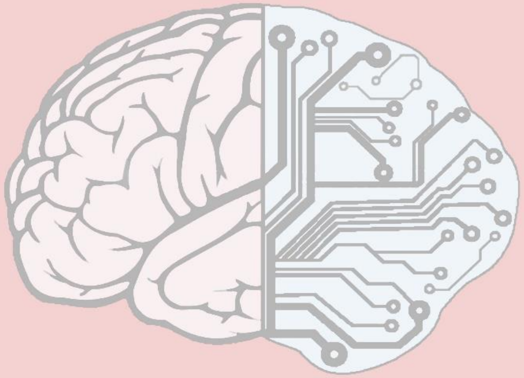


2019 Computational Psychiatry Summer (pre-)Course

Introduction to the Bayesian approach

Modelling principles in actions

Model-Based fMRI

Vincenzo G. Fiore, PhD
Mount Sinai School of Medicine

1) Belief updating: a Bayesian perspective

- A world of probabilities
- Conditional and independent probabilities
- How to assign probabilities to an hypothesis: an example
- From priors to posteriors, to new priors and again new posteriors

2) Modelling behaviour: a comparison

- RL approach: expected values, rewards, prediction error, value updates
- Bayesian approach: inference, probabilities, evidence and belief updates

3) Model Based fMRI

- Prediction error example
- Uncertainty example

Part I: A world of probabilities

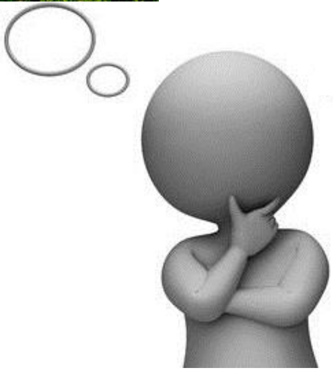
Data

Part I: A world of probabilities

Data

Hypothesis

100%?

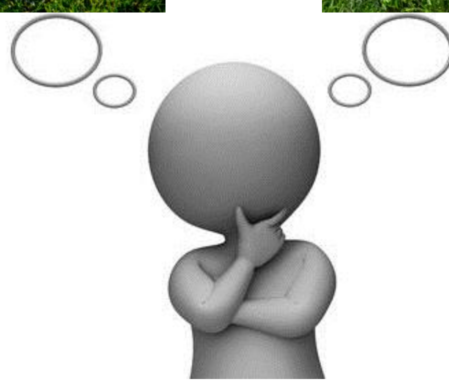


Part I: A world of probabilities

Data

Hypotheses

50%?



50%?

Part I: A world of probabilities

Data

Was the street also wet?
Are sprinkles usually working at this time of the day?
Is rain likely in this time of the year?
Bayesian Inference!

Hypotheses

25%?

75%?

Part I: Conditional and independent probabilities

Bayes' theorem:

$$P(H|D) = \frac{P(D|H)\,P(H)}{P(D)}$$

Part I: Conditional and independent probabilities

Bayes' theorem:

Conditional Probability of data (D) to occur, if the hypothesis (H) is correct.

$$P(H|D) = \frac{P(D|H)\,P(H)}{P(D)}$$

Conditional Probability of hypothesis (H) to be true, given the data (D).

Part I: Conditional and independent probabilities

Bayes' theorem:

Conditional Probability of data (D) to occur, if the hypothesis (H) is correct.

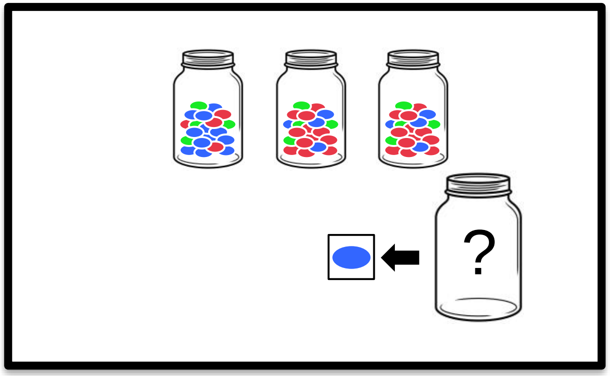
$$P(H|D) = \frac{P(D|H)\,P(H)}{P(D)}$$

Independent Probability to observe the hypothesis (H), in the environment.

Conditional Probability of hypothesis (H) to be true, given the data (D).

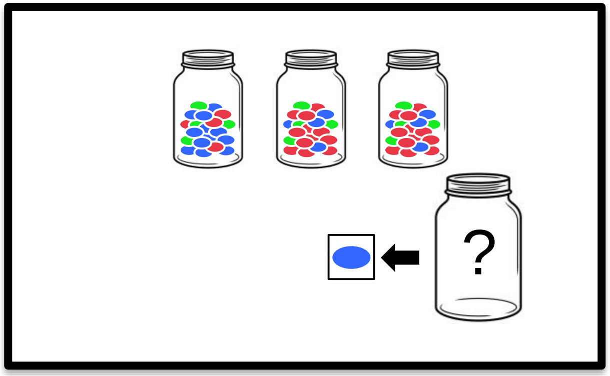
Independent Probability to observe the data (D), in the environment.

Part I: example



1. Experiment: establish from which jar the coloured bead is extracted from.
2. Evidence (data): the extracted coloured beads.
3. Hypothesis: the bead comes from the Blue jar, or any Red jar (2 hypotheses).

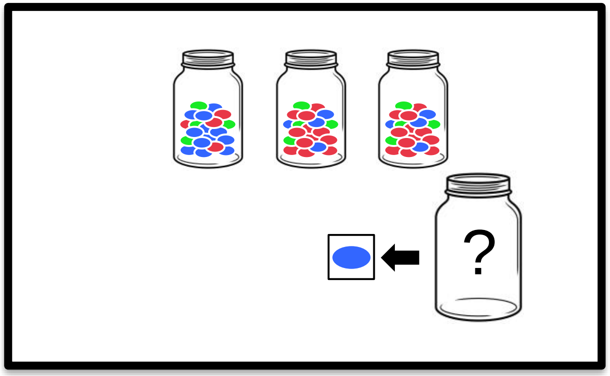
Part I: example



1. Experiment: establish from which jar the coloured bead is extracted from.
2. Evidence (data): the extracted coloured beads.
3. Hypothesis: the bead comes from the Blue jar, or any Red jar (2 hypotheses).
4. Prior probability: the distribution of probability, prior to collecting evidence.

$$\mathrm{P}(B_j) = \frac{1}{3} \qquad \mathrm{P}(R_j) = \frac{2}{3}$$

Part I: example



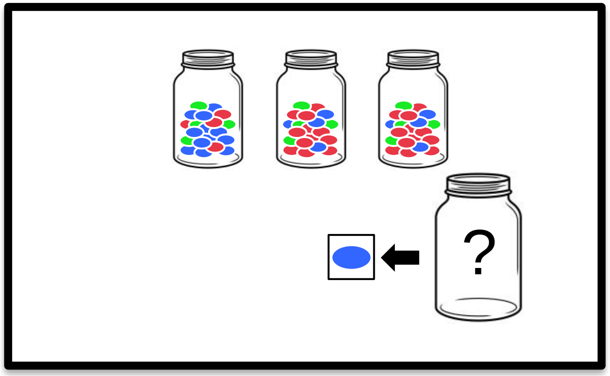
1. Experiment: establish from which jar the coloured bead is extracted from.
2. Evidence (data): the extracted coloured beads.
3. Hypothesis: the bead comes from the Blue jar, or any Red jar (2 hypotheses).
4. Prior probability: the distribution of probability, prior to collecting evidence.

$$P(B_j) = \frac{1}{3} \qquad P(R_j) = \frac{2}{3}$$

5. Likelihood: the probability of the data (bead extracted=blue), assuming each hypothesis is correct.

$$P(b|B_j)=0.8 \qquad P(b|R_j)=0.1$$

Part I: example



1. Experiment: establish from which jar the coloured bead is extracted from.
2. Evidence (data): the extracted coloured beads.
3. Hypothesis: the bead comes from the Blue jar, or any Red jar (2 hypotheses).
4. Prior probability: the distribution of probability, prior to collecting evidence.

$$\mathrm{P}(B_j) = \frac{1}{3} \qquad \mathrm{P}(R_j) = \frac{2}{3}$$

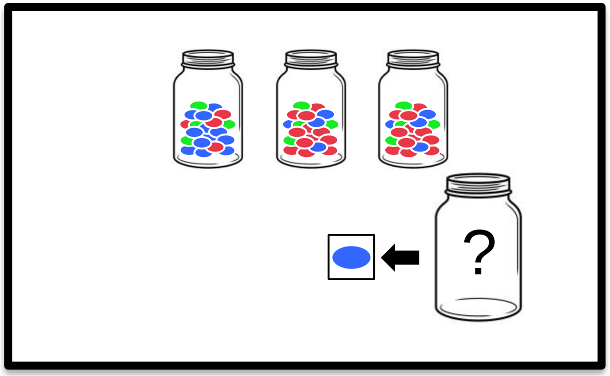
5. Likelihood: the probability of the data (bead extracted=blue), assuming each hypothesis is correct.

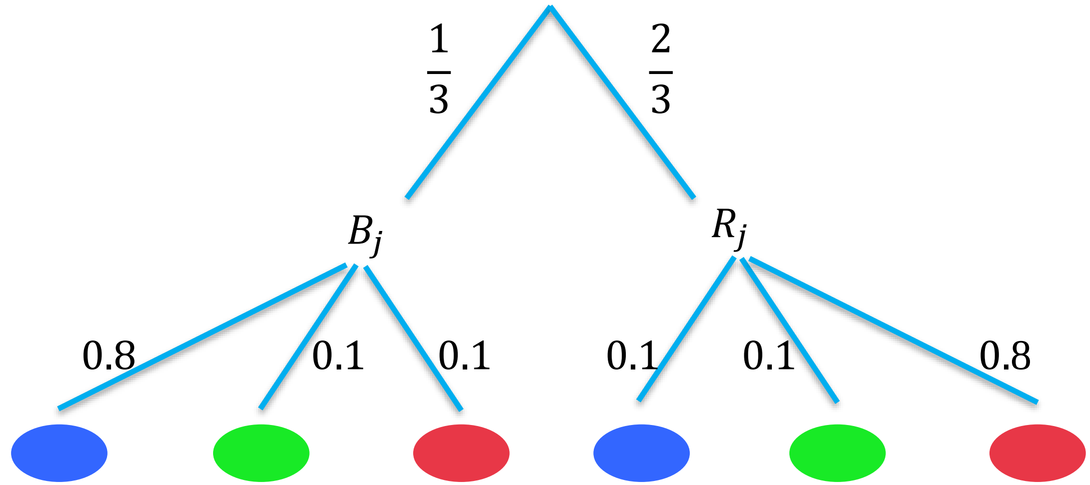
$$\mathrm{P}(\mathrm{b}|B_j)=0.8 \qquad \mathrm{P}(\mathrm{b}|R_j)=0.1$$

6. Posterior probability: the probability of each hypothesis, given the data (bead extracted=blue).

$$\mathrm{P}(B_j|\mathrm{b})=? \qquad \mathrm{P}(R_j|\mathrm{b})=?$$

Part I: example





4. Prior probability: the distribution of probability, prior to collecting evidence.

$$P(B_j) = \frac{1}{3} \qquad P(R_j) = \frac{2}{3}$$

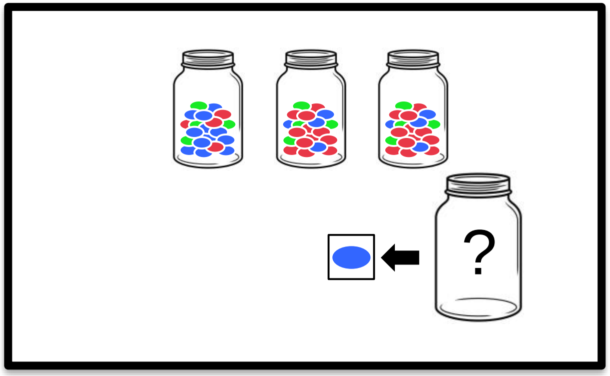
5. Likelihood: the probability of the data (bead extracted=blue), assuming each hypothesis is correct.

