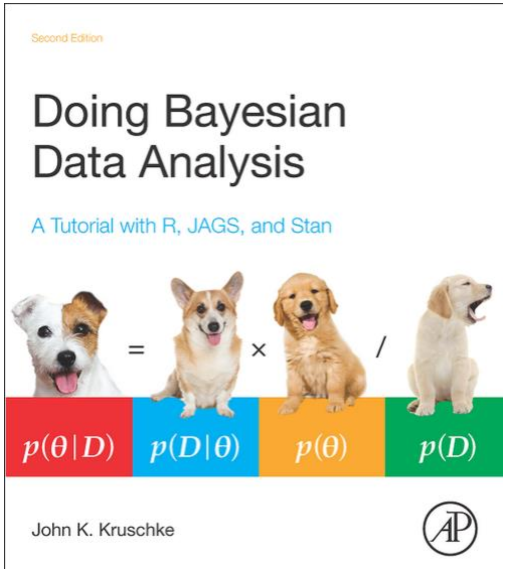
Bayes: an alternative to NHST



Outline

Intro to Bayesian inference

Bayes factor *applied*

- Bayesian *model comparison*

Bayesian sampling *applied*

- Bayesian *estimation* using MCMC

More examples (?)

Bayes rule

Bayesian inference is a re-allocation of credibility

- Examples in R code later..

Likelihood Prior belief

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

Posterior belief

Evidence
⇔
Marginal likelihood

Example

Assume rare disease

- P(θ = ☹) = 0.01
- P(θ = ☺) = 1 – P(θ = ☹) = 0.99

There is a test for the disease

- Test has 99% 'hit rate', 5% false alarm
- P(T = +| θ = ☹) = 0.99
- P(T = +| θ = ☺) = 0.05

What is the probability of having the disease given a positive test??

- i.e. P(θ = ☹|T = +)

Bayes rule *continuous* data

$$P(\theta|D) = \frac{P(D|\theta) \cdot P(\theta)}{P(D)} = \frac{P(D|\theta) \cdot P(\theta)}{\int P(D|\theta_i) \cdot P(\theta_i)\, d\theta_i}$$

Why bother?

- Because integral often cannot be calculated!
- Difference between *exact* and *approximate* solutions

Common approximations:

- Variational Bayes
- Sampling: **Markov Chain Monte Carlo** (**MCMC**)

Why Bayes?

Two views: *falsification* vs. *model evidence*

- *Null Hypothesis Significance Testing* is restricted to rejecting the null
  - “straw-man null“
  - Focus of interest
- Bayes: test for evidence in favour of null
  - Perform model comparison using **Bayes factors**

Bayes factor

Basic idea: compare evidence for null and alternative

$$P(Null) \sim normal(0, \sigma^2)$$
$$P(Alternative) \sim Cauchy()$$

$$BF = \frac{\sum P(D|Alternative) \cdot P(Alternative)}{\sum P(D|Null) \cdot P(Null)}$$

Compare model evidences

More sophisticated:

$$\frac{P(Alternative)}{P(Null)} \quad \text{x} \quad BF \quad = \frac{P(Alternative|Data)}{P(Null|Data)}$$

Prior Odds **Model evidences** **Posterior Odds**

Why Bayes?

Two views: *falsification* vs. *model evidence*

– Bayes: test for evidence in favour of null

Bayes *supersedes* NHST:

Journal of Experimental Psychology: General
2013, Vol. 142, No. 2, 573–603

0096-3445/13/$12.00 DOI: 10.1037/a0029146

Bayesian Estimation Supersedes the *t* Test

John K. Kruschke
Indiana University, Bloomington

Why Bayes?

Bayes *supersedes* NHST:

- NHST has several problems, one of them being that it relies on *sampling intentions*
- Bayes provides more information + credibility!

Journal of Experimental Psychology: General
2013, Vol. 142, No. 2, 573–603

0096-3445/13/$12.00 DOI: 10.1037/a0029146

Bayesian Estimation Supersedes the *t* Test

John K. Kruschke
Indiana University, Bloomington

Sampling intentions - sketch

Assume we test for bias of a coin

- Estimate bias θ
- $H_0 : \theta = 0.5$
- $H_A : \theta \neq 0.5$
  (two-sided test, reject if in left or right 2.5%)

Assume you throw 24 times and there are 7 heads

- Is the coin fair?

We estimate $P(\,{}^{24}_{7\ or\ less}|\theta = 0.5)$

Sampling intentions - sketch

However, there are three ways to acquire the data:

- Throw 24 times, then stop & estimate
- Throw until you have 7 heads
- Throw and stop after 5 minutes

…why should that make a difference?

Sampling intentions - sketch

Throw 24 times, then stop & estimate

⇔ what is the likelihood of getting 7 or less heads when we throw 24 times, given that $\theta = 0.5$?

