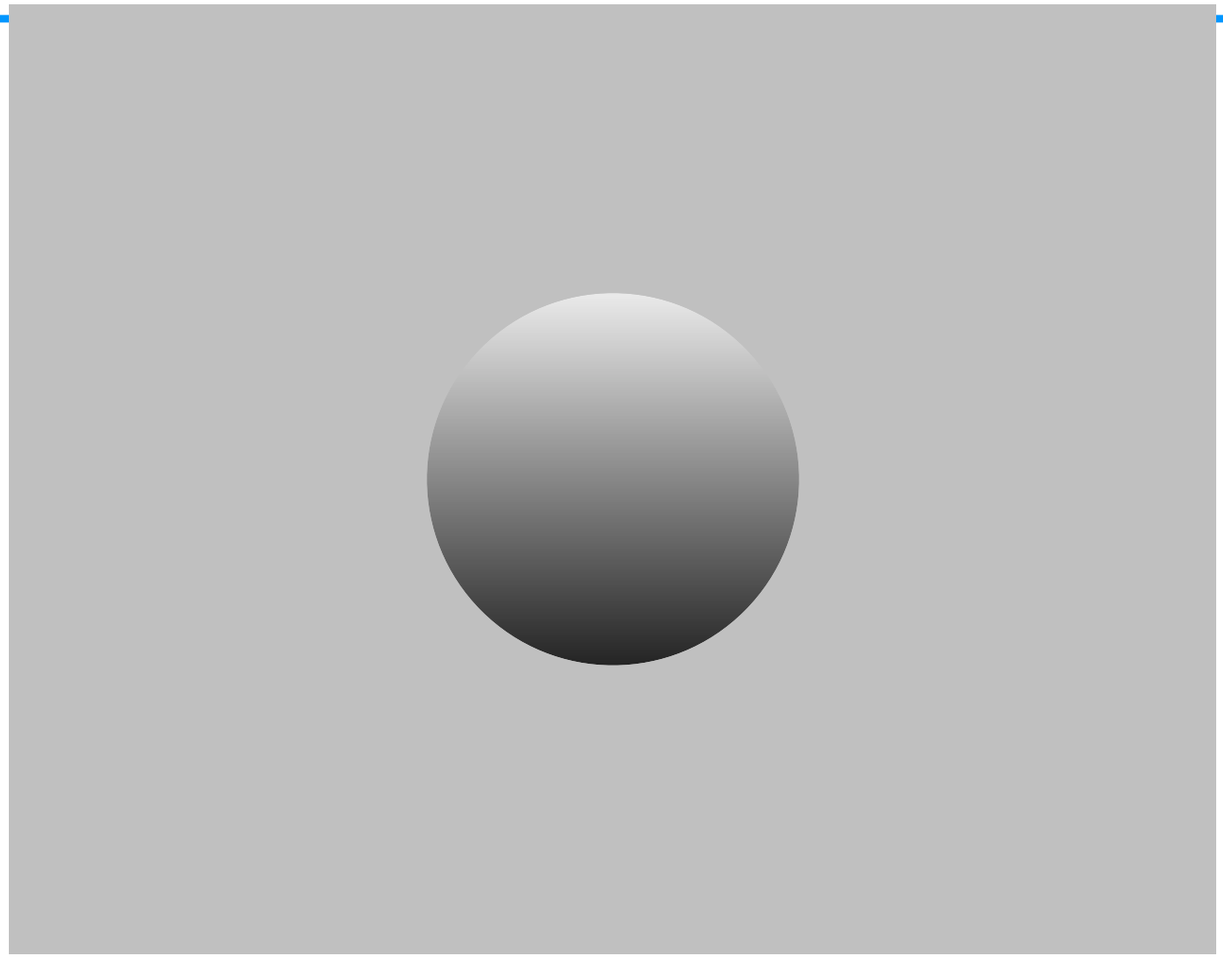


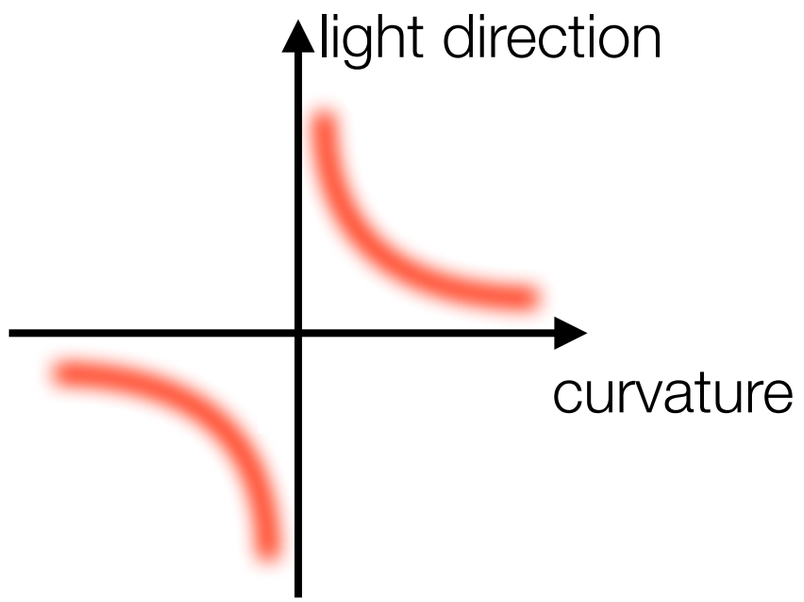
Bayesian inference





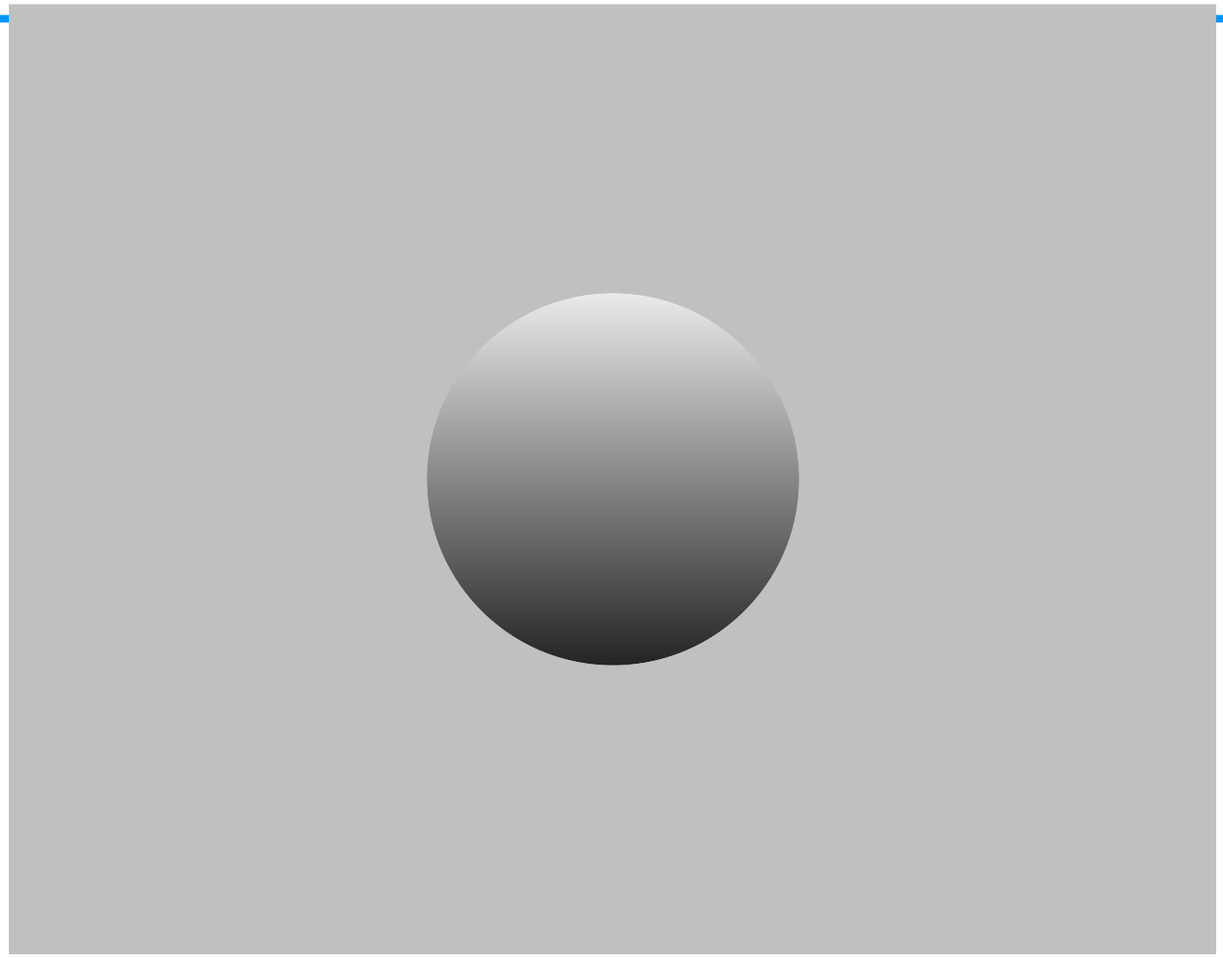
Bayesian inference

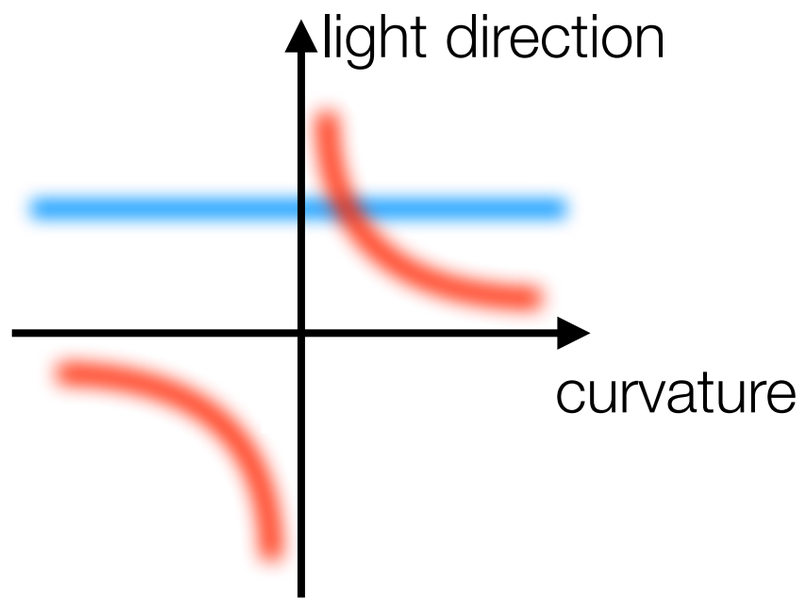