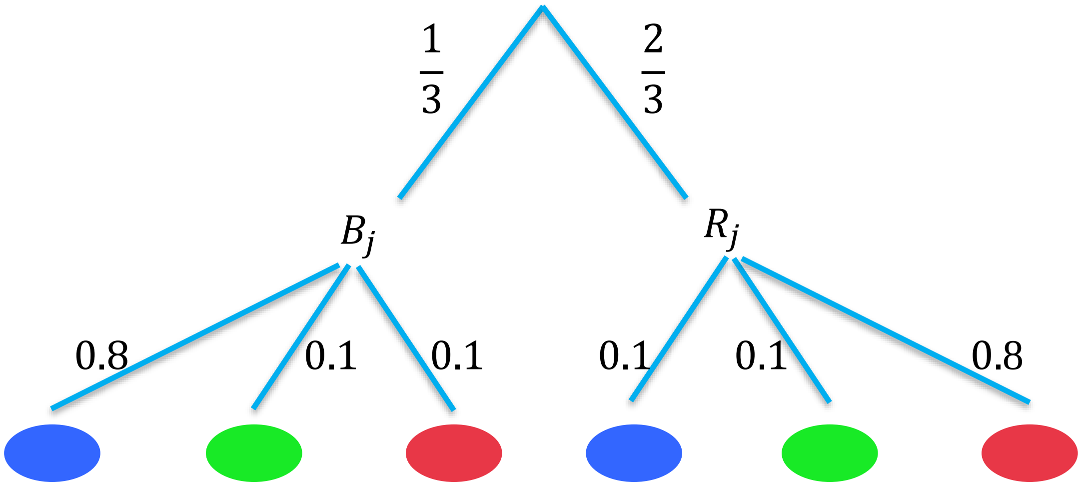
$$P(b|B_j)=0.8 \qquad P(b|R_j)=0.1$$

6. Posterior probability: the probability of each hypothesis, given the data (bead extracted=blue).

$$P(B_j|b)=? \qquad P(R_j|b)=?$$

Part I: example



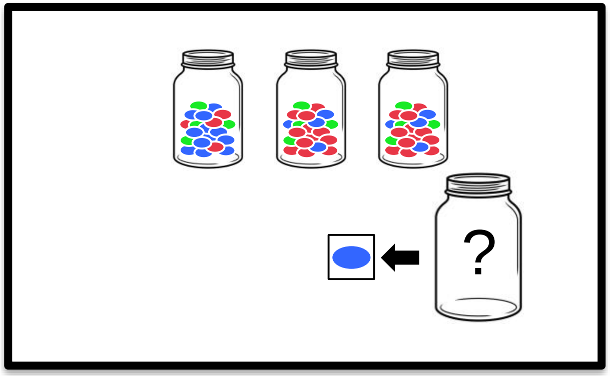


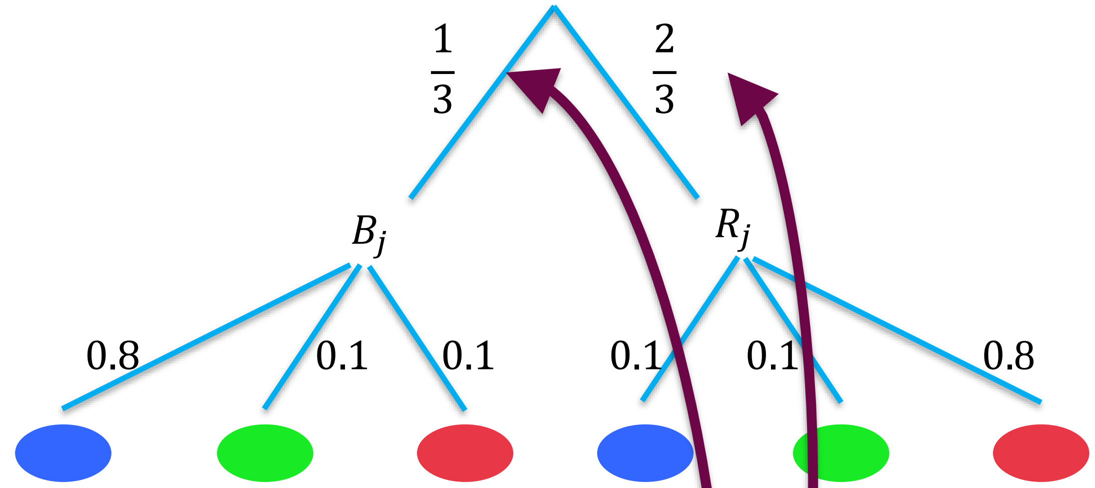
6. Posterior probability: the probability of each hypothesis, given the data (bead extracted=blue).

$$P(B_j|b) = \frac{P(b|B_j)\,P(B_j)}{P(b)} = \frac{0.8 \cdot 0.33}{(0.8 \cdot 0.33 + 0.1 \cdot 0.66)} = 0.8$$

$$P(R_j|b) = \frac{P(b|R_j)\,P(R_j)}{P(b)} = \frac{0.1 \cdot 0.66}{(0.8 \cdot 0.33 + 0.1 \cdot 0.66)} = 0.2$$

Part I: example





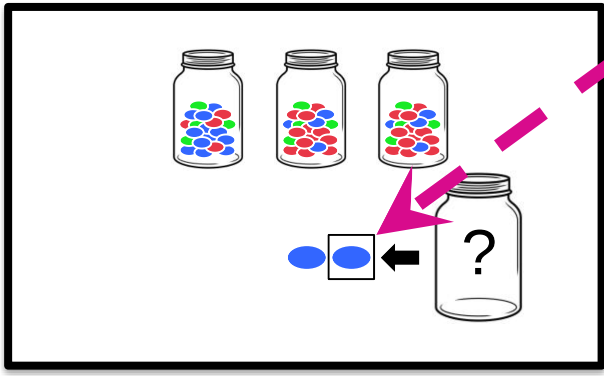
6. Posterior probability: the probability of each hypothesis, given the data (bead extracted=blue).

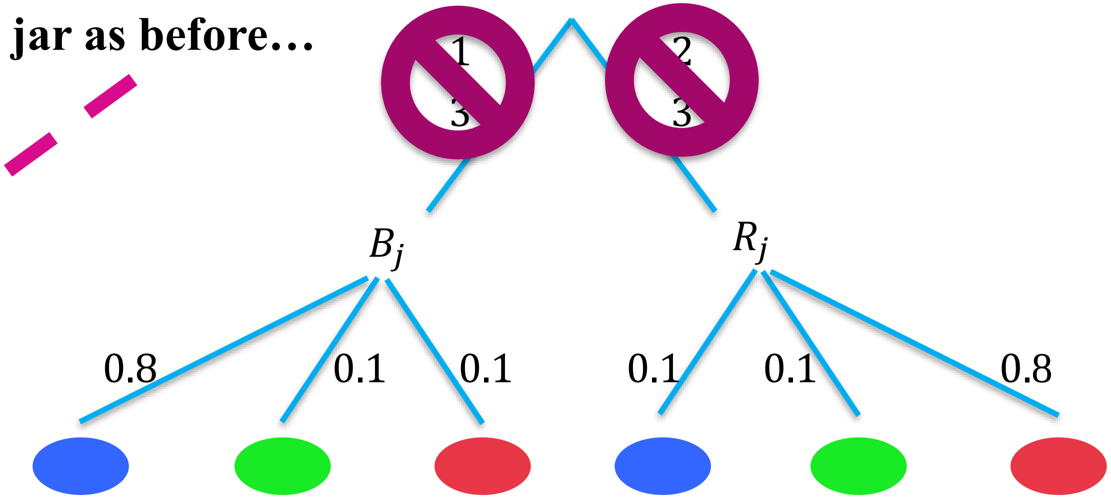
$$\mathrm{P}(B_j|b) = \frac{P(b|B_j)\,P(B_j)}{P(b)} = \frac{0.8 \cdot 0.33}{(0.8 \cdot 0.33 + 0.1 \cdot 0.66)} = 0.8$$

$$\mathrm{P}(R_j|b) = \frac{P(b|R_j)\,P(R_j)}{P(b)} = \frac{0.1 \cdot 0.66}{(0.8 \cdot 0.33 + 0.1 \cdot 0.66)} = 0.2$$

Part I: example

If extracted from the same jar as before…





6. Posterior probability: the probability of each hypothesis, given the data (beads extracted=blue+blue).

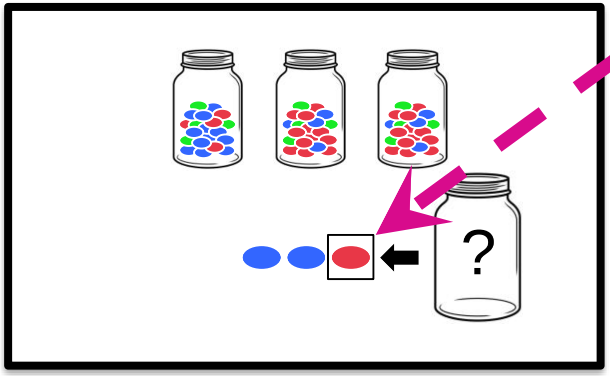
$$\mathrm{P}(B_j|b) = \frac{P(b|B_j)\,P(B_j)}{P(b)} = \frac{0.8 \cdot \mathbf{0.8}}{(0.8 \cdot \mathbf{0.8} + 0.1 \cdot \mathbf{0.2})} = 0.97$$

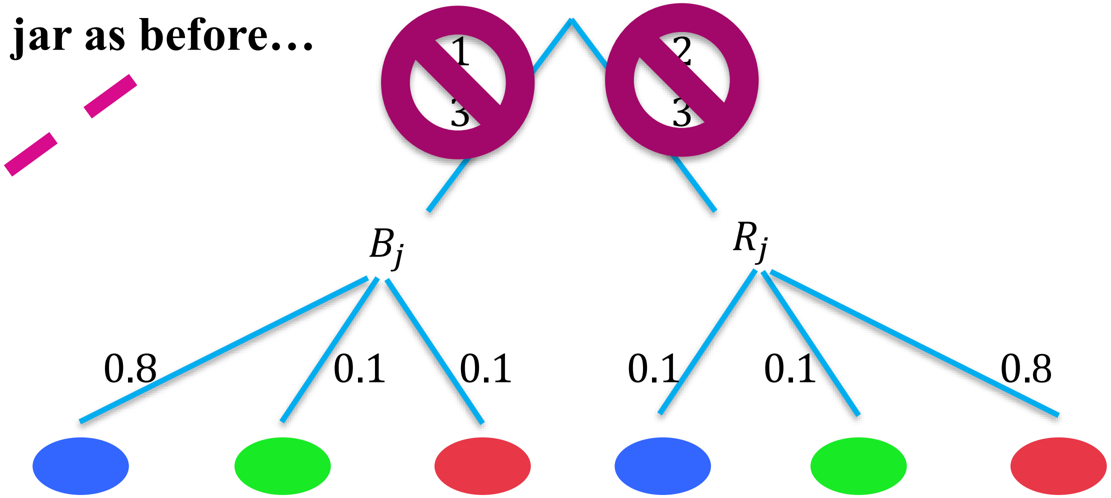
$$\mathrm{P}(R_j|b) = \frac{P(b|R_j)\,P(R_j)}{P(b)} = \frac{0.1 \cdot \mathbf{0.2}}{(0.8 \cdot \mathbf{0.8} + 0.1 \cdot \mathbf{0.2})} = 0.03$$

…priors must be updated!

Part I: example

If extracted from the same jar as before…





6. Posterior probability: the probability of each hypothesis, given the data (beads extracted=blue+blue+red).

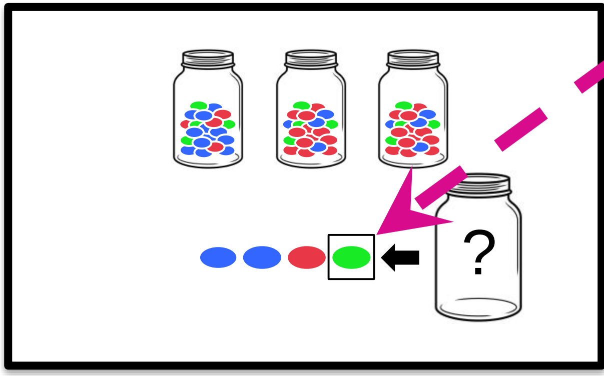
$$P(B_j|r) = \frac{P(r|B_j)\,P(B_j)}{P(r)} = \frac{0.1 \cdot \mathbf{0.97}}{(0.1 \cdot \mathbf{0.97} + 0.8 \cdot \mathbf{0.03})} = 0.8$$

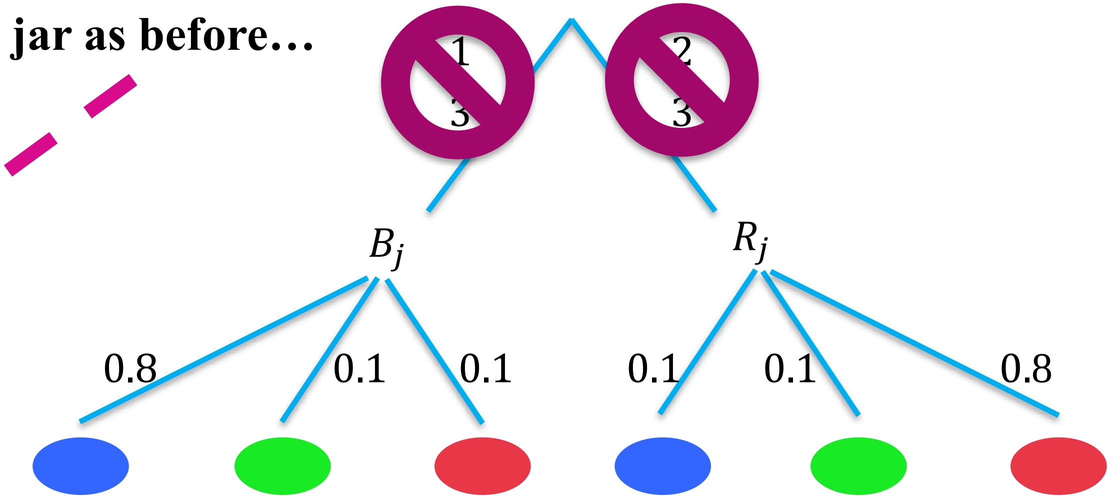
$$P(R_j|r) = \frac{P(r|R_j)\,P(R_j)}{P(r)} = \frac{0.8 \cdot \mathbf{0.03}}{(0.1 \cdot \mathbf{0.97} + 0.8 \cdot \mathbf{0.03})} = 0.2$$

…priors must be updated!