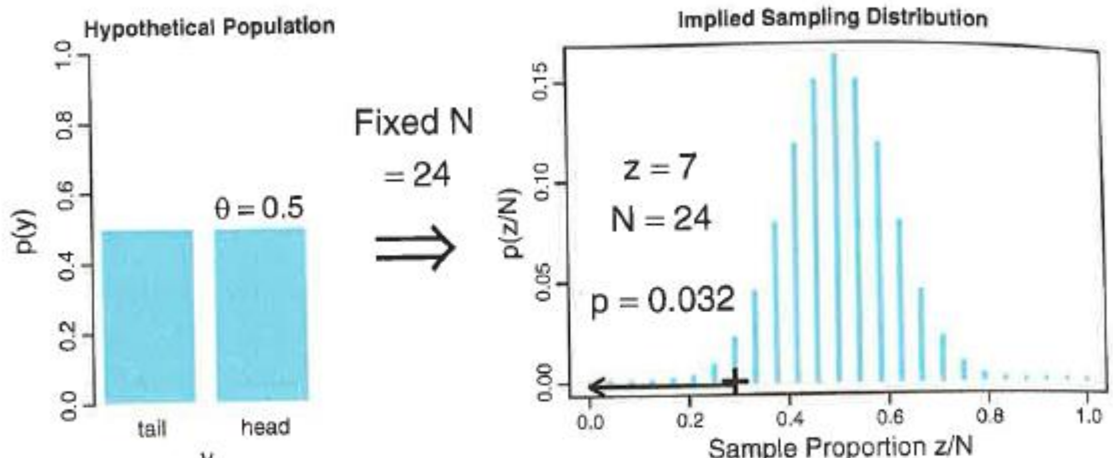


Figure 11.3 The imaginary cloud of possible outcomes when *N* is fixed. The null hypothesis likelihood distribution and parameter are shown on the left. The stopping intention is shown in the middle. The sampling distribution and *p* value are shown on the right. Compare with Figures 11.4 and 11.5.

$$P(z|N,\theta) = \binom{N}{z} \cdot \theta^z \cdot (1-\theta)^{N-z}$$

Sampling intentions - sketch

Throw until you have 7 heads

⇔ what is the probability of needing 24 throws or more until we get 7 heads, given that $\theta = 0.5$?

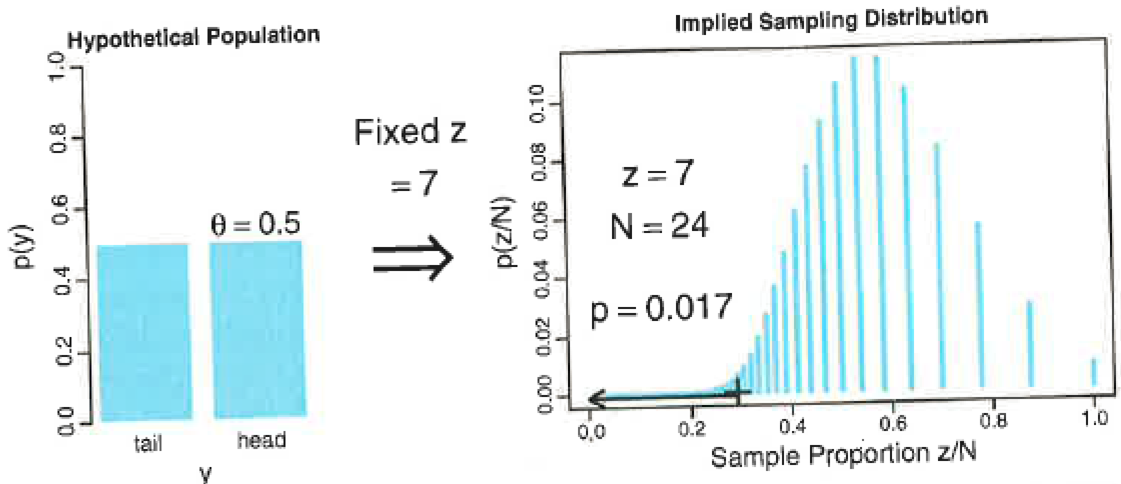


Figure 11.4 The imaginary cloud of possible outcomes when z is fixed. The null hypothesis likelihood distribution and parameter are shown on the left. The stopping intention is shown in the middle. The sampling distribution and p value are shown on the right. Compare with Figures 11.3 and 11.5.

$$P(z|N,\theta) = \binom{N-1}{z-1} \cdot \theta^{z} \cdot (1-\theta)^{N-z}$$

Sampling intentions - sketch

Throw and stop after 5 minutes

⇔ what is the probability of ending up with 24 throws of which 7 are heads, given that $\theta = 0.5$?

