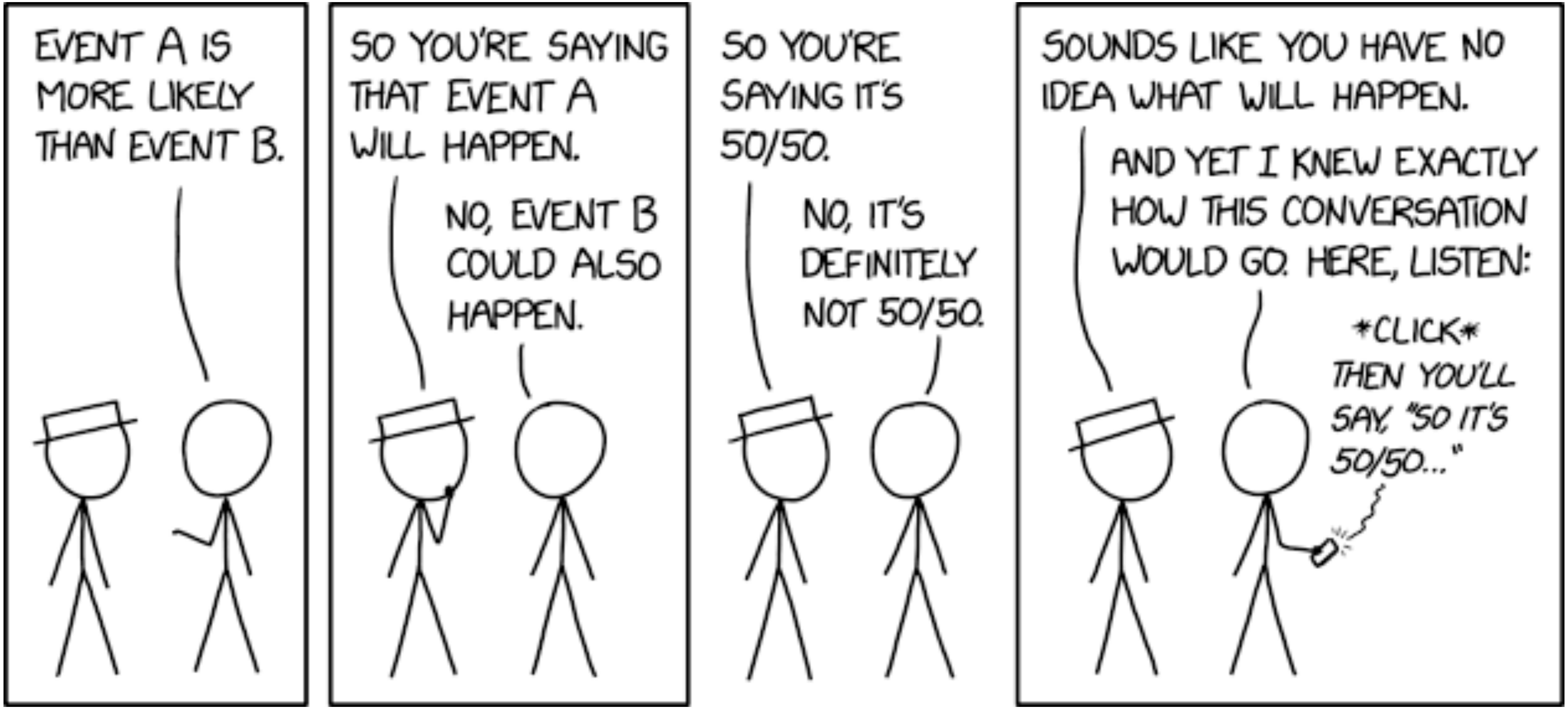


download these slides from: github.com/qtabs/compneuro/04_pc.{key, pdf}

Questions from last session

Q3.1 (Miguel)

According to the Bayesian Brain Hypothesis, our only access to reality is via our observations y. Can we be sure there's a reality, and that that reality is common to everyone else, without observing it?