Part I: example

If extracted from the same jar as before…





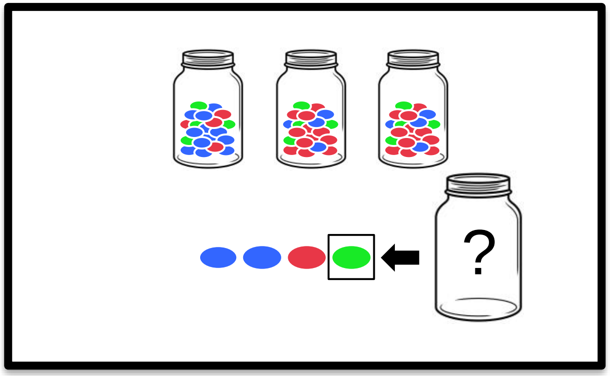
6. Posterior probability: the probability of each hypothesis, given the data (beads extracted=blue+blue+red+green).

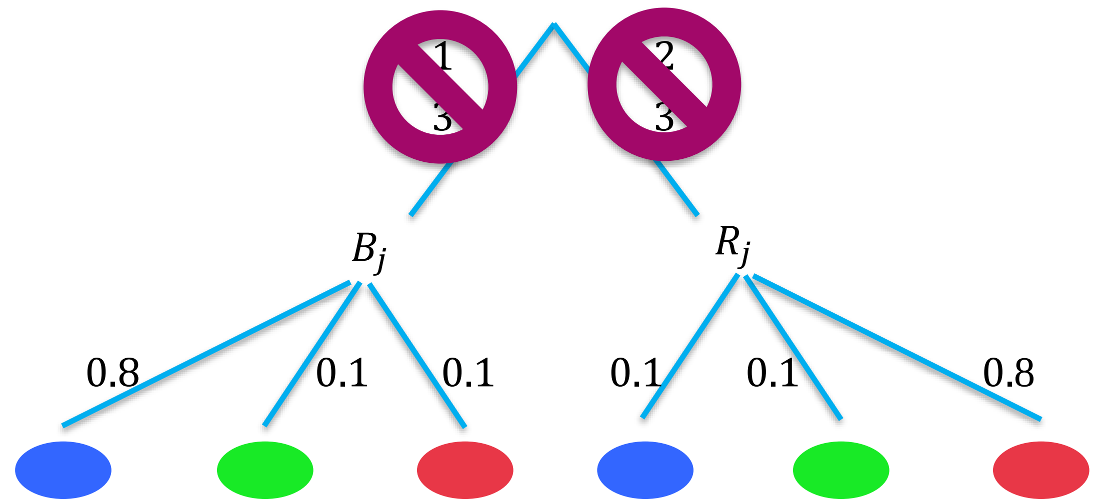
$$P(B_j|g) = \frac{P(g|B_j)\,P(B_j)}{P(g)} = \frac{0.1 \cdot \mathbf{0.8}}{(0.1 \cdot \mathbf{0.8} + 0.1 \cdot \mathbf{0.2})} = 0.8$$

$$P(R_j|g) = \frac{P(r|R_j)\,P(R_j)}{P(g)} = \frac{0.1 \cdot \mathbf{0.2}}{(0.1 \cdot \mathbf{0.8} + 0.1 \cdot \mathbf{0.2})} = 0.2$$

…priors must be updated!

Part I: example





6. Posterior probability: the probability of each hypothesis, given the data (beads extracted=blue+blue+red+green).

$$P(B_j|g) = \frac{P(g|B_j)\,P(B_j)}{P(g)} = \frac{0.1 \cdot \mathbf{0.8}}{(0.1 \cdot \mathbf{0.8} + 0.1 \cdot \mathbf{0.2})} = 0.8$$

$$P(R_j|g) = \frac{P(r|R_j)\,P(R_j)}{P(g)} = \frac{0.1 \cdot \mathbf{0.2}}{(0.1 \cdot \mathbf{0.8} + 0.1 \cdot \mathbf{0.2})} = 0.2$$

{ *Data can be meaningless*!

Part I Summary

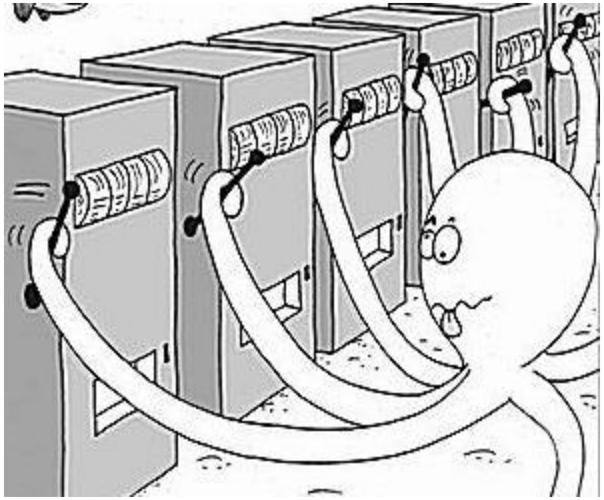
- Biological agents continuously collect information from the environment to form and update their own beliefs.
- In Bayesian terms, beliefs are organized as distribution of probabilities.
- These distributions can be estimated using Bayes' theorem, assuming:
  - optimal behaviour, relative to the objectives and the information available.
  - Probability distributions for all events in the environment are known… **or guessed?**

Part II: modelling principles in action

A two armed bandit example

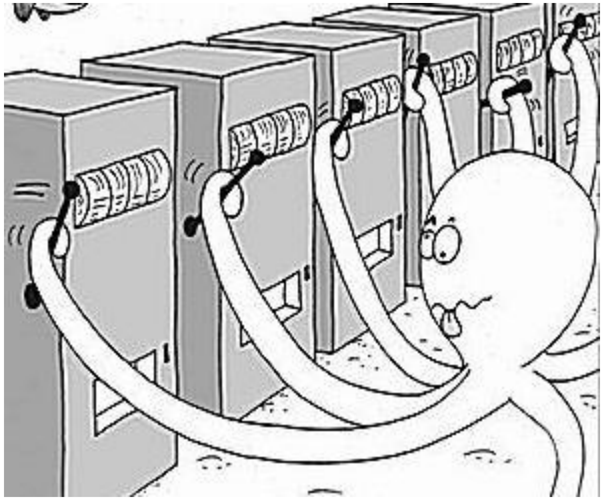
1. RL approach: expected values, rewards, prediction error, value updates
2. Bayesian approach: inference, probabilities, evidence and belief updates