evidence

Bayesian inference

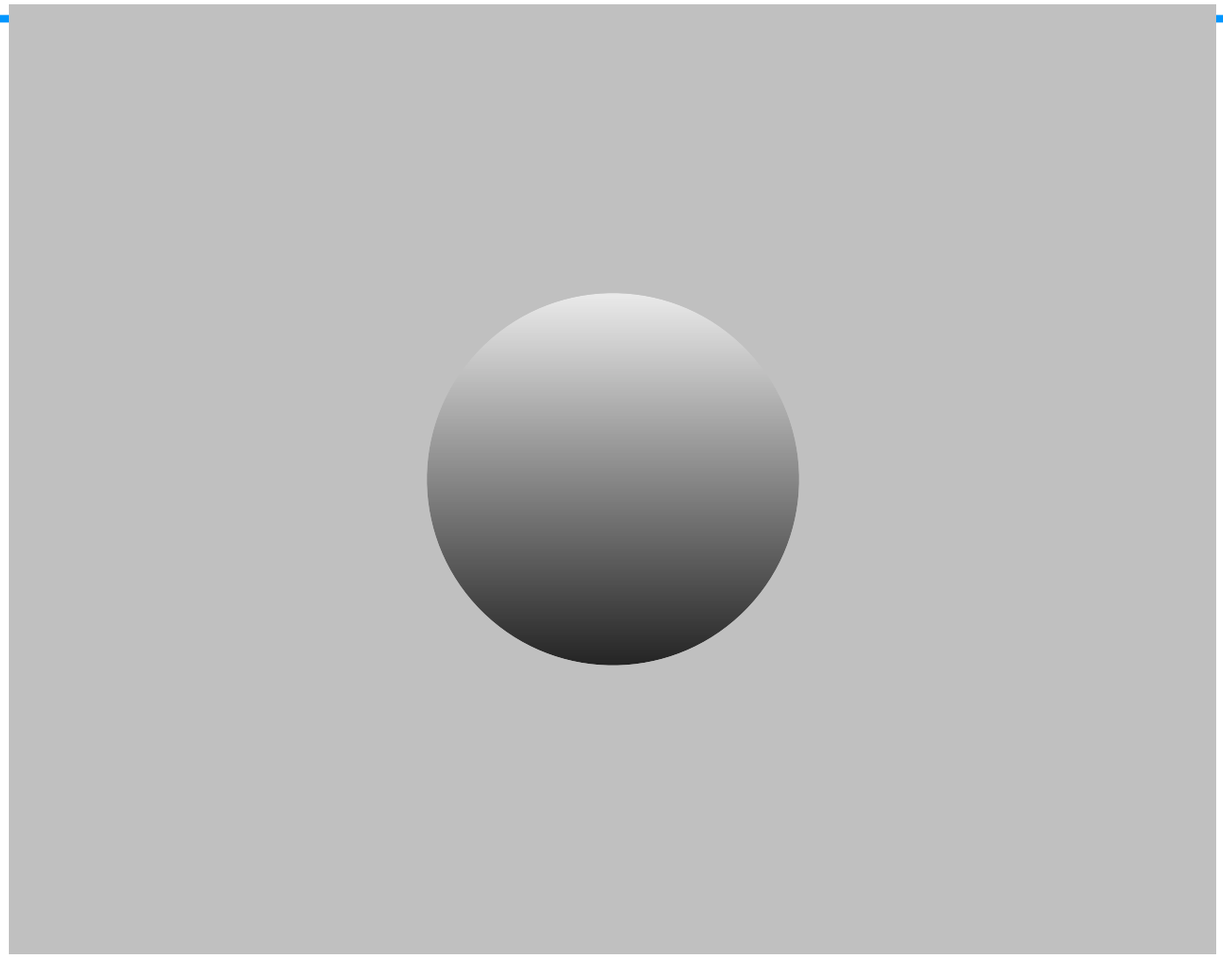


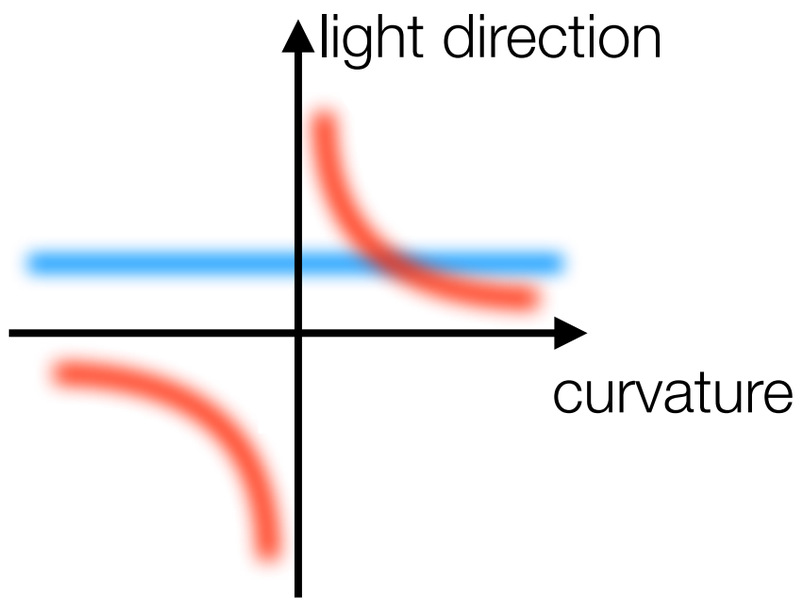


evidence

expectation

Bayesian inference