Q3.2 (Francesca)

Calibration is a measure of the match between the uncertainty of a belief and the actual variability of the events. How could miscalibration affect perception? Find some situations in which humans are often miscalibrated?

Q3.3 (Izei)

We use Gaussian distributions a lot because of two reasons: 1) they are easy to operate with; 2) they are maximum entropy distributions for continuous variables. Find out what (2) means and why that's a desirable property. What would be the maximum entropy distribution for a discrete variable? What's common between that distribution and the Gaussian distribution?

Q3.4 (Anastasiaa)

If different individuals have different priors, whose perception is "correct"? Does the Bayesian framework imply that perception might be constructed? What would then be the role of our culture and upbringing in our perception?

Q3.5 (Shiru)

Where do our priors come from? Evolution? Development? Culture? Can we possibly determine what is the source of our priors?

Q3.6 (Sebastian)

Some people model stereotyped beliefs as reasonings based on the wrong assumption that $P(A|B) = P(B|A)$. Try to use this model to explain some stereotypes. What are the limits for the model? Why would humans, if the Bayesian Brain Hypothesis is correct, often make that mistake (hint: start by considering when does $P(A|B) = P(B|A)$)?

Representational learning and Predictive Coding

Computational Neuroscience - Lecture 4

Alejandro Tabas

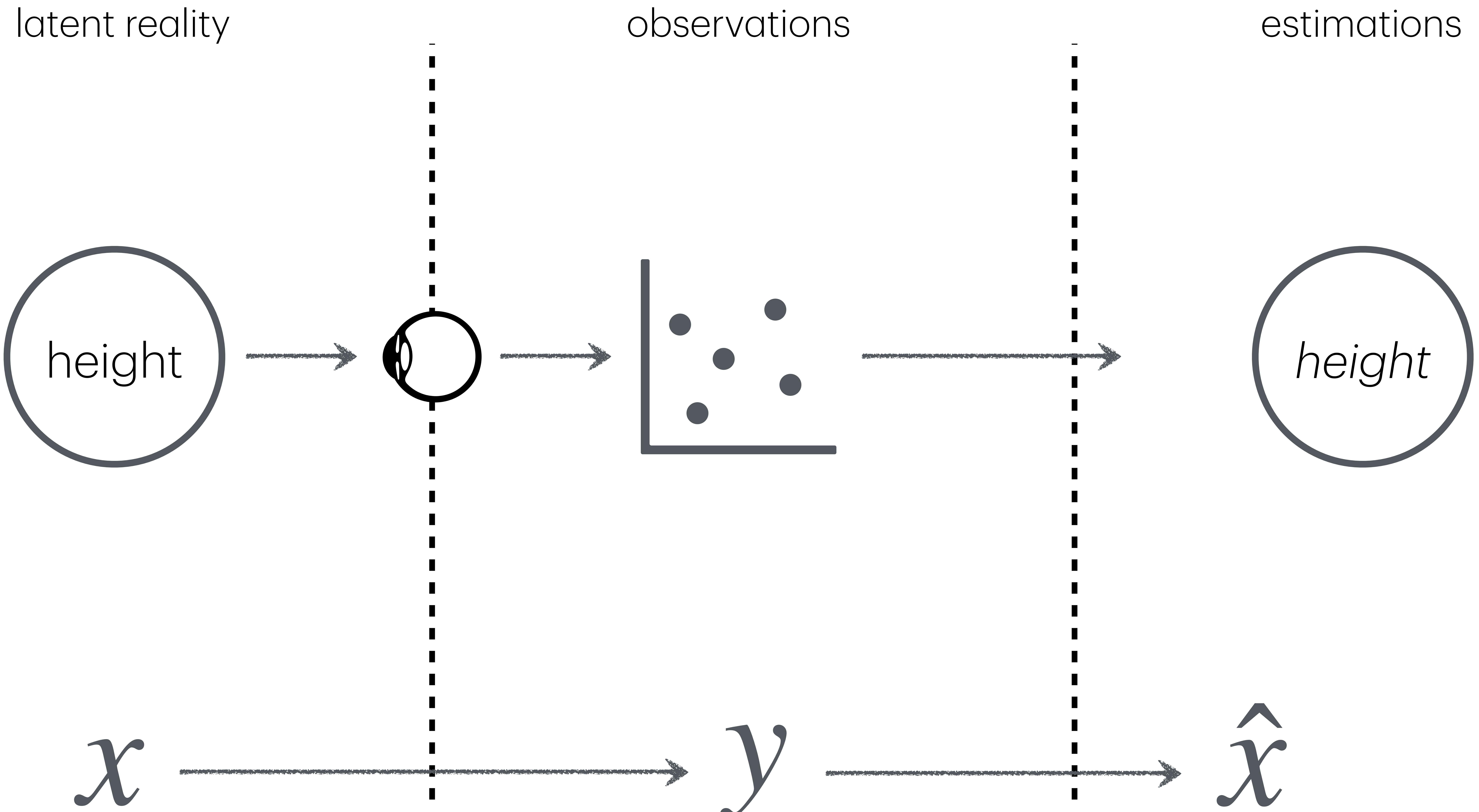
Representational learning and Predictive Coding