Part II: modelling principles in action

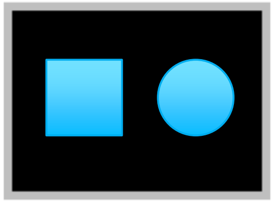


The multi armed bandit

Part II: modelling principles in action

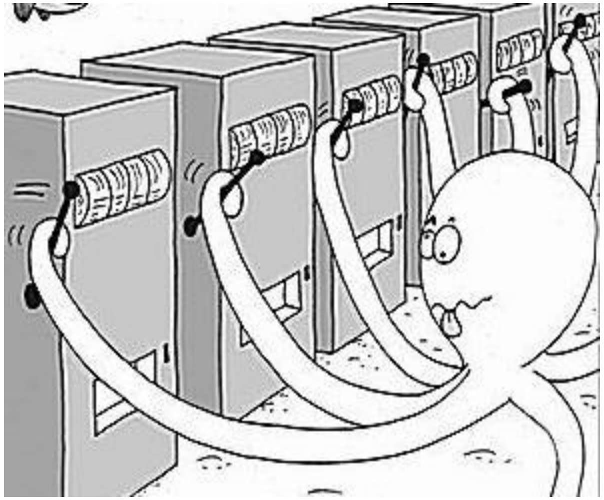


The multi armed bandit

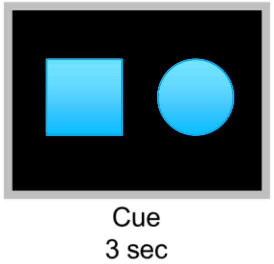


State 1: two actions available

Part II: modelling principles in action

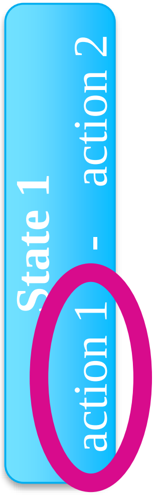


The multi armed bandit

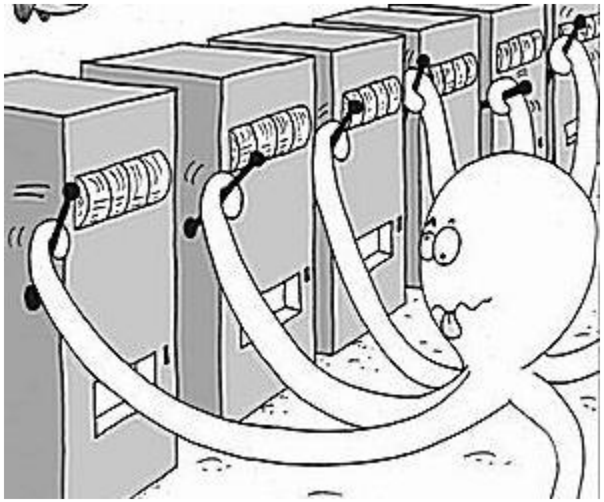


State 1: two actions available

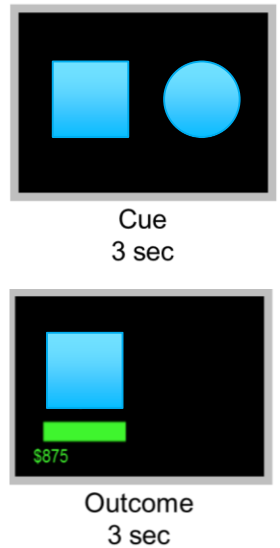
The agent selects one action



Part II: modelling principles in action



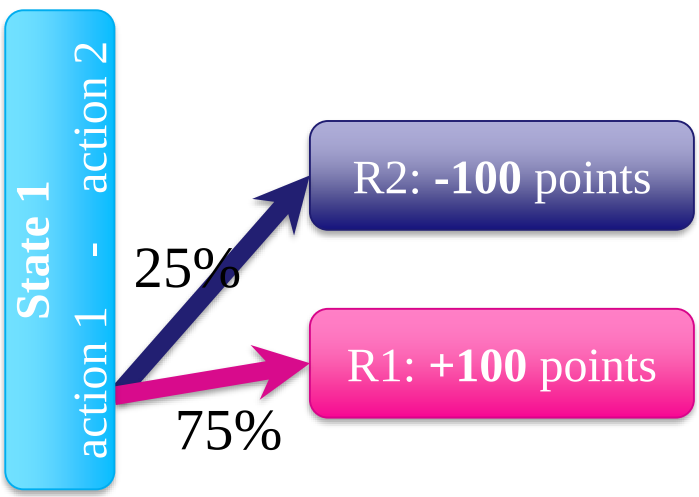
The multi armed bandit



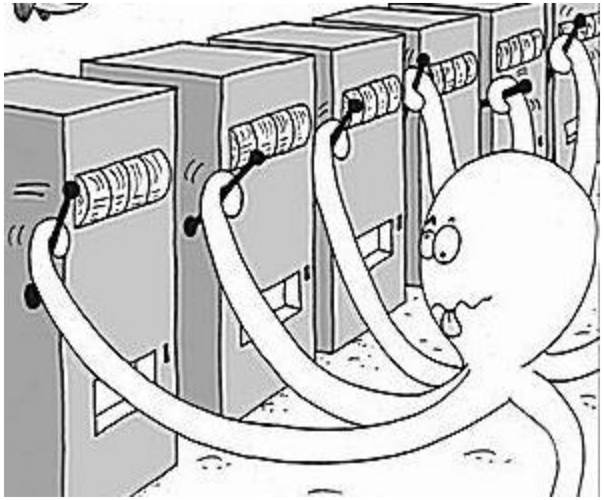
State 1: two actions available

The agent selects one action

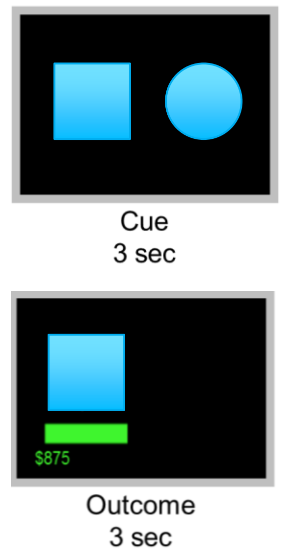
State 2: stochastic reward



Part II: modelling principles in action



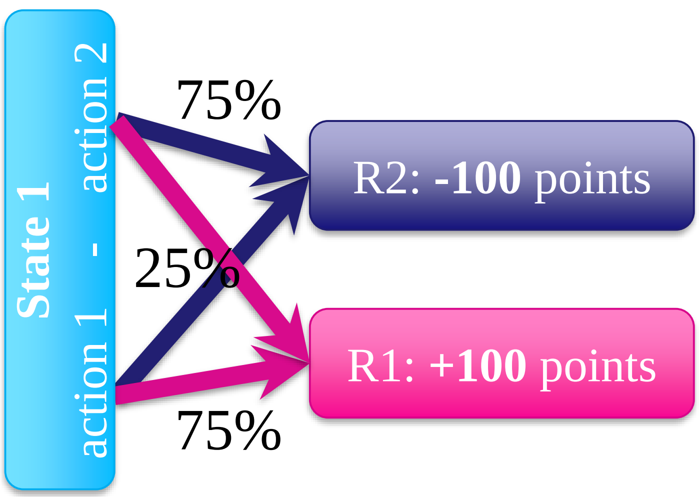
The multi armed bandit



State 1: two actions available

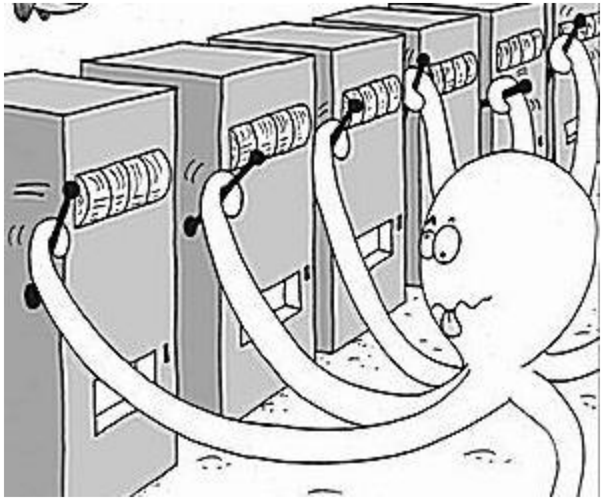
The agent selects one action

State 2: stochastic reward

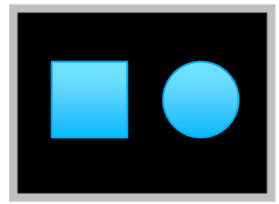


Classic reward configuration,
perfect for RL model

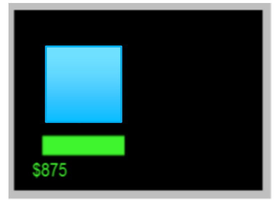
Part II: modelling principles in action



The multi armed bandit



Cue
3 sec

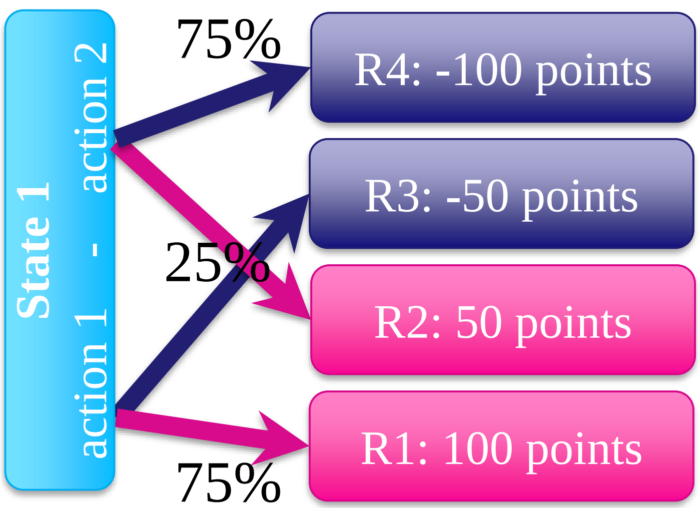


Outcome
3 sec

State 1: two actions available

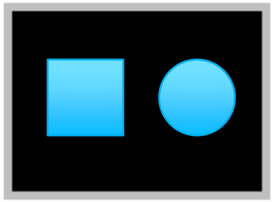
The agent selects one action

State 2: stochastic reward



Unconventional reward configuration:
R1: high positive outcome
R2: low positive outcome
R3: low negative outcome
R4: high negative outcome

Part II: modelling principles in action



Cue
3 sec

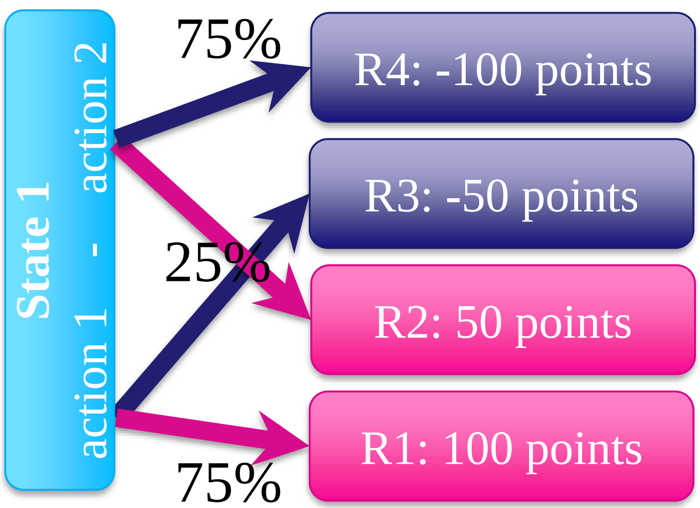
Old value for the selected action

$$Q_{a_{t+1}} = Q_{a_t} + \alpha\left(r_t - Q_{a_t}\right)$$

Updated value for the selected action

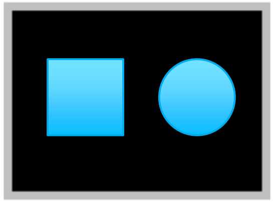
Prediction error (reward – old value)

Trial n	Q value Action 1	reward
1	0	100
2	50	100
3	75	-50
4	12.5	100
5	56.25	50
6	53.12	-100
7	-23.43	-100
8	-61.71	-100
9	-80.85	-50
10	-65.42	



Expected value optimal choice: 62.5
Expected value suboptimal choice: -62.5

Part II: modelling principles in action



Cue
3 sec

Learning rate

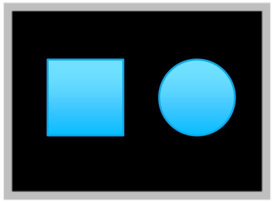
$$Q_{a_{t+1}} = Q_{a_t} + \alpha(r_t - Q_{a_t})$$

Trial n	Q value Action 1	reward
1	0	100
2	50	100
3	75	-50
4	12.5	100
5	56.25	50
6	53.12	-100
7	-23.43	-100
8	-61.71	-100
9	-80.85	-50
10	-65.42	

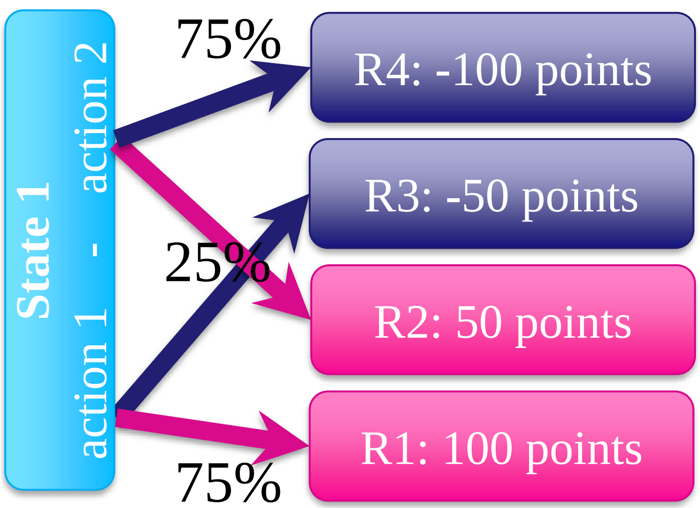
The free parameters regulate the behaviour of the artificial agent.
To describe the behaviour of actual participants, we explore which values for the free parameters best fit their behaviour.
E.g. in the example, $\boldsymbol{\alpha = 0.5}$

Expected value optimal choice: 62.5
Expected value suboptimal choice: -62.5

Part II: modelling principles in action



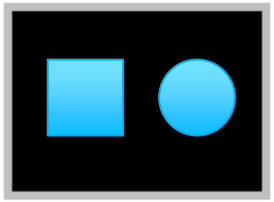
If the structure of reward is known. this is sufficient evidence to change policy, immediately.



Trial n	Q value Action 1	reward
1	0	100
2	50	100
3	75	-50
4	12.5	100
5	56.25	50
6	53.12	-100
7	-23.43	-100
8	-61.71	-100
9	-80.85	-50
10	-65.42	

Expected value optimal choice: 62.5
Expected value suboptimal choice: -62.5

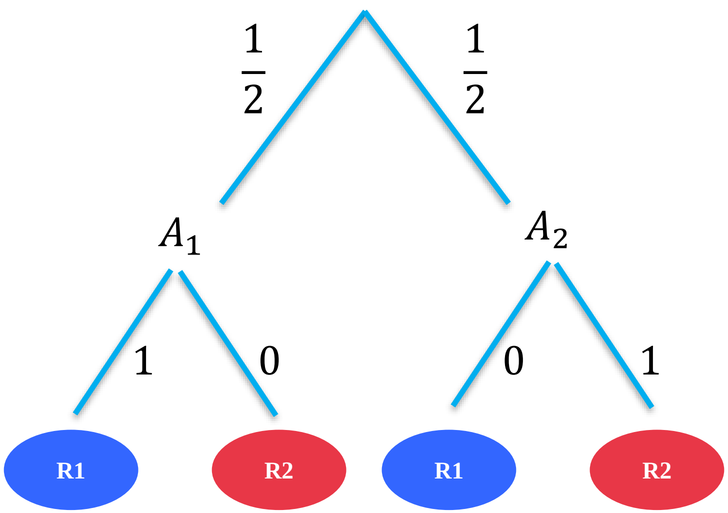
Part II: modelling principles in action



Cue
3 sec

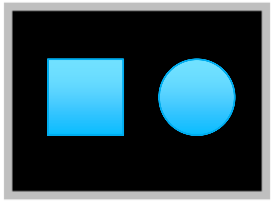
$$P(A_1|R_1) = \frac{P(R_1|A_1)\,P(A_1)}{P(R_1)} = \frac{1 \cdot 0.5}{(1 \cdot 0.5 + 0 \cdot 0.5)} = 1$$

Trial n	Bayesian update	reward
1	0.5	100 (R_1)
2	1	100(R_1)
3	1	-50(R_1)
4	1	100(R_1)
5	1	50(R_2)
6	0	-100(R_2)
7	0	-100(R_2)
8	0	-100(R_2)
9	0	-50(R_1)
10	1	



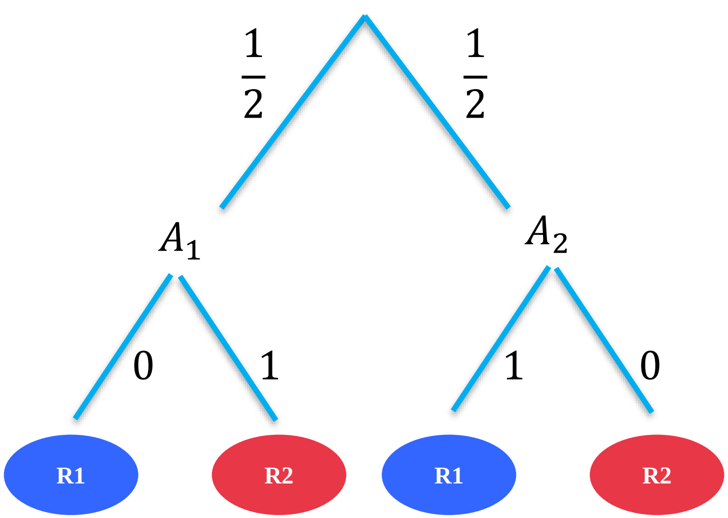
Data is unequivocal, so the probability distribution is immediately updated to 100% vs 0%

Part II: modelling principles in action



Cue
3 sec

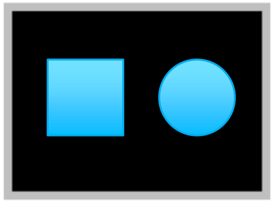
$$P(A_1|R_2) = \frac{P(R_2|A_1)\,P(A_1)}{P(R_2)}$$
$$= \frac{1 \cdot 0.5}{(1 \cdot 0.5 + 0 \cdot 0.5)} = 1$$