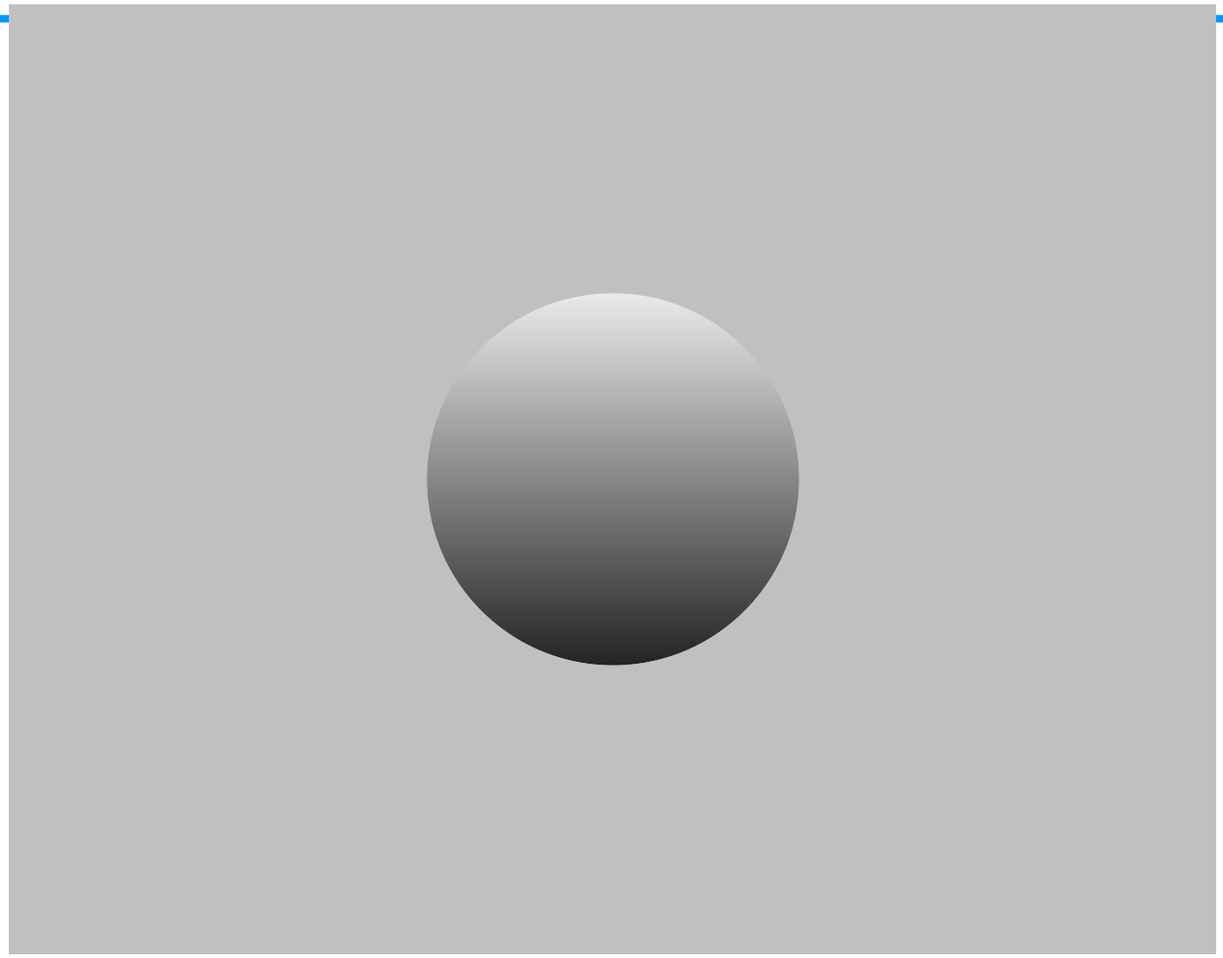


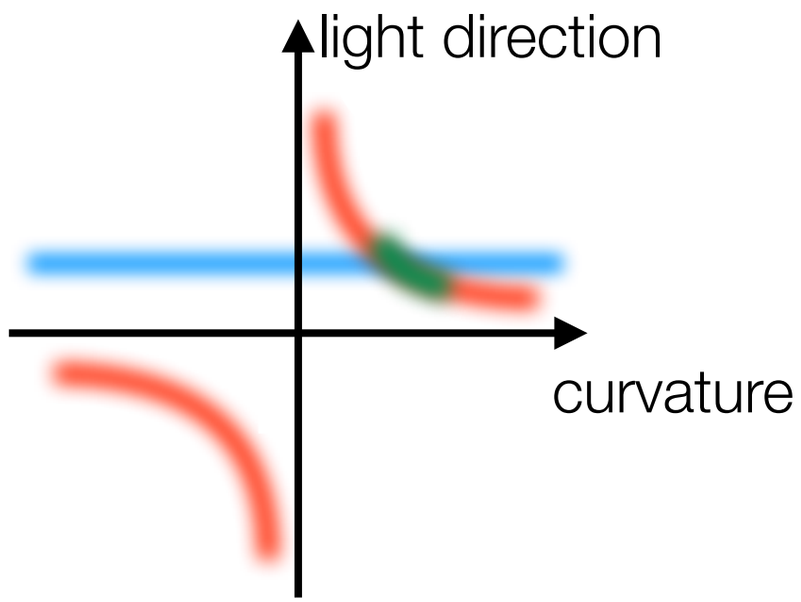


evidence

expectation

Bayesian inference



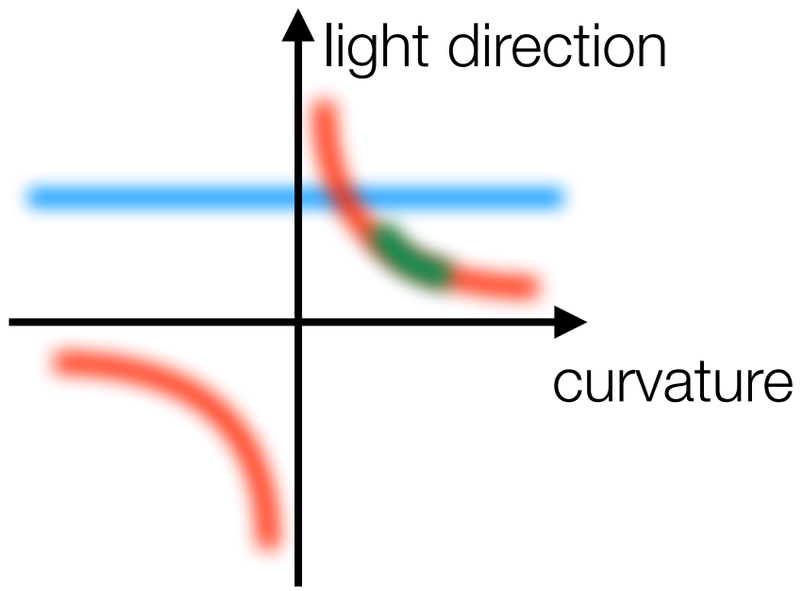


evidence

expectation

inference

Bayesian inference