1. Representational learning
2. Predictive coding as representational learning
3. Perceptual inference
4. Predictive coding as variational perceptual inference
5. Empirical evidence for and controversies of predictive coding
6. The sampling hypothesis
7. Conclusion

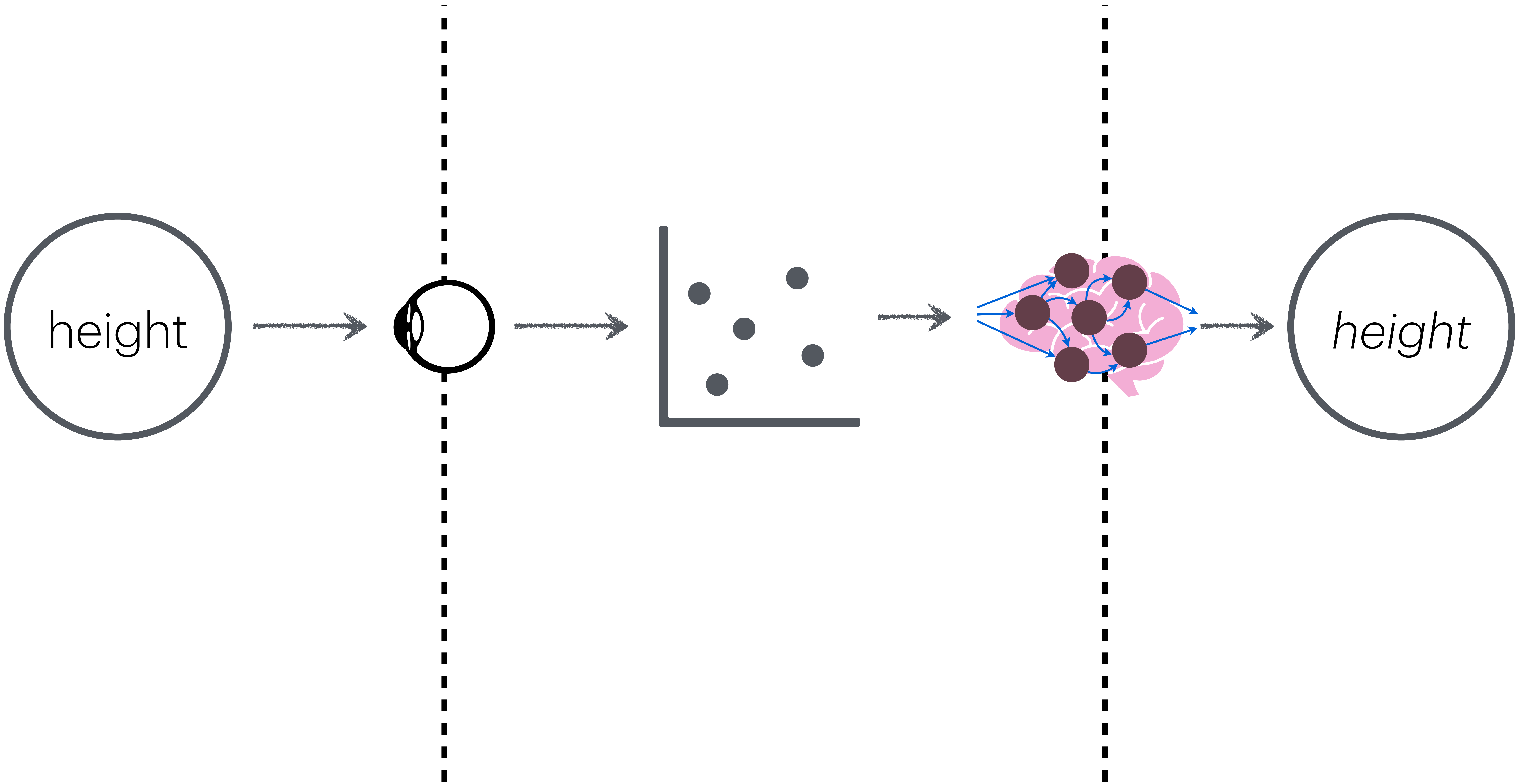
Representational learning and Predictive Coding