Trial n	Bayesian update	reward
1	0.5	100 (R_1)
2	1	100(R_1)
3	1	-50(R_1)
4	1	100(R_1)
5	1	50(R_2)
6	0	-100(R_2)
7	0	-100(R_2)
8	0	-100(R_2)
9	0	-50(R_1)
10	1	

Data is unequivocal, so the probability distribution is immediately updated to 100% vs 0%

Part II: modelling principles in action

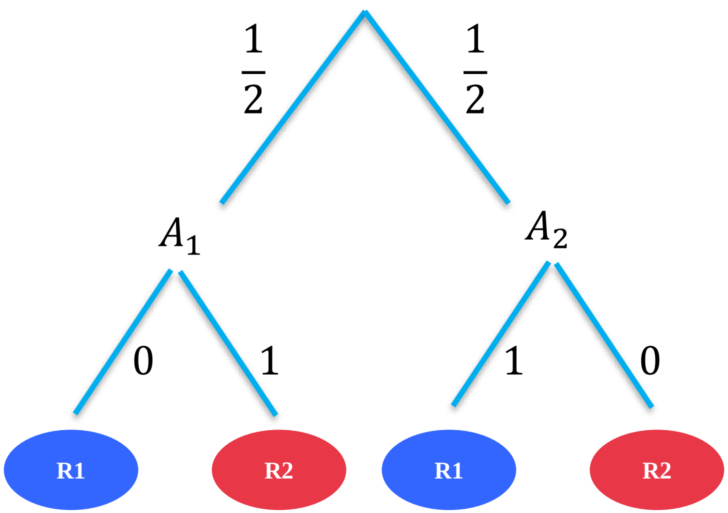


Cue
3 sec

$$P(A_1|R_2) = \frac{P(R_2|A_1)\,P(A_1)}{P(R_2)}$$

$$= \frac{1 \cdot 0.5}{(1 \cdot 0.5 + 0 \cdot 0.5)} = 1$$

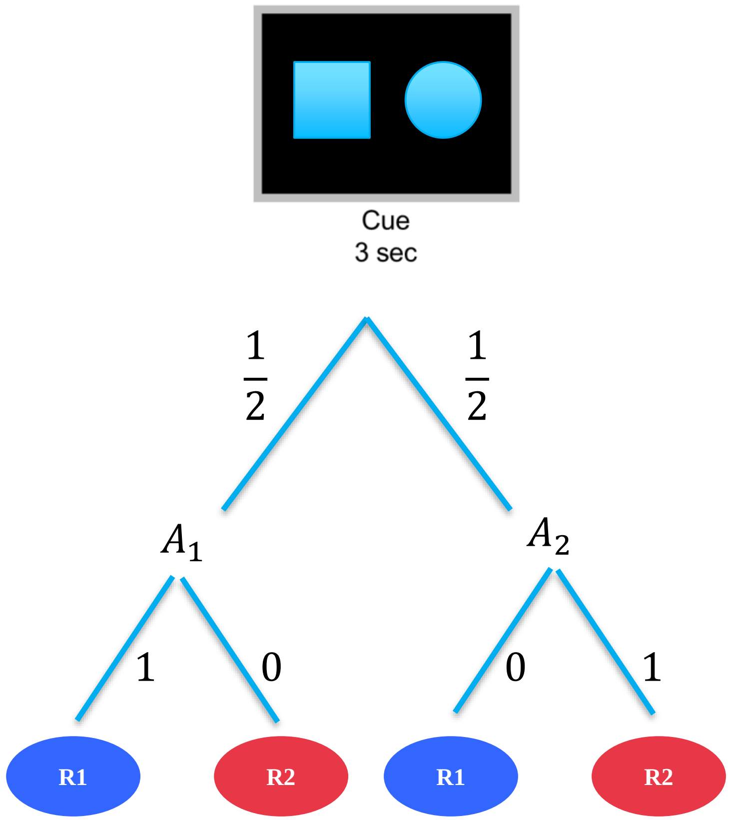
$$\boldsymbol{P(R_1|A_1) = 0}$$

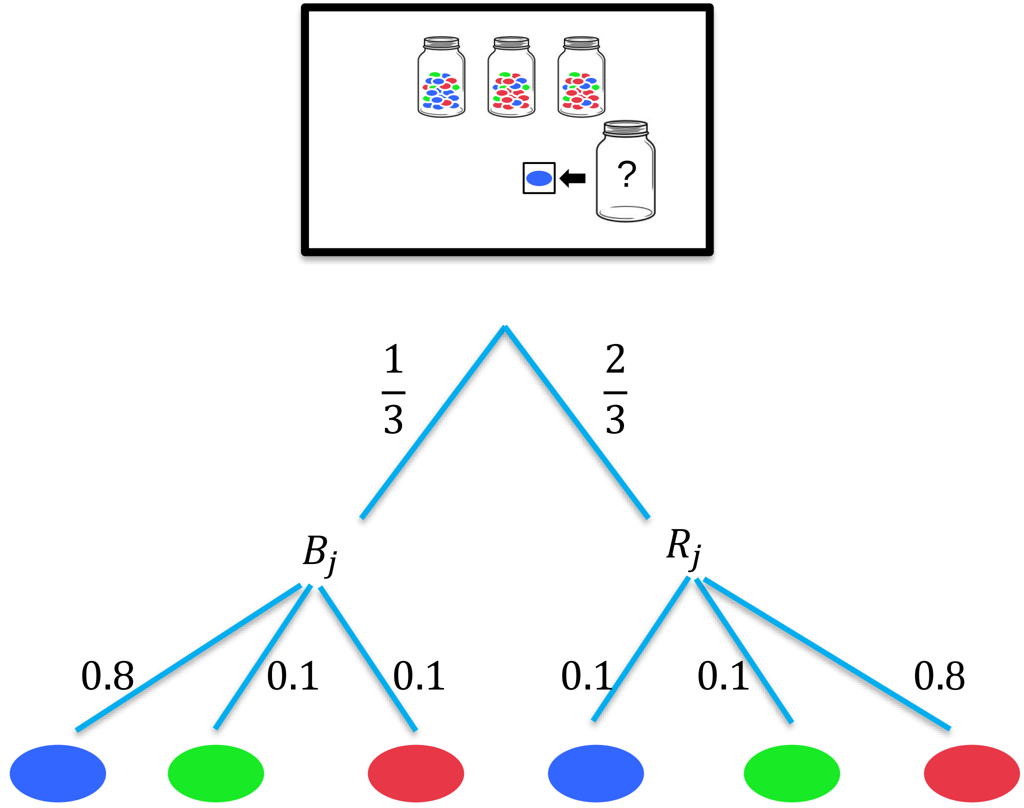


Trial n	Bayesian update	reward
1	0.5	100 (R_1)
2	1	100(R_1)
3	1	-50(R_1)
4	1	100(R_1)
5	1	50(R_2)
6	0	-100(R_2)
7	0	-100(R_2)
8	0	-100(R_2)
9	0	-50(R_1)
10	1	

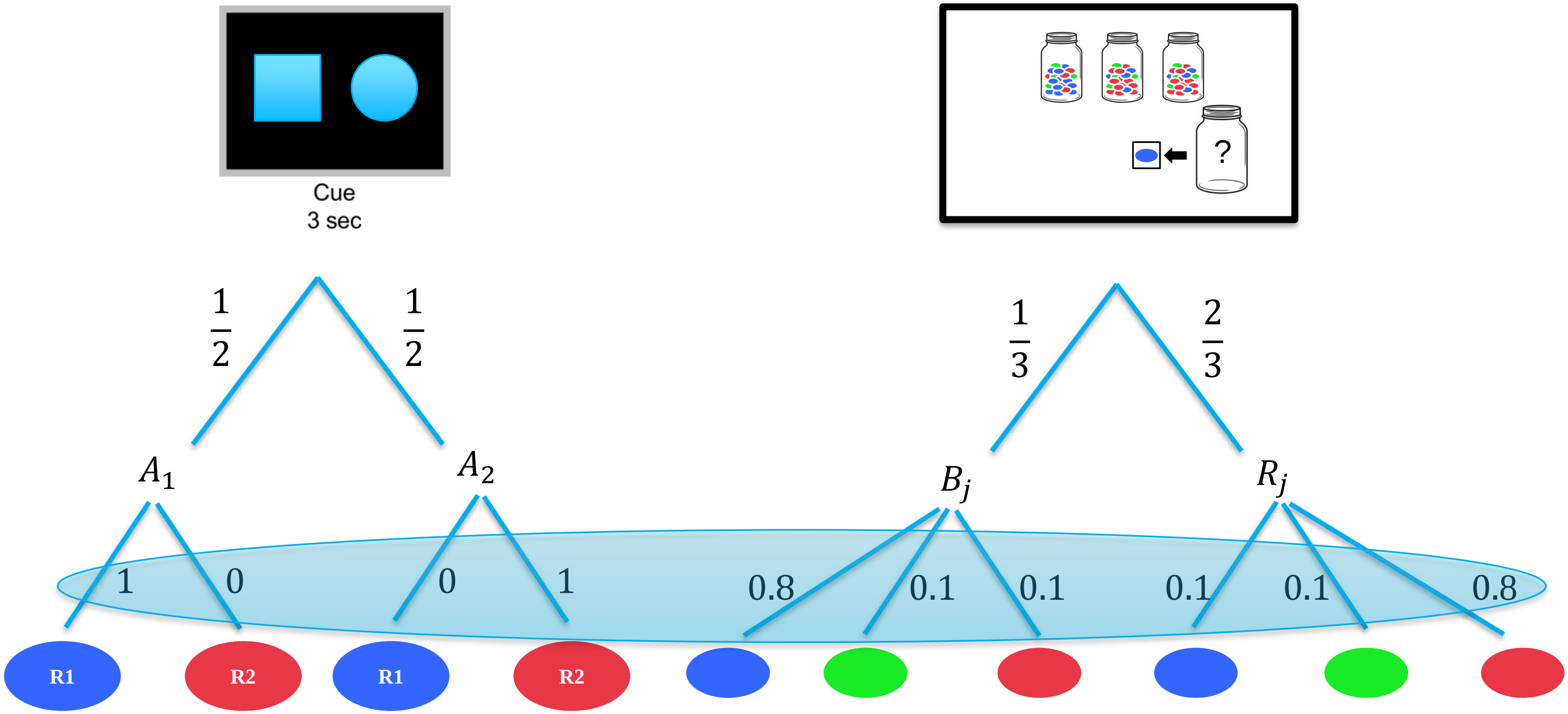
Data is unequivocal, so the probability distribution is immediately updated to 100% vs 0%

Part II: modelling principles in action





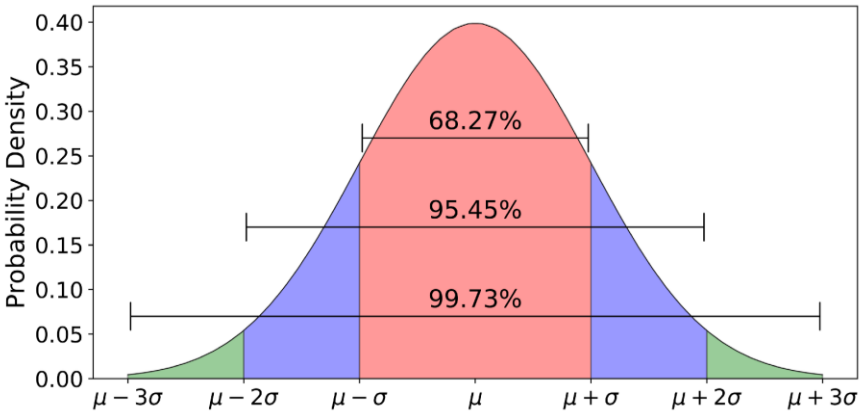
Part II: modelling principles in action

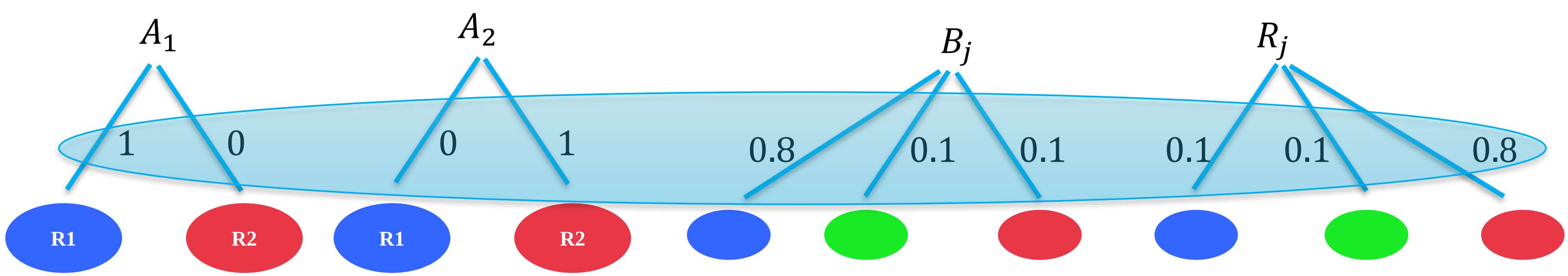


What if these probability distributions are not known?
(i.e. real world scenario)

Part II: modelling principles in action

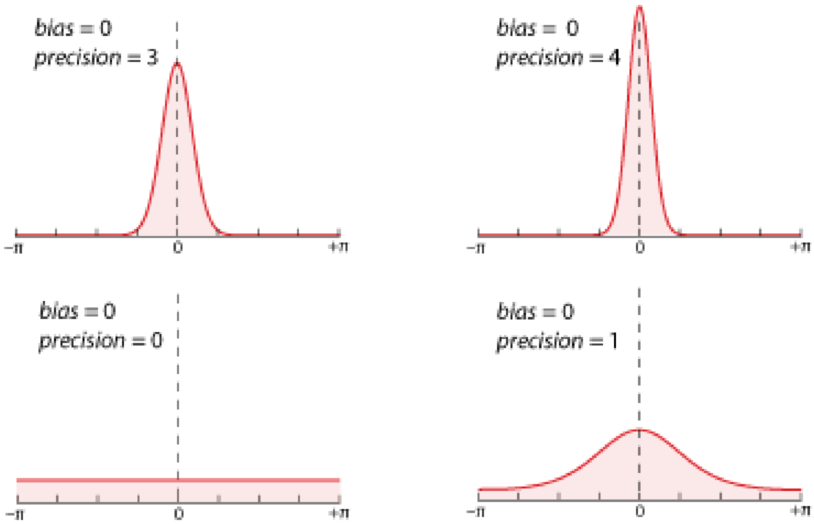
Bayesian modelling usually assumes that the beliefs of biological agents are normally distributed.
Thus, subjects differ depending on their assumptions about how the events are distributed (i.e. values for the standard deviation, σ).