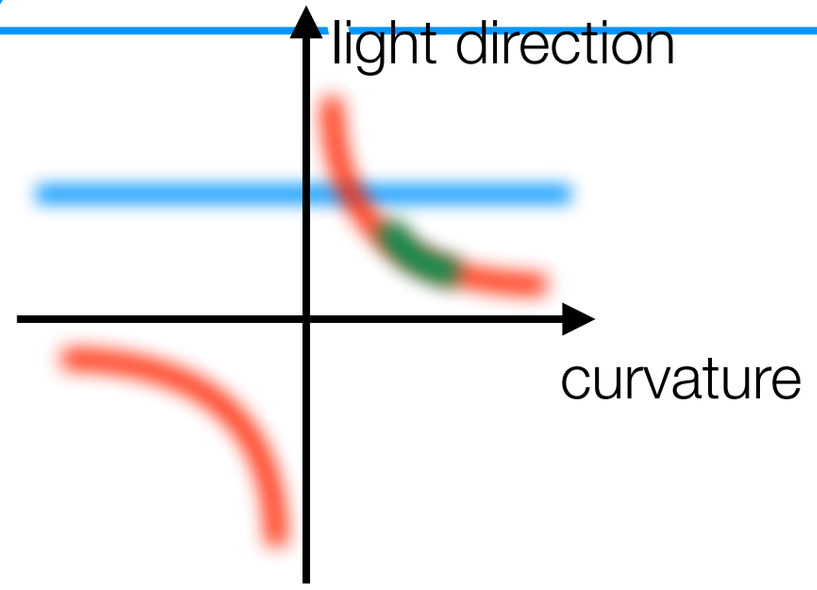


evidence

expectation

inference

Bayesian inference

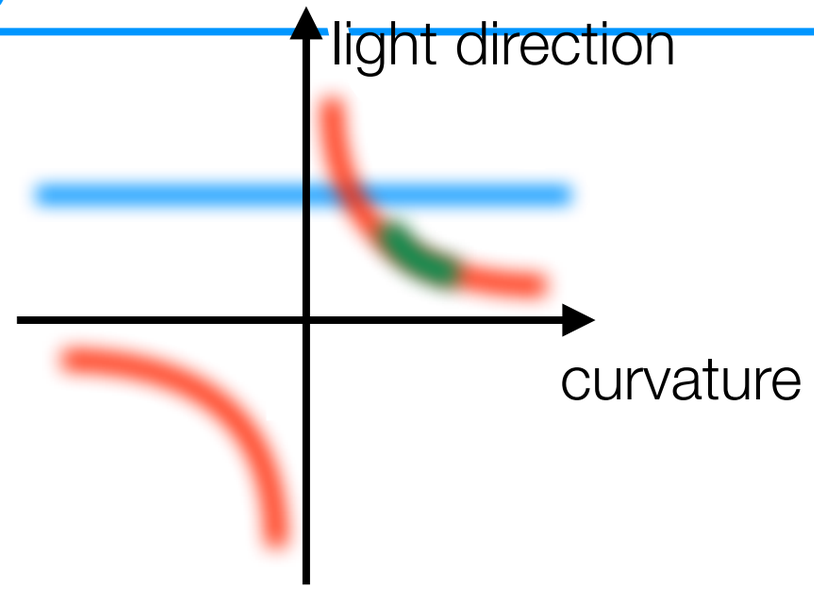


evidence

expectation

inference

Bayesian inference