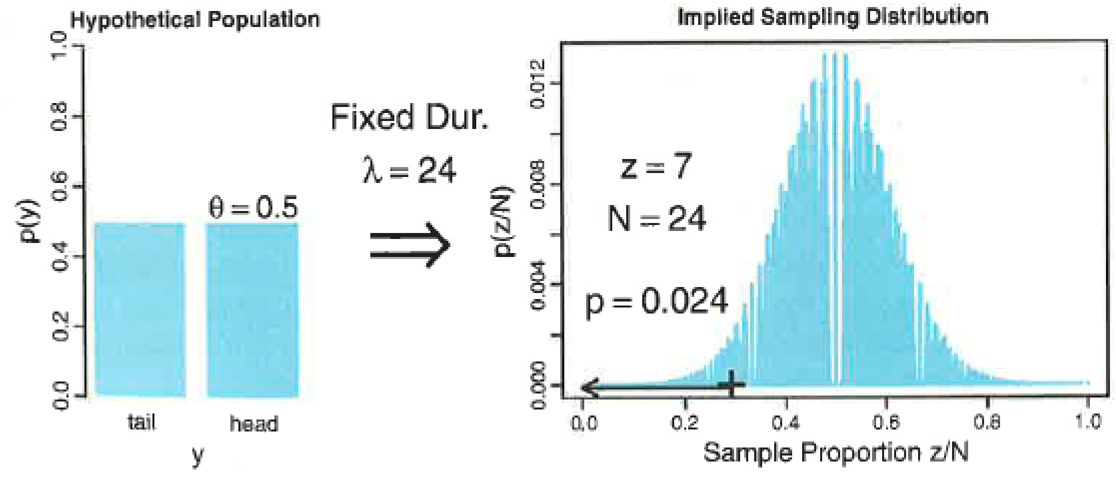


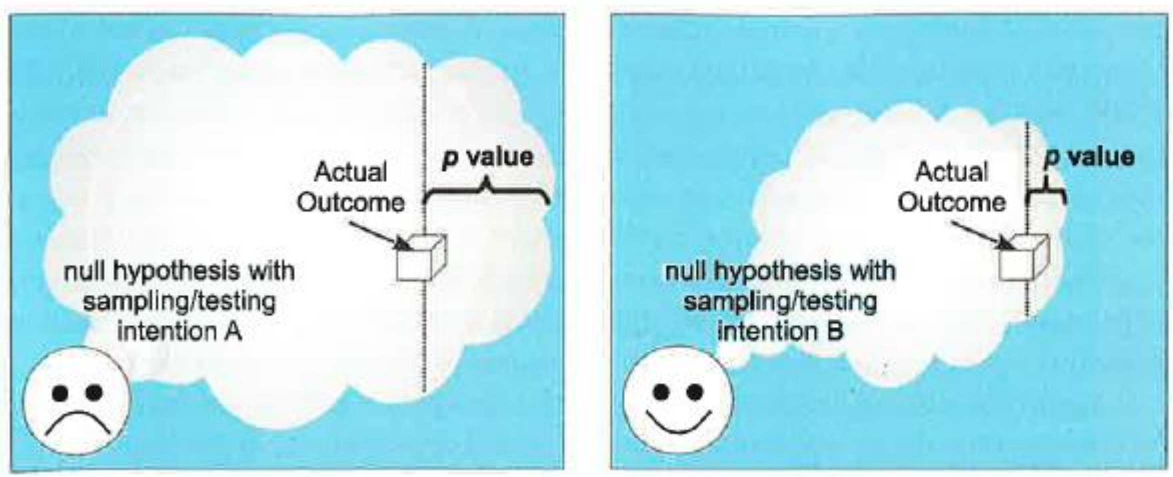
Figure 11.5 The imaginary cloud of possible outcomes when duration is fixed. The null hypothesis likelihood distribution and parameter are shown on the left. The stopping intention is shown is the middle. The sampling distribution and *p* value are shown on the right. Sample sizes are drawn randomly from a Poisson distribution with mean λ. Compare with Figures 11.3 and 11.4.

Sampling intentions - sketch

Thus, we end up with three different p-values *that are all three correct*

Esp. severe if N is small

– Also in other problems bigger differences



Sampling intentions - sketch

This also extends to confidence intervals

⇔ the range of parameter values that would not be rejected.

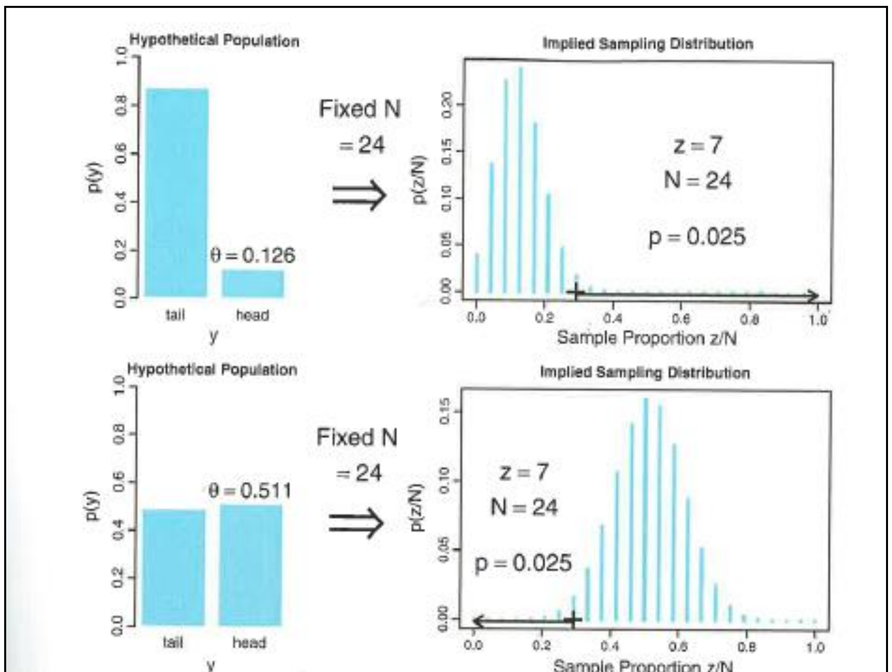


Figure 11.8 95% confidence interval when N is fixed by the experimenter's intention extends from $\theta = 0.126$ (top row) to $\theta = 0.511$ (bottom row). Compare with Figures 11.9 and 11.10.

vs.