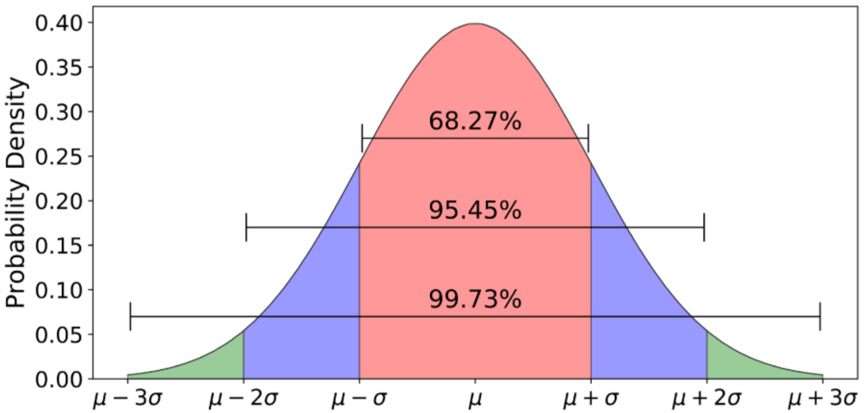


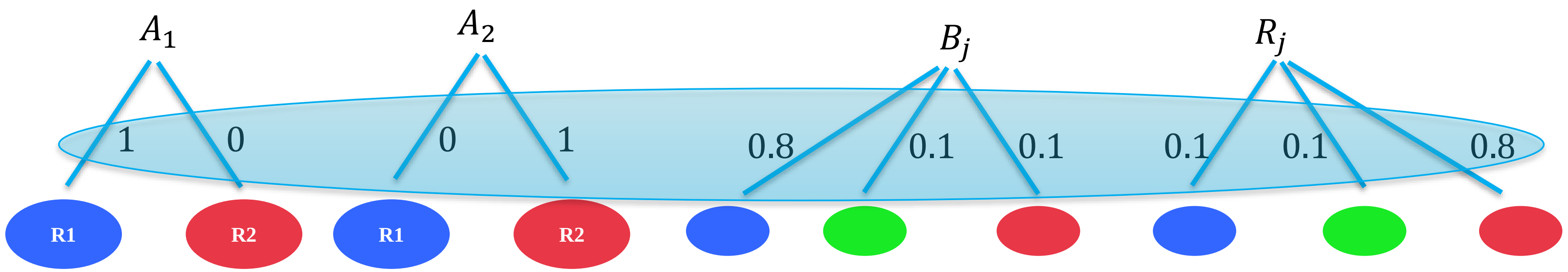


What if these probability distributions are not known?
(i.e. real world scenario)

Part II: modelling principles in action



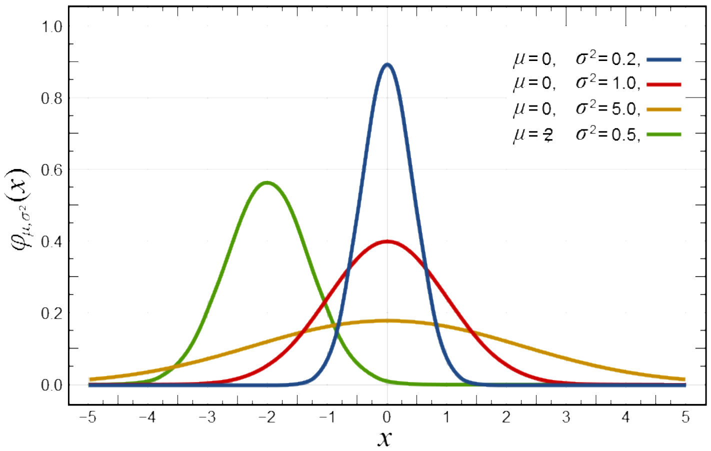


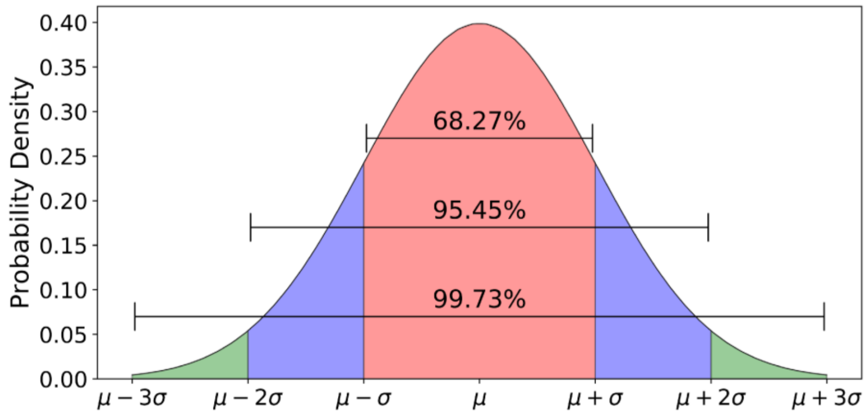


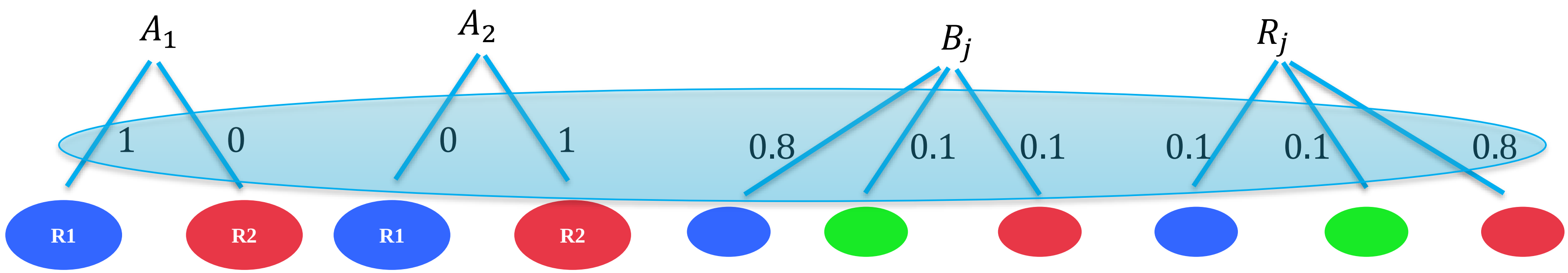
The lower the σ, the higher the "precision".

Part II: modelling principles in action

Probability density function



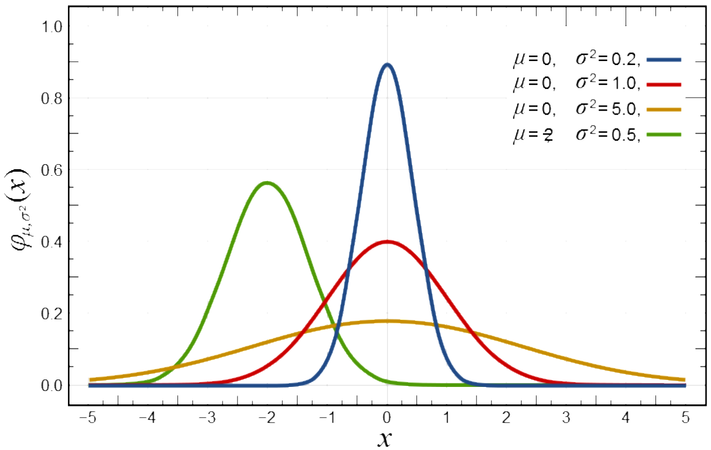




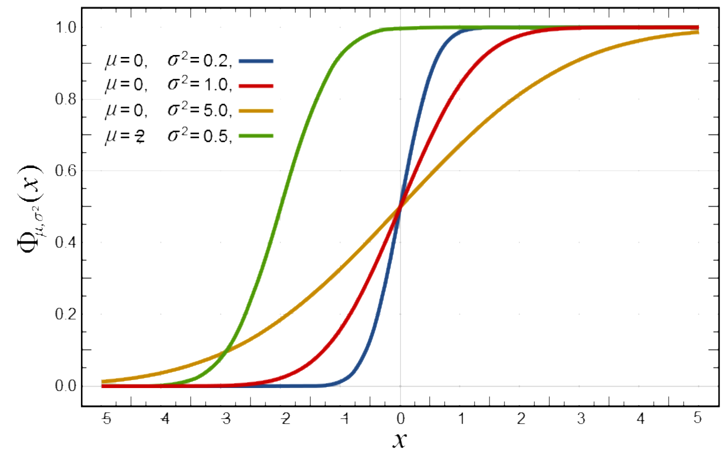
The lower the σ, the higher the "precision".

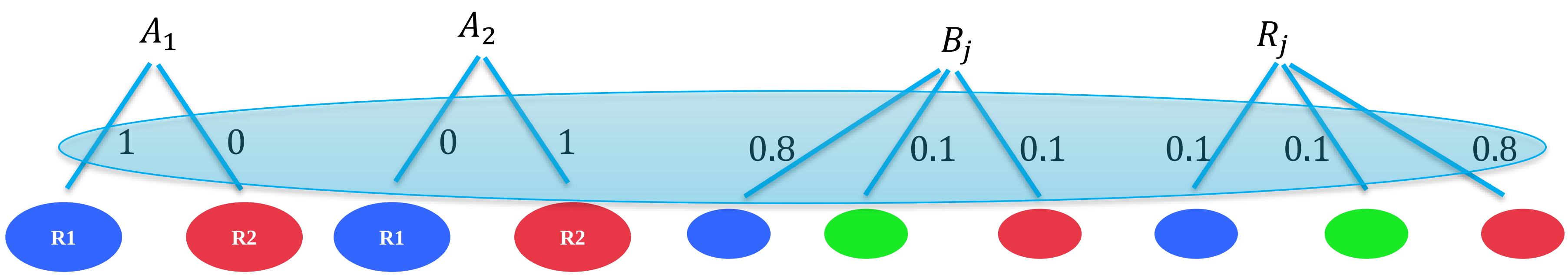
Part II: modelling principles in action

Probability density function



Cumulative distribution function

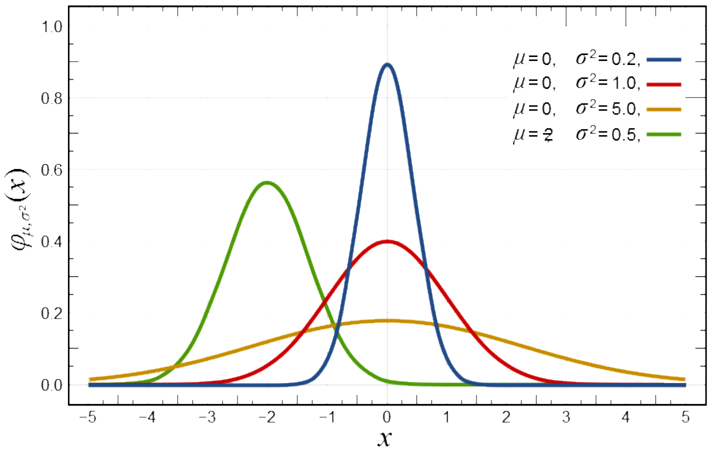




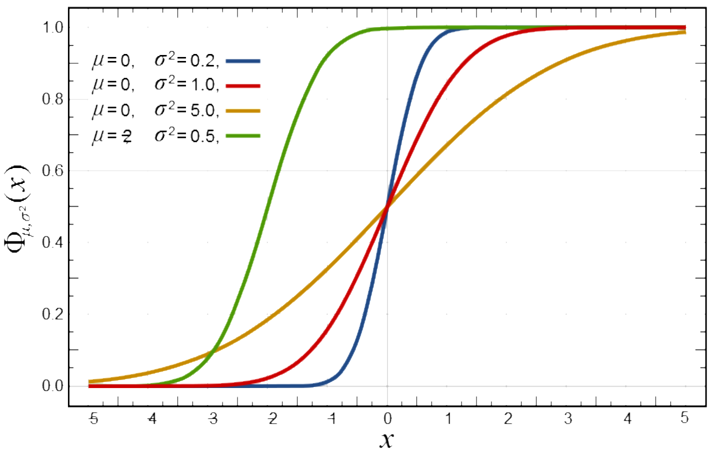
The lower the σ, the higher the "precision".
The higher the "precision", the faster the update
(akin high values for the learning rate, with a twist!).