1. **Representational learning**
2. Predictive coding as representational learning
3. Perceptual inference
4. Predictive coding as variational perceptual inference
5. Empirical evidence for and controversies of predictive coding
6. The sampling hypothesis
7. Conclusion

Perception = estimate hidden causes

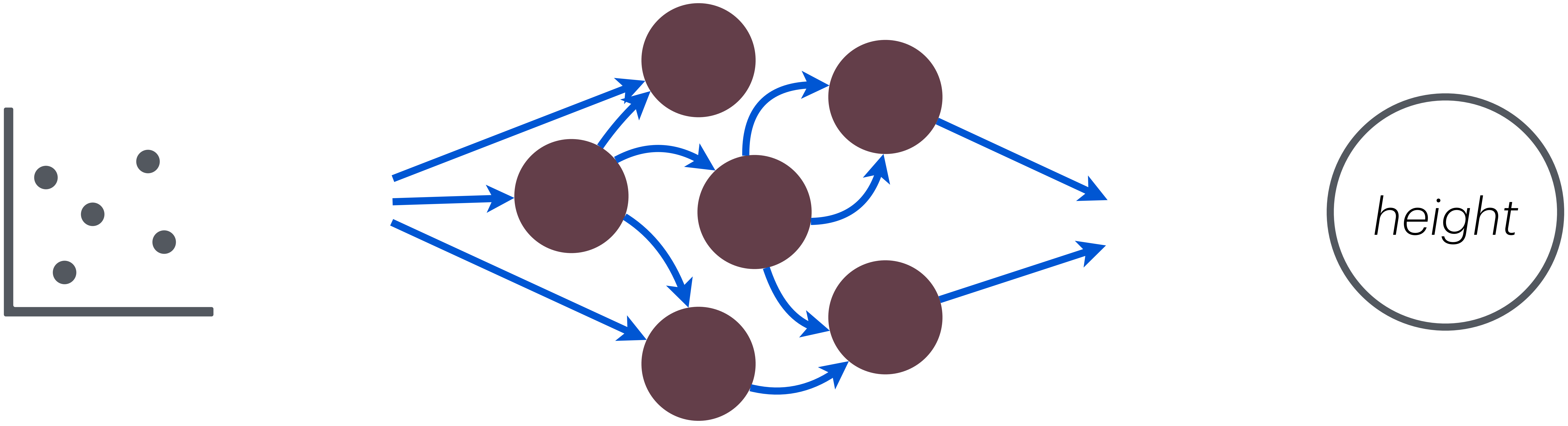


How can the brain learn this mapping?