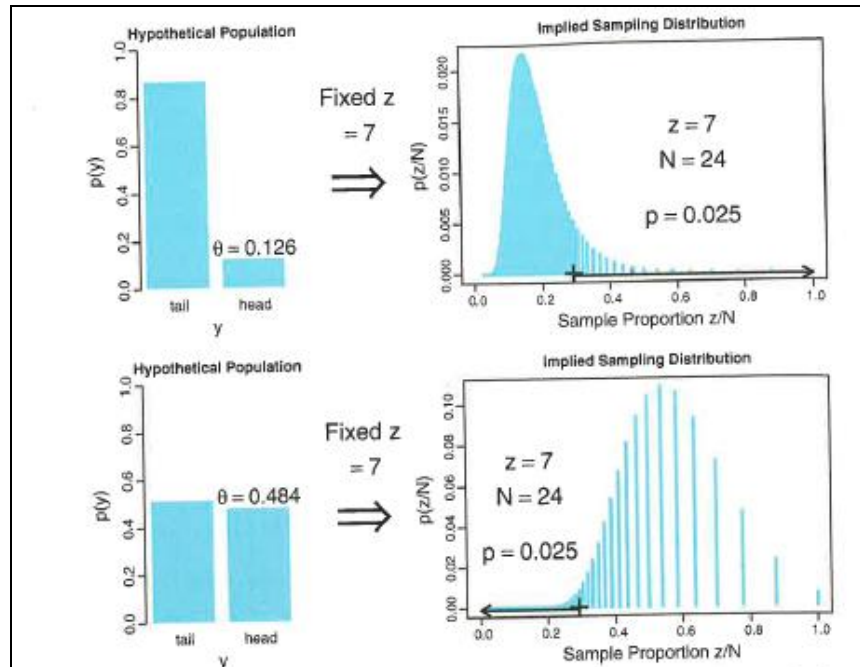


Figure 11.9 95% confidence interval when z is fixed by the experimenter's intention extends from $\theta = 0.126$ (top row) to $\theta = 0.484$ (bottom row). Compare with Figures 11.8 and 11.10.

Sampling intentions - summary

P-value depends on how we define the hypothesis space

- Severity depends on problem and sample size!
- Equally affects confidence intervals!

Furthermore, confidence intervals provide no information about credibility

- Unlike HDI in Bayes estimation

Lastly, what happens if we have a strong prior expectation about a bias, such as flipping a nail

- Throwing 7 heads out of 24 flips and not rejecting hypothesis that nail is fair??

Why Bayes?

Bayes *supersedes* NHST:

- Bayes provides more information + credibility!

1st approach: Bayes factor

- Compare models

2nd (better) approach: parameter estimation

- Define **Region of practical equivalence** (*ROPE*)
  - Small range of values considered practical equivalent to null value
- Define **Highest density interval** (*HDI*)
  - 95% of most credible points/parameter values

Why Bayes?

2nd (better) approach: parameter estimation

- Define **Region of practical equivalence** (*ROPE*)
- Define **Highest density interval** (*HDI*)

Decision-rule

- If ROPE contains complete HDI, then we accept null
- If HDI lies outside ROPE, we accept alternative
- If intersect, no conclusive evidence

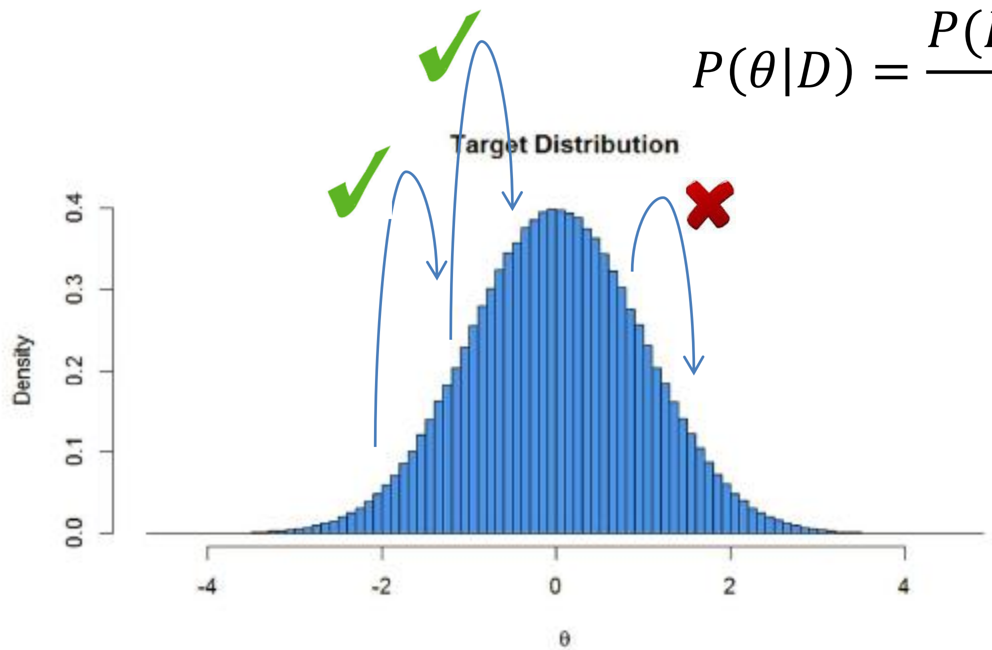
Bayesian estimation

How do we get HDIs for certain parameters?

- Via **Markov-Chain Monte-Carlo** ($MCMC$) sampling

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

MCMC?



Adapted from Benjamin Scheibehenne

Bayesian estimation: compare 2 groups

Basic idea: estimate 5 parameters

– Mean 1
– Mean 2
– Sd 1
– Sd 2
– ‘normality parameter’