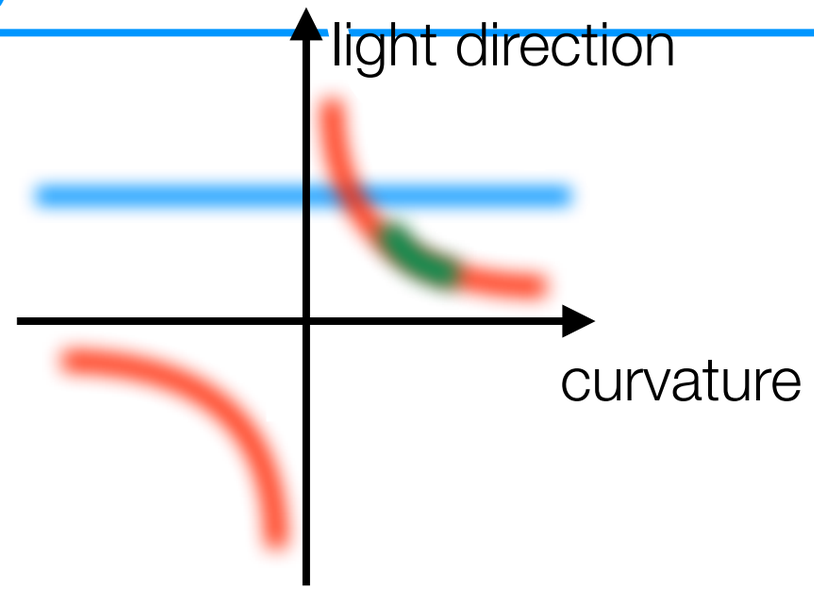
evidence

expectation

inference

$$P(\text{feature}\,|\,\text{stimulus}) \propto P(\text{stimulus}\,|\,\text{feature}) \times P(\text{feature})$$