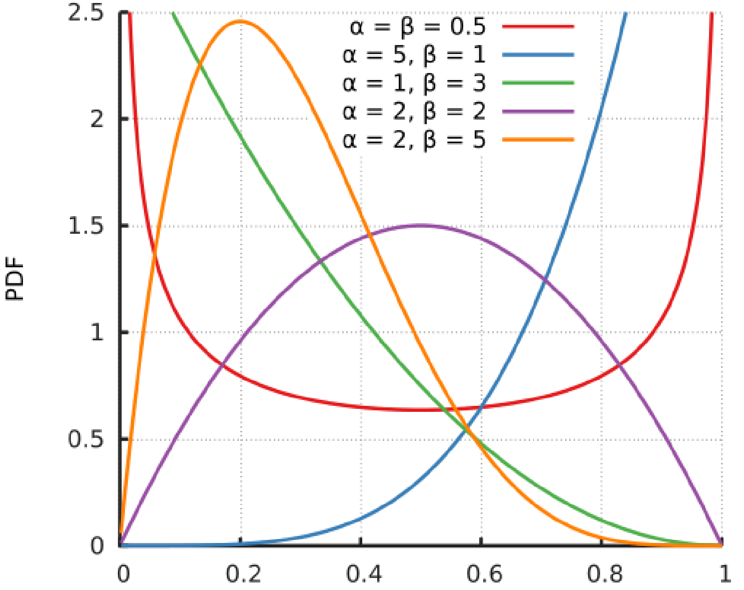
Part II: modelling principles in action

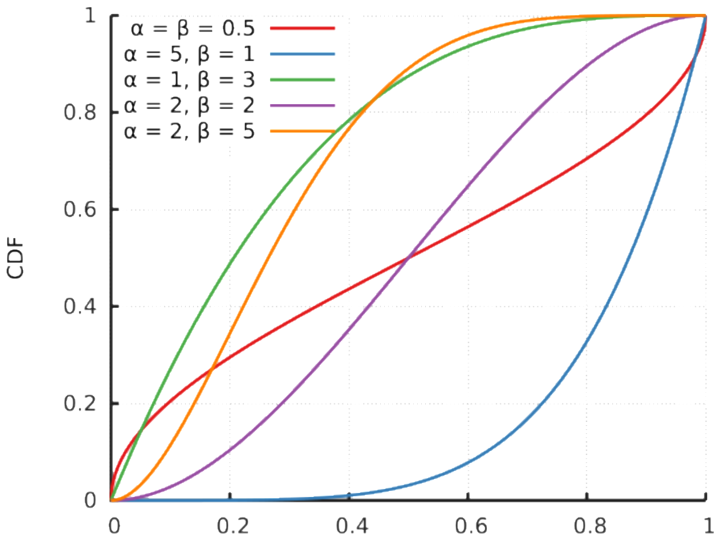
Probability density function



Cumulative distribution function





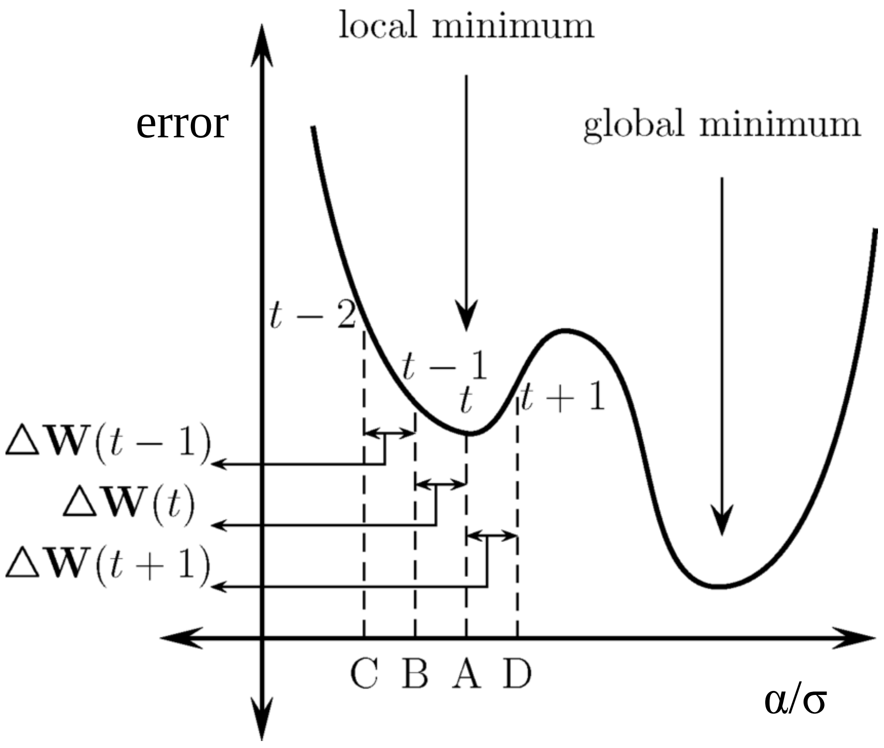


Part II: modelling principles in action

Hunting for chimeras: space of parameters, *optimal* values, research methods.

$$Q_{a_{t+1}} = Q_{a_t} + \alpha(r_t - Q_{a_t})$$

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



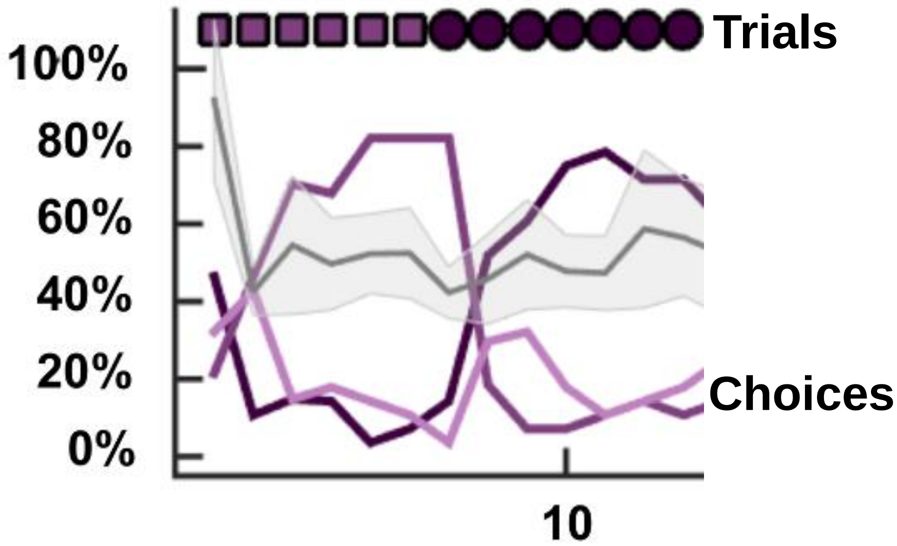
Part II Summary

- Both RL and Bayesian models attempt to establish a mechanism to “solve” the given problem and find an optimal behaviour that would maximise the given objective.
- The behaviour expressed by the model is then tuned to the specific choice selections of each participant, establishing their update pace, on the basis of subject-specific parameters.
- Different environments pose different challenges to the same models, making some constructs more or less likely to fit the behaviour of the biological agents.

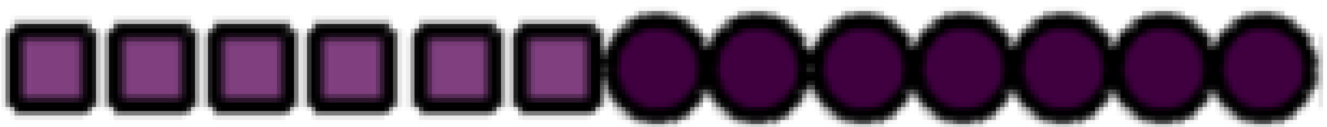
Part III: model based fMRI

1. Contrast based analysis
2. RL approach: expected values, rewards, prediction error, value updates
3. Bayesian approach: inference, probabilities, evidence and belief updates