Mean 1, Mean 2: Use to compute (HDI on) **mean difference**

Sd 1, Sd 2: Use to compute (HDI on) difference of **standard deviation**

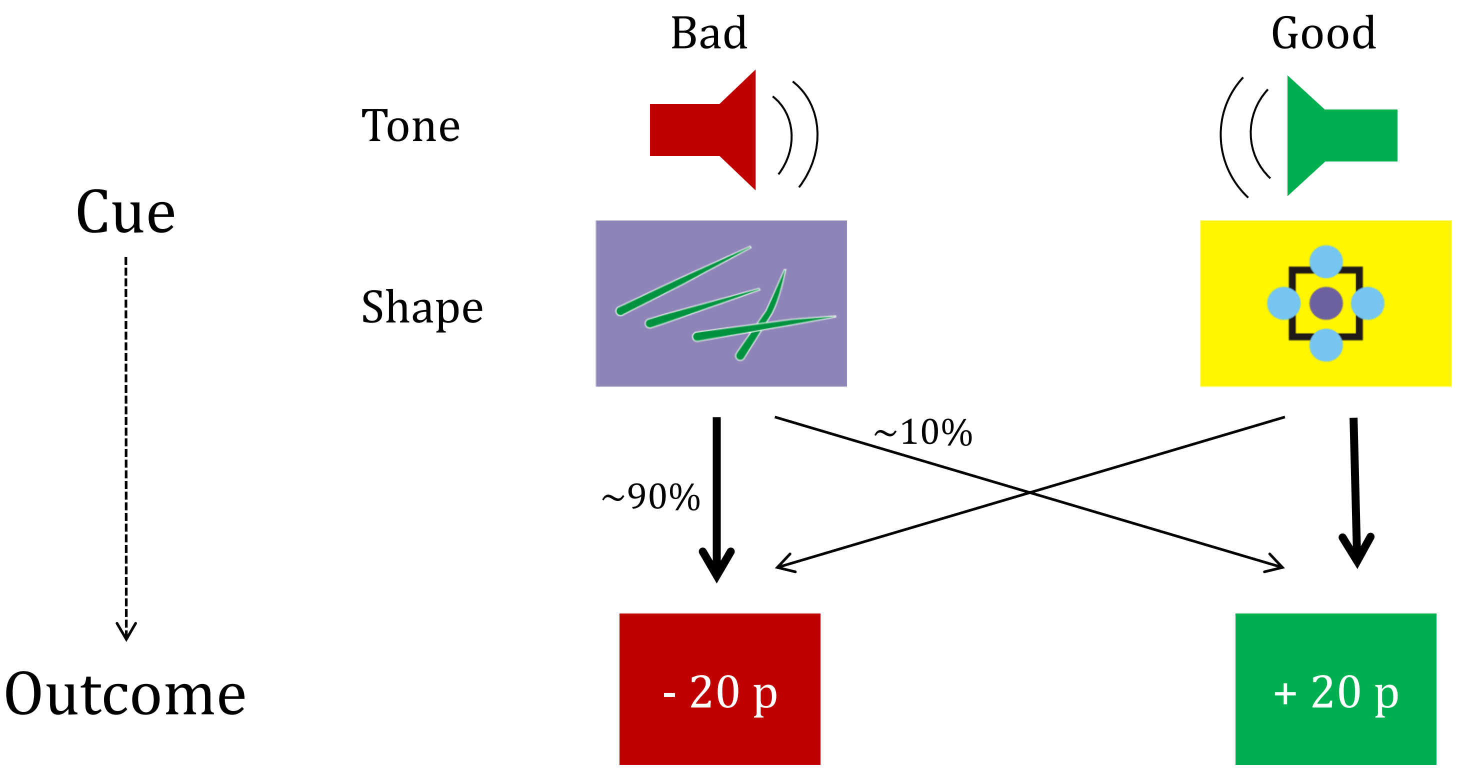
Mean 1, Mean 2, Sd 1, Sd 2: Use to compute **effect size**

‘normality parameter’: Distribution of data, esp. account for **outliers**

+ ‘posterior predictive check’

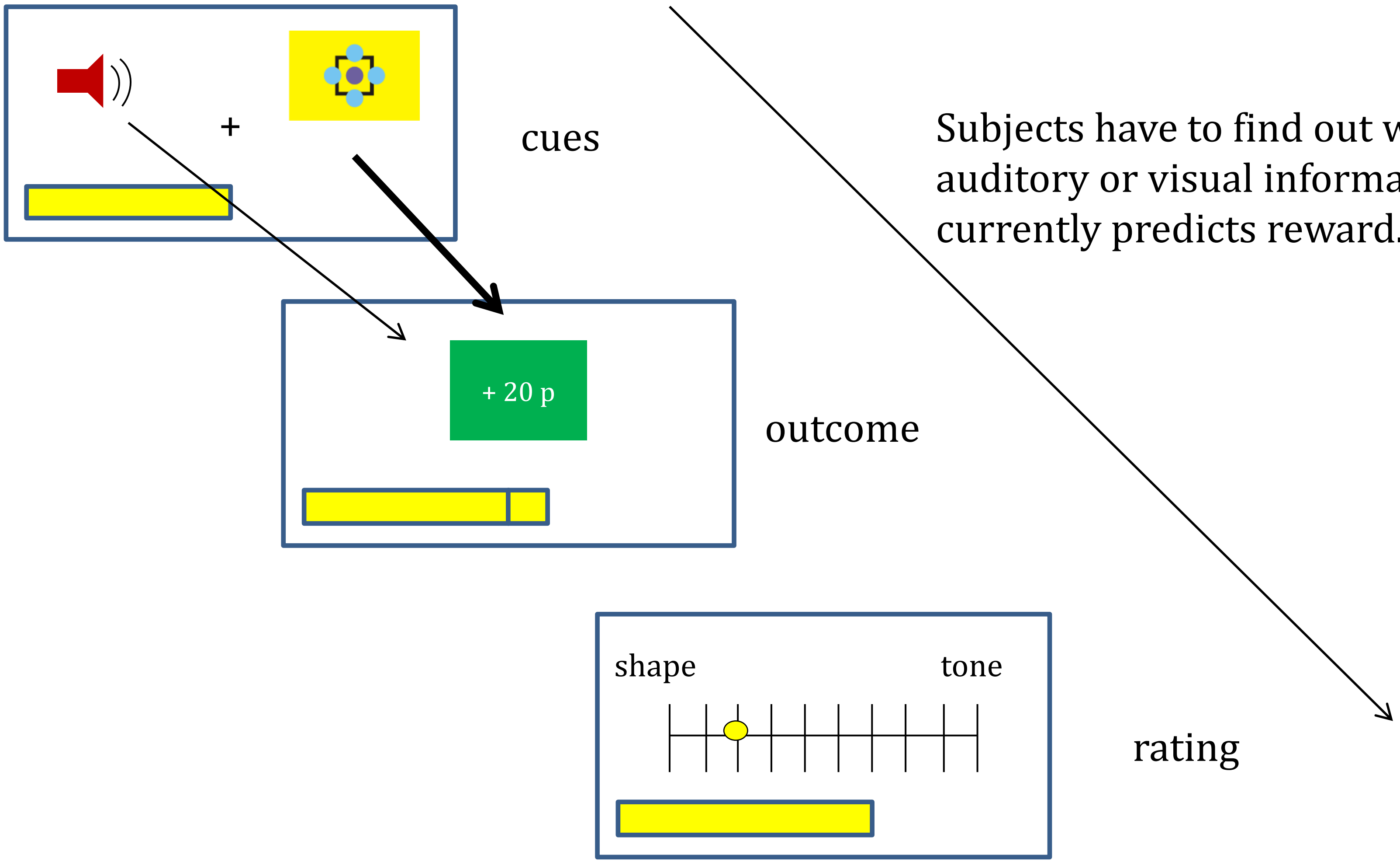
Different perspective: Bayesian modelling of belief updates