Bayesian inference



evidence

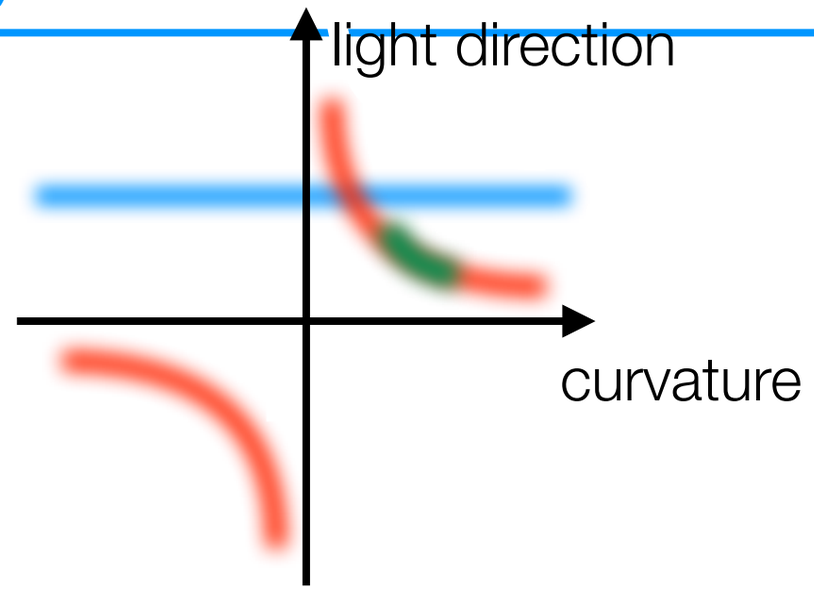
expectation

inference

$$P(\text{feature}\,|\,\text{stimulus}) \propto P(\text{stimulus}\,|\,\text{feature}) \times P(\text{feature})$$

posterior: inference

Bayesian inference



evidence

expectation

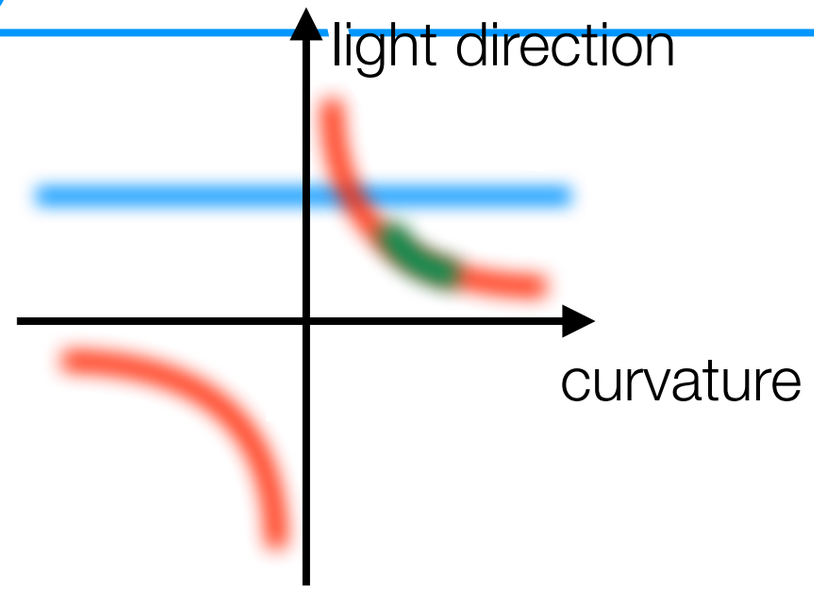
inference

$$P(\text{feature}\,|\,\text{stimulus}) \propto P(\text{stimulus}\,|\,\text{feature}) \times P(\text{feature})$$

posterior: inference

likelihood: evidence

Bayesian inference



evidence

expectation

inference

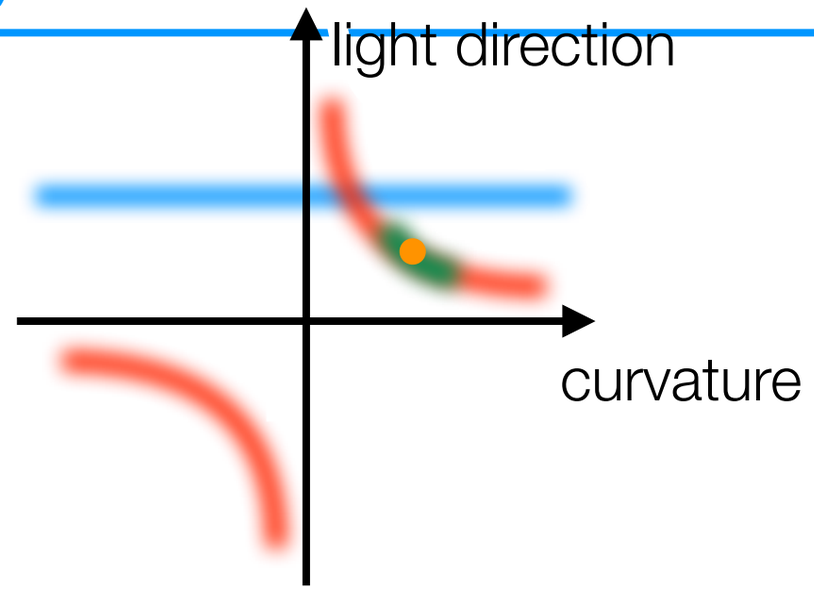
$$P(\text{feature}\,|\,\text{stimulus}) \propto P(\text{stimulus}\,|\,\text{feature}) \times P(\text{feature})$$