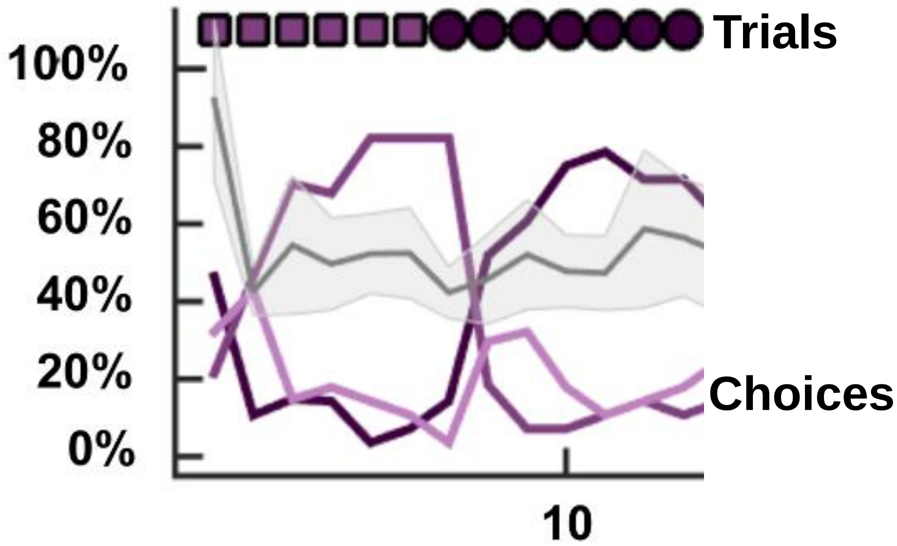
Example: Contrast analysis

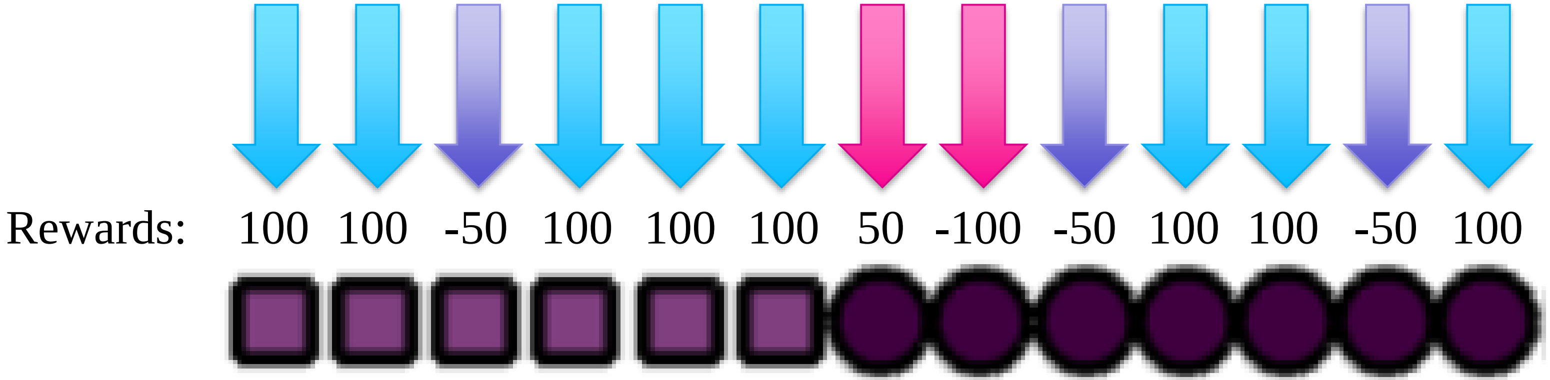


Rewards: 100 100 -50 100 100 100 50 -100 -50 100 100 -50 100

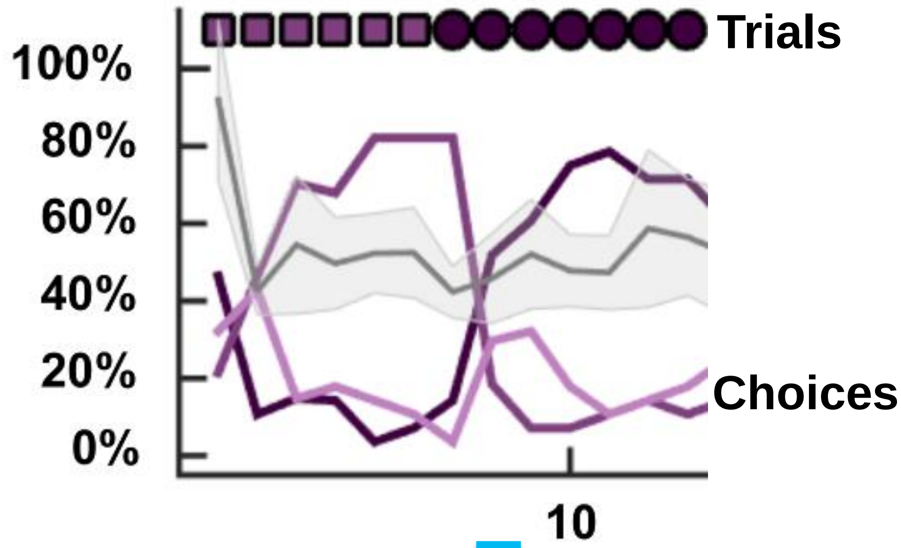


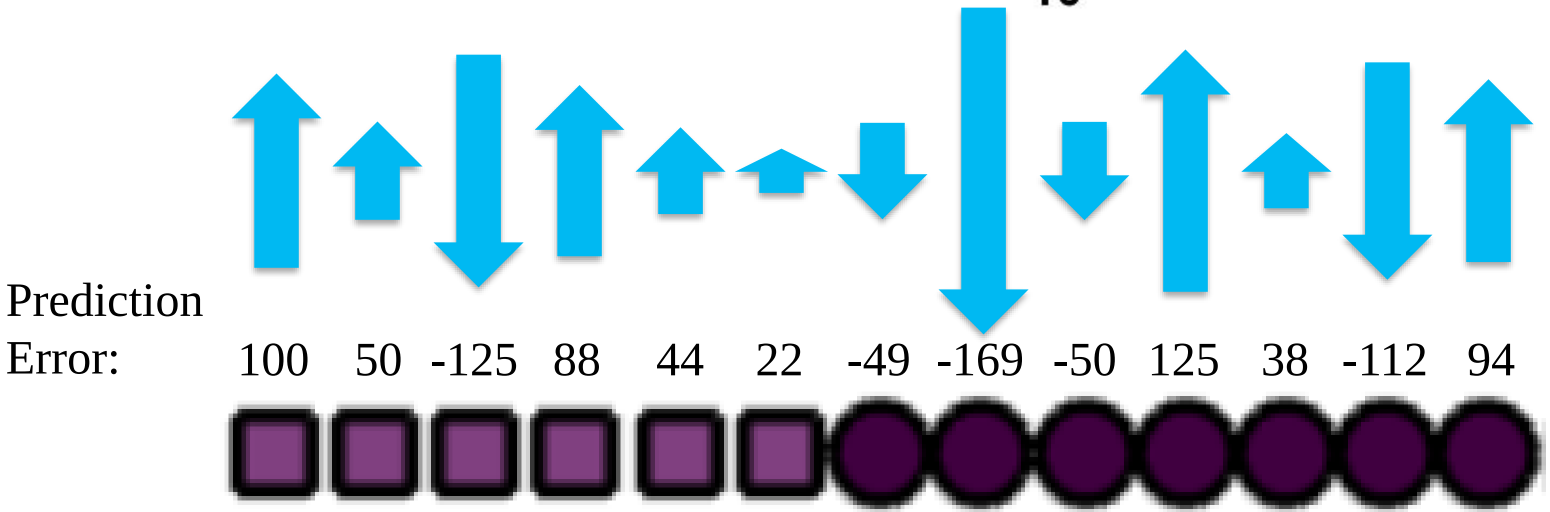
Example: Contrast analysis



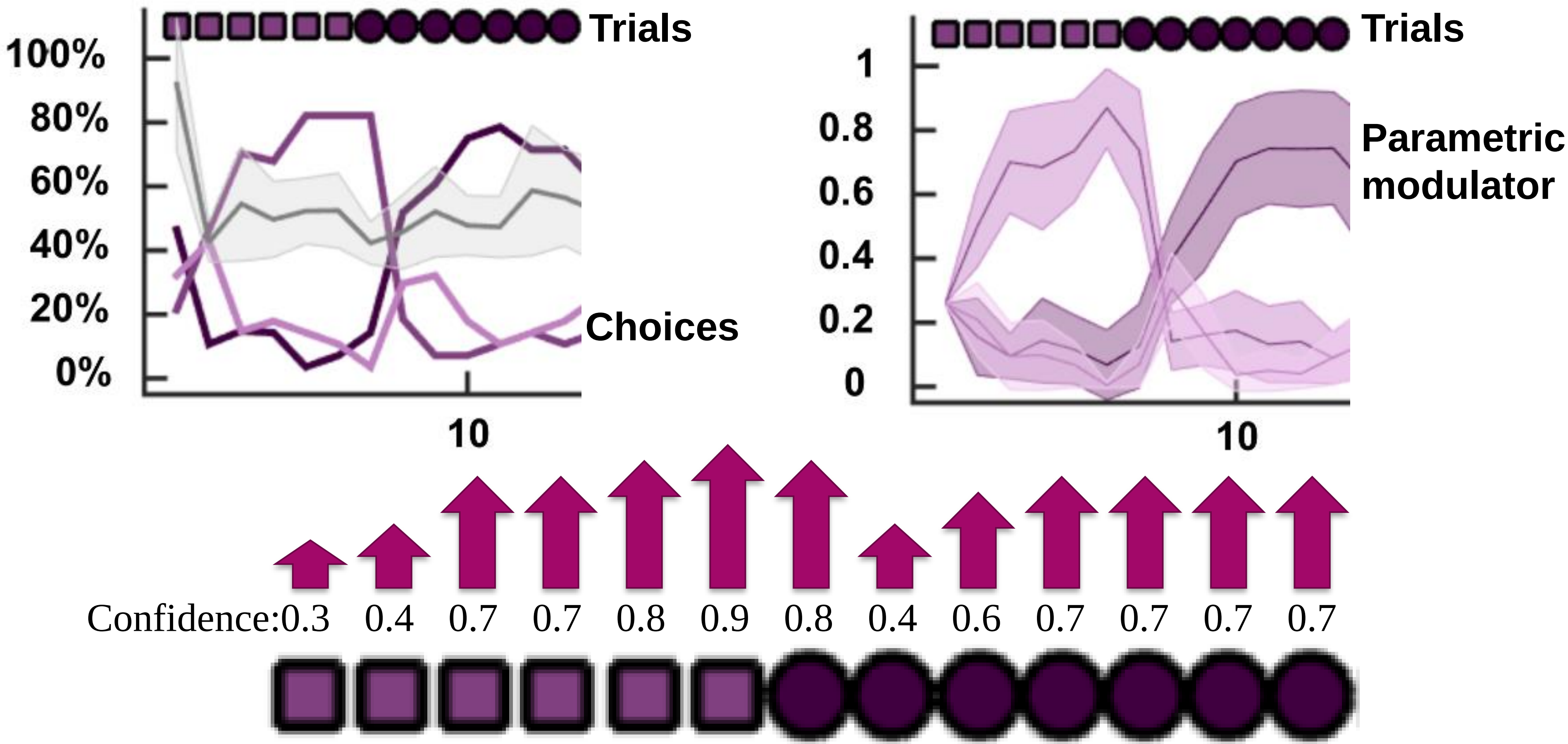


Example: Prediction Error as parametric modulator



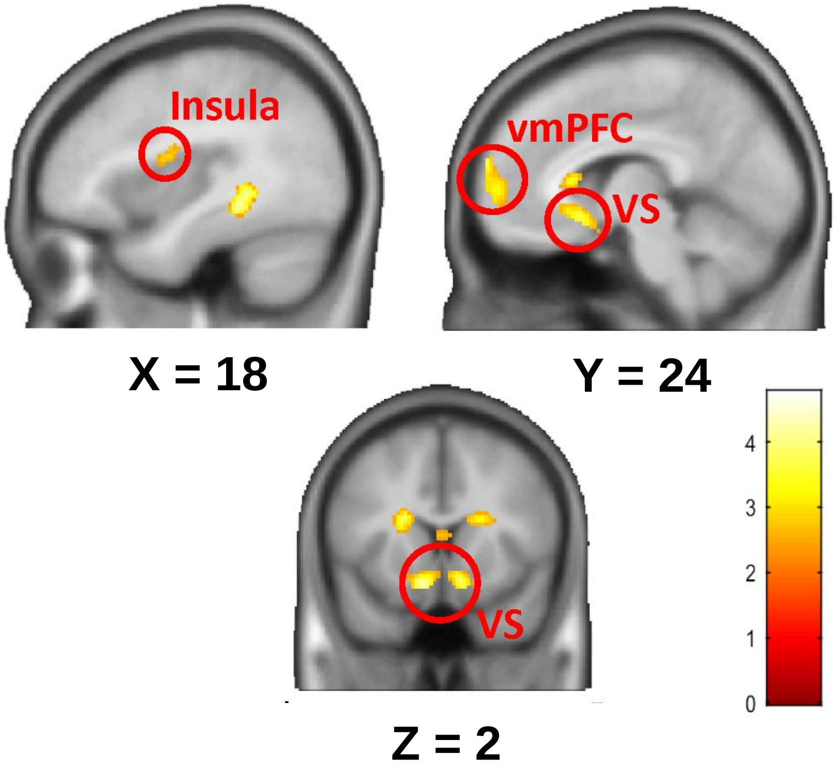


Example: Confidence as parametric modulator

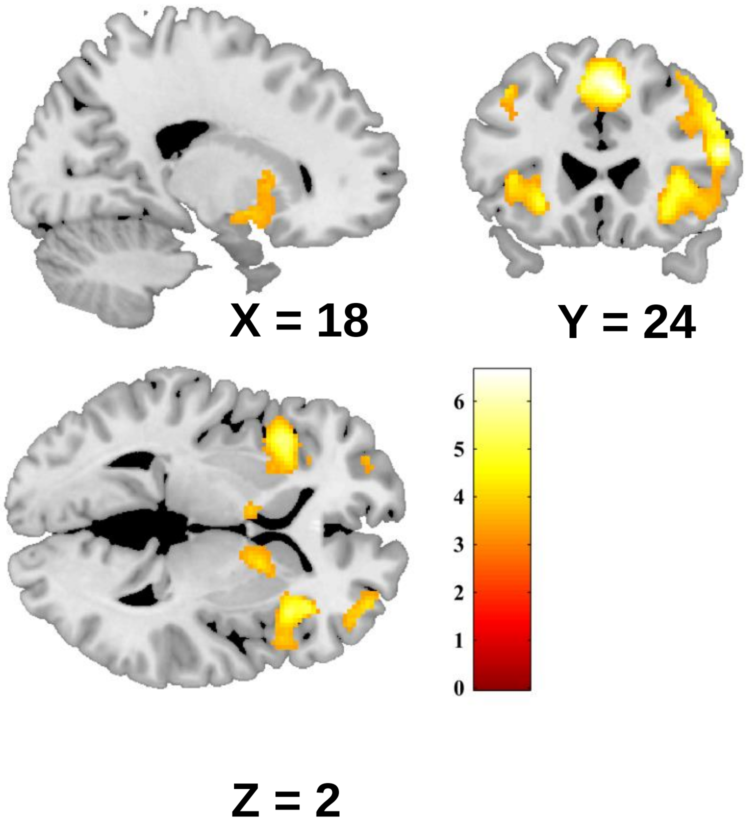


Finder Keeper

Prediction error signal

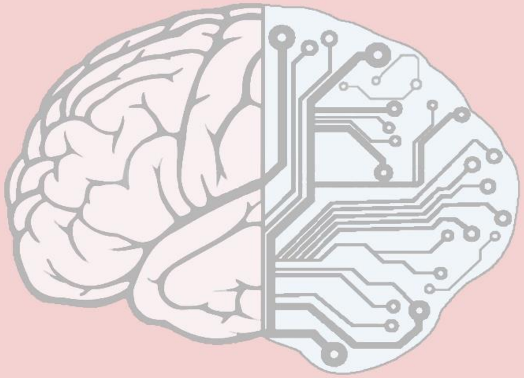


Uncertainty signal



Part III Summary

- Different models address different questions. However, models can systematically fail in replicating a behaviour making any analysis based on them, meaningless.
- Model-based fMRI consists in estimating weights related to cognitive processes, associated (frequently) with a choice behaviour. These weights are then applied to the fMRI signal on a trial-by-trial basis, avoiding reduced sampling and arbitrary trial selections.
- The weights are subject-specific: the model is tuned to replicate the choice selections of each participant in a study, thus (hopefully) estimating the cognitive processes.



2019 Computational Psychiatry Summer (pre-)Course

Introduction to the Bayesian approach

Modelling principles in actions

Model-Based fMRI

Vincenzo G. Fiore, PhD
Mount Sinai School of Medicine

Mount Sinai