Assigning relevance to cues



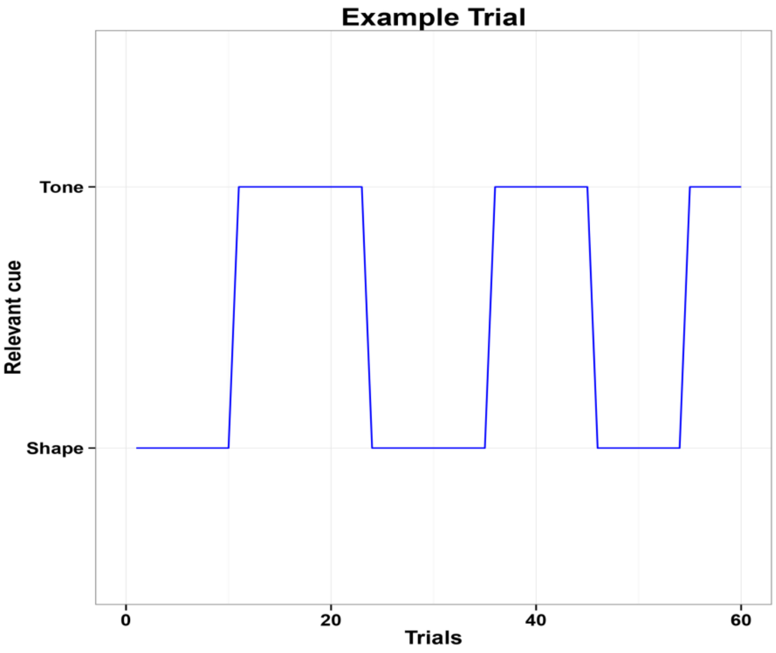
Contingencies stay the same!