posterior: inference

likelihood: evidence

prior : expectations

Bayesian inference



evidence

expectation

inference

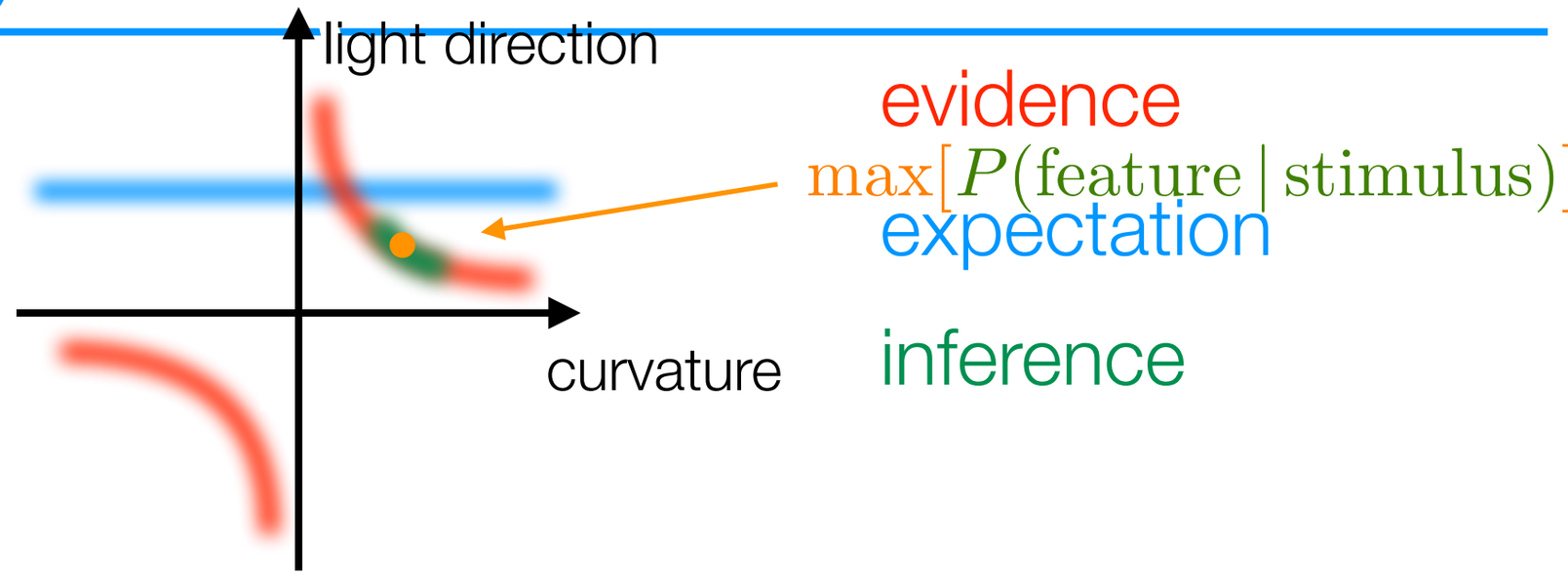
$$P(\text{feature}\,|\,\text{stimulus}) \propto P(\text{stimulus}\,|\,\text{feature}) \times P(\text{feature})$$