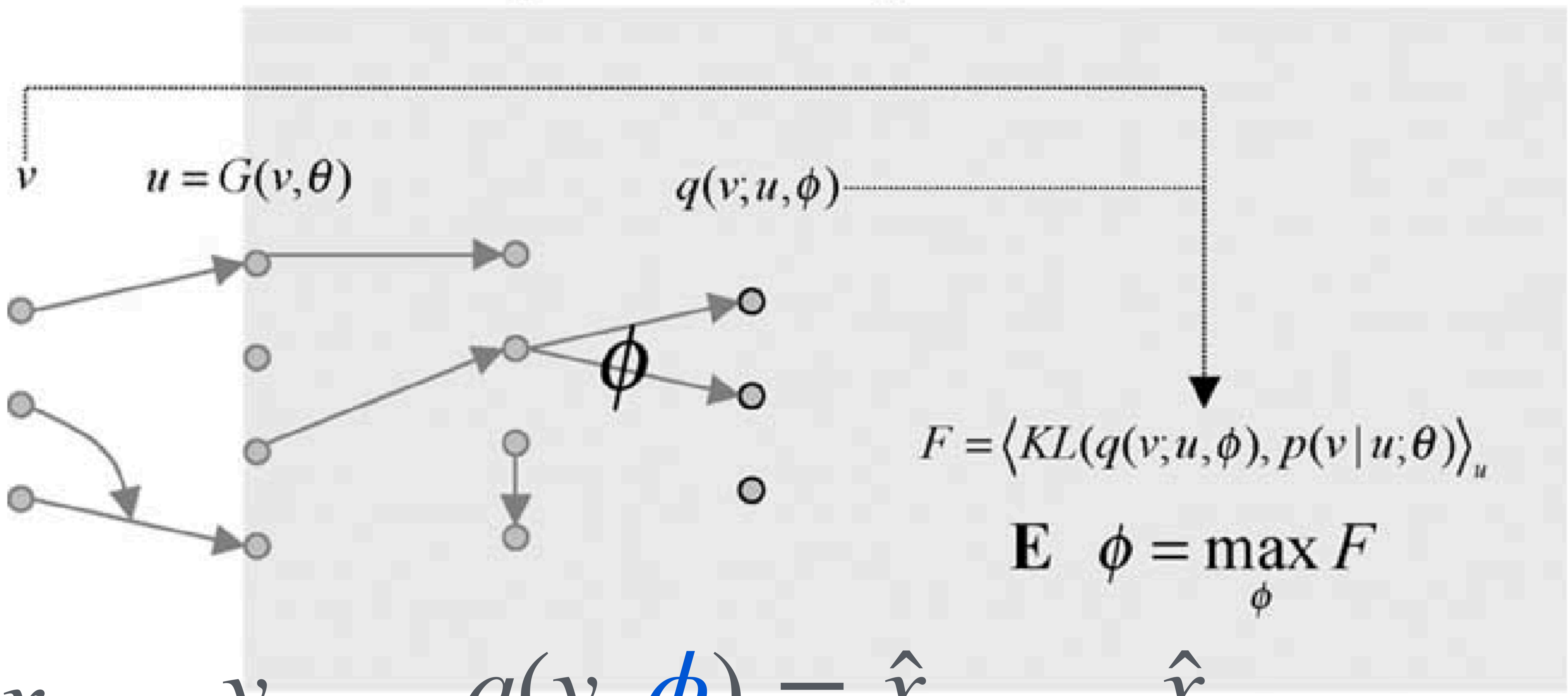
Supervised learning

$$\phi = \operatorname{argmin}_{\phi} \left(x - q(y, \phi)\right)^2$$



y $\quad$ $q(y, \phi) = \hat{x}$ $\quad$ $\hat{x}$

Supervised learning