Task

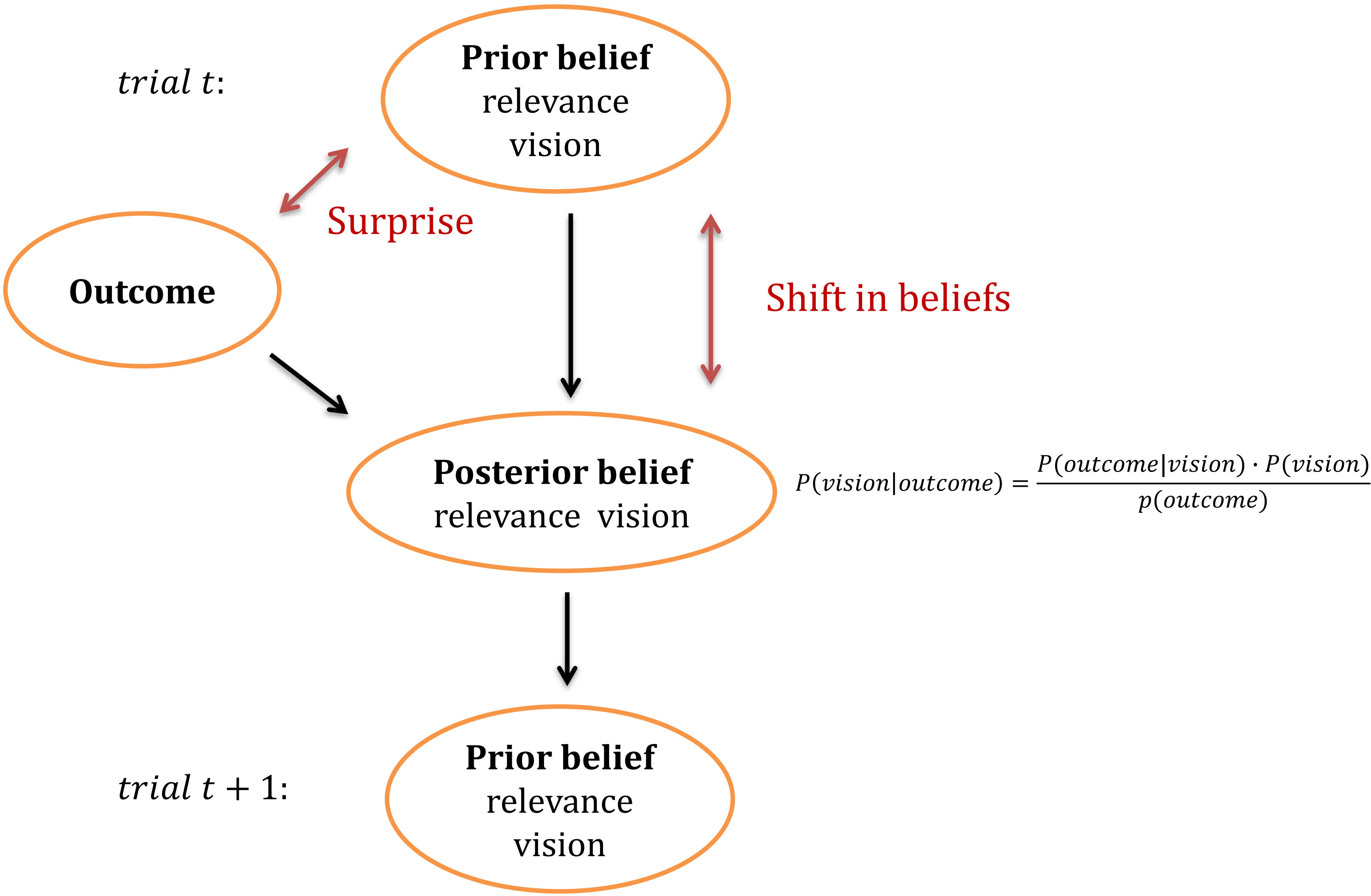


Subjects have to find out whether auditory or visual information currently predicts reward.