posterior: inference

likelihood: evidence

prior : expectations

Bayesian inference

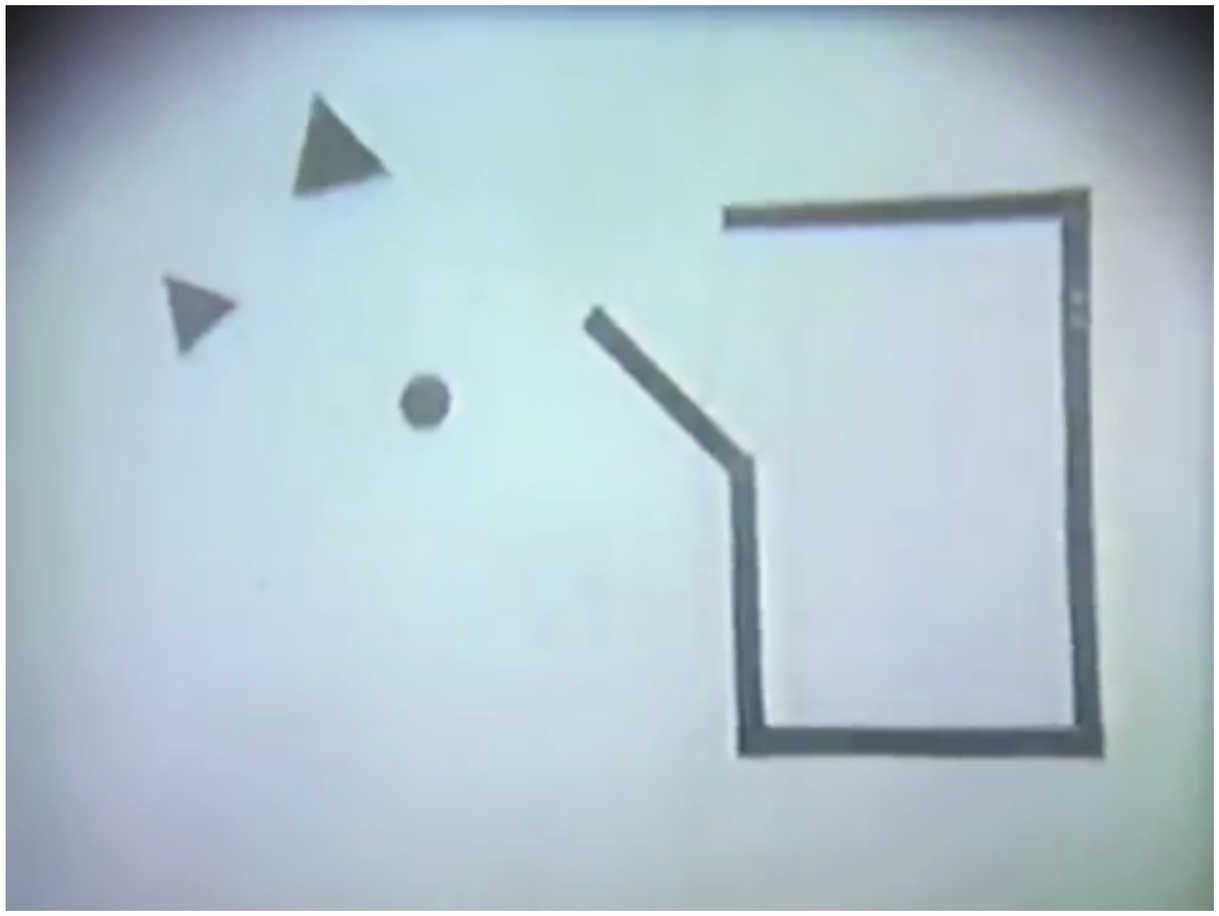


$$P(\text{feature} \mid \text{stimulus}) \propto P(\text{stimulus} \mid \text{feature}) \times P(\text{feature})$$

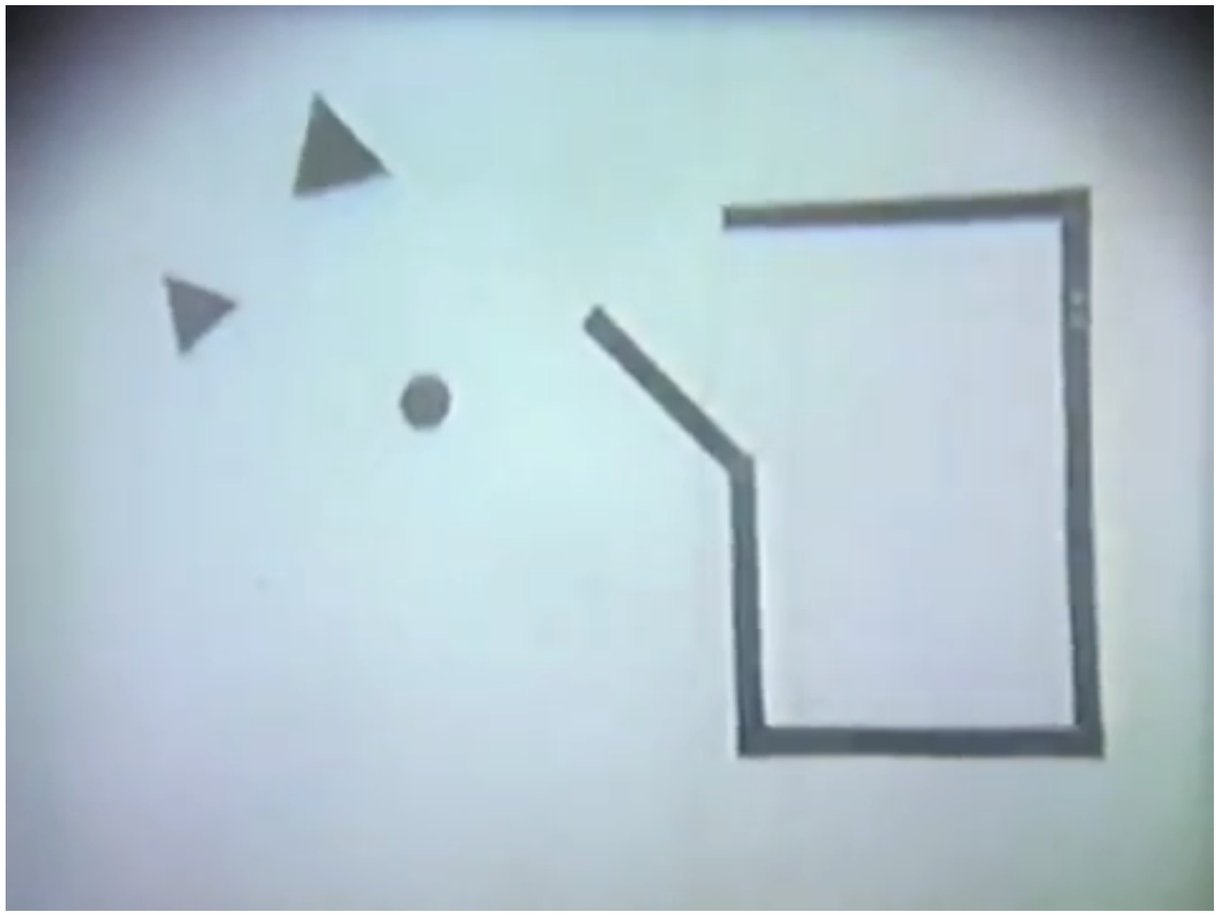
posterior: inference

likelihood: evidence

prior : expectations



Heider & Simmel, 1944



Heider & Simmel, 1944