$x \qquad y \qquad q(y, \phi) = \hat{x} \qquad \hat{x}$

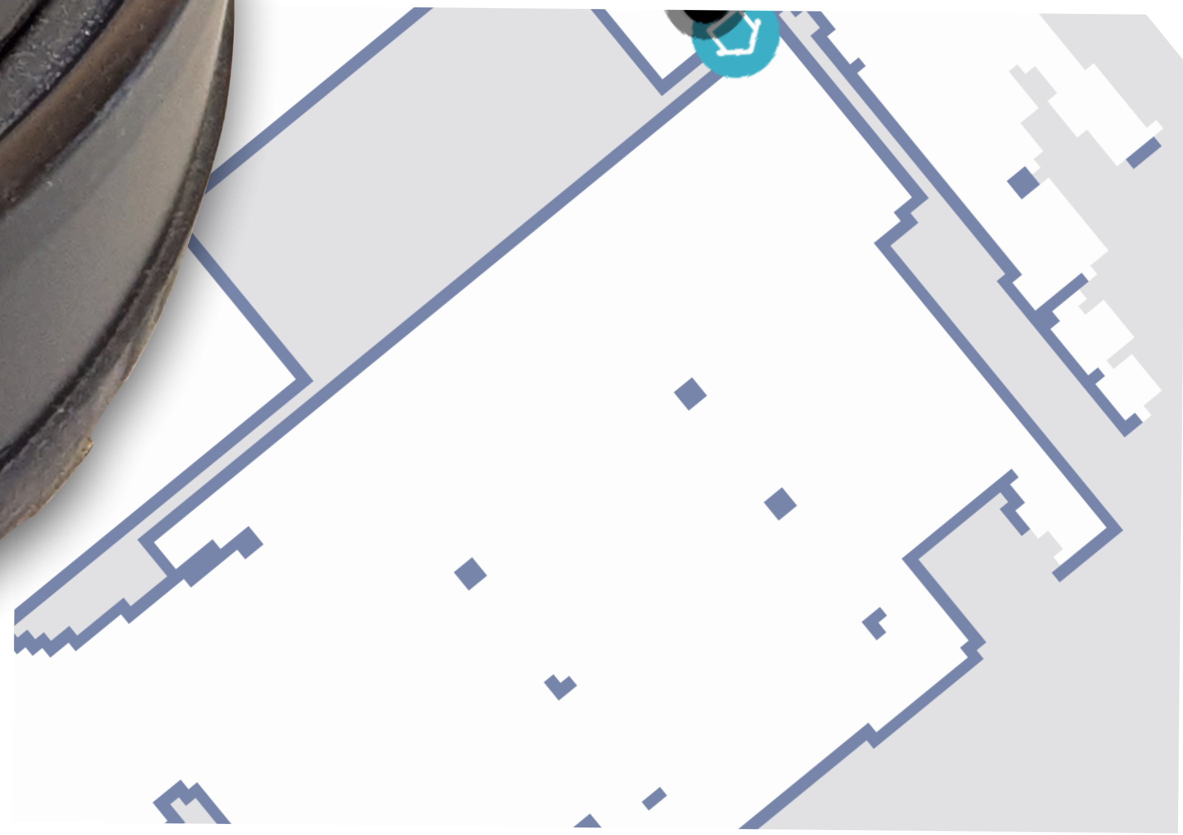
Reality is generally not accessible

latent reality

estimations

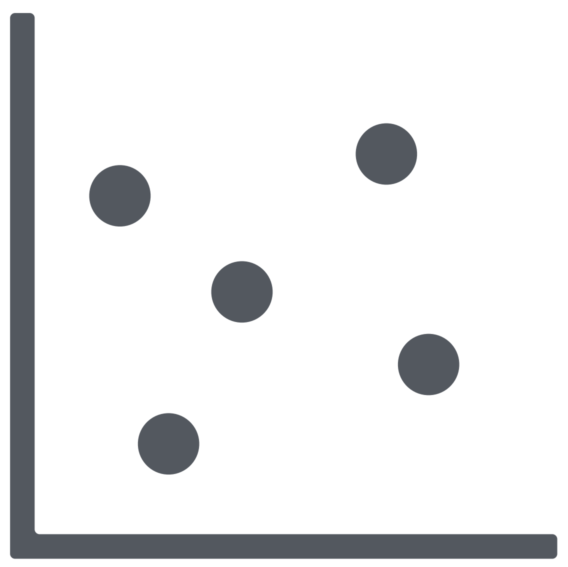
?