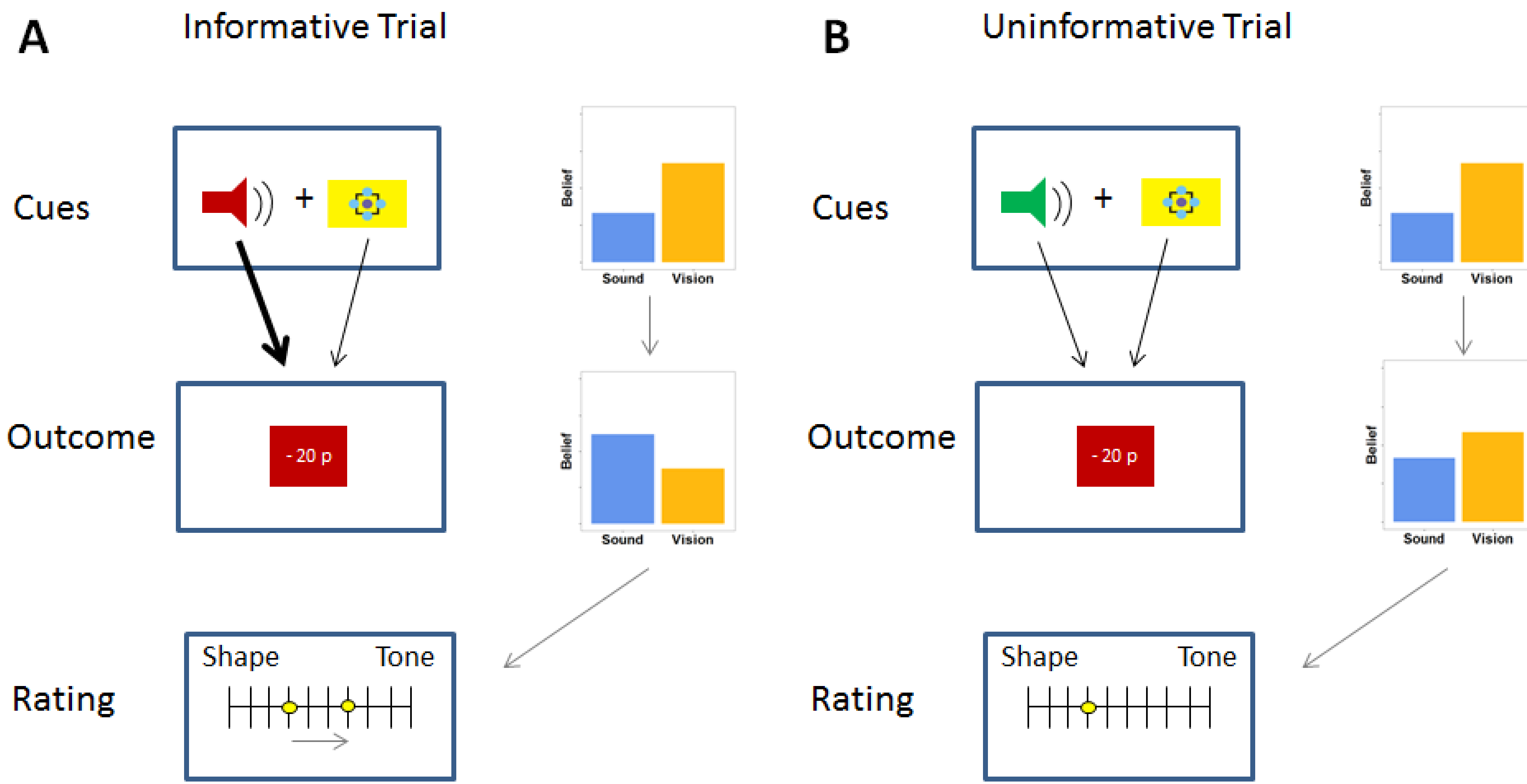
Task: changing relevance



Updates of beliefs



Information

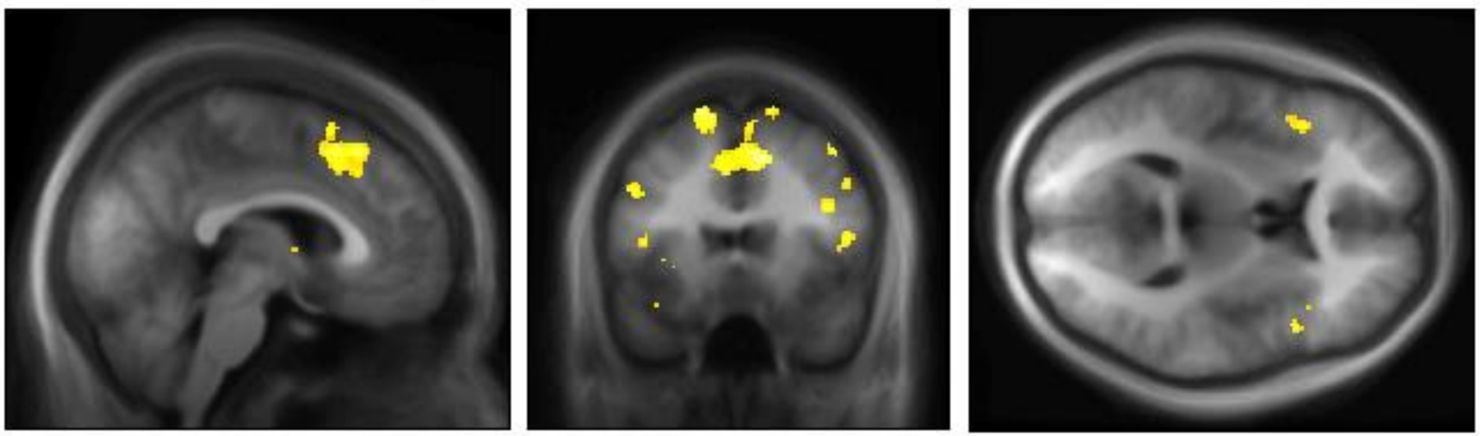


Shift in beliefs

Surprising, but no shift

Results: Surprise

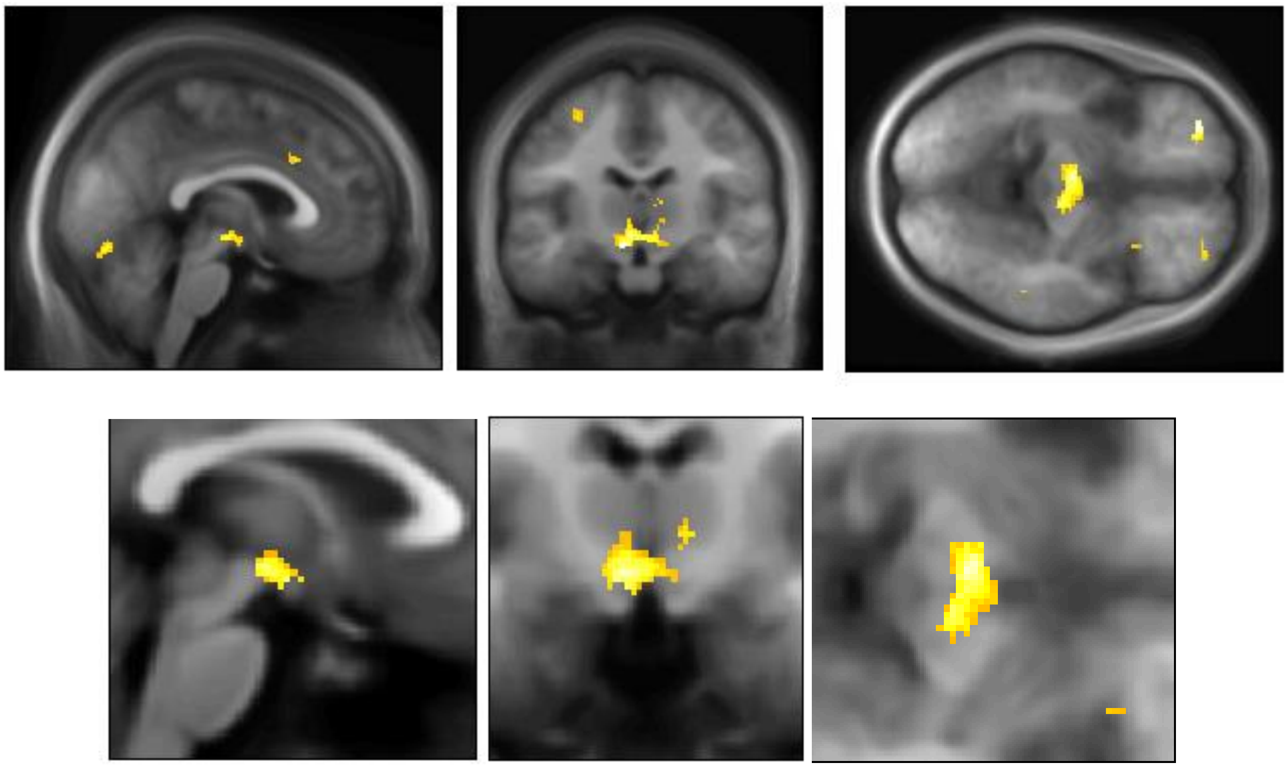
- How unexpected is outcome given expectations?



Effects in **Cingulate Cortex** (peak MNI 10 15 44, $T_{peak} = 4.35$, $p_{cluster} < 0.01$), and **bilateral Insula**

Results: Shifts in beliefs

- Where are shifts in beliefs encoded?



Effects in **Substantia Nigra/Ventral Tegmental Area** (small-volume corrected, peak MNI -8 -16 -10,$T_{peak} = 5.17$,$p_{peak} = 0.002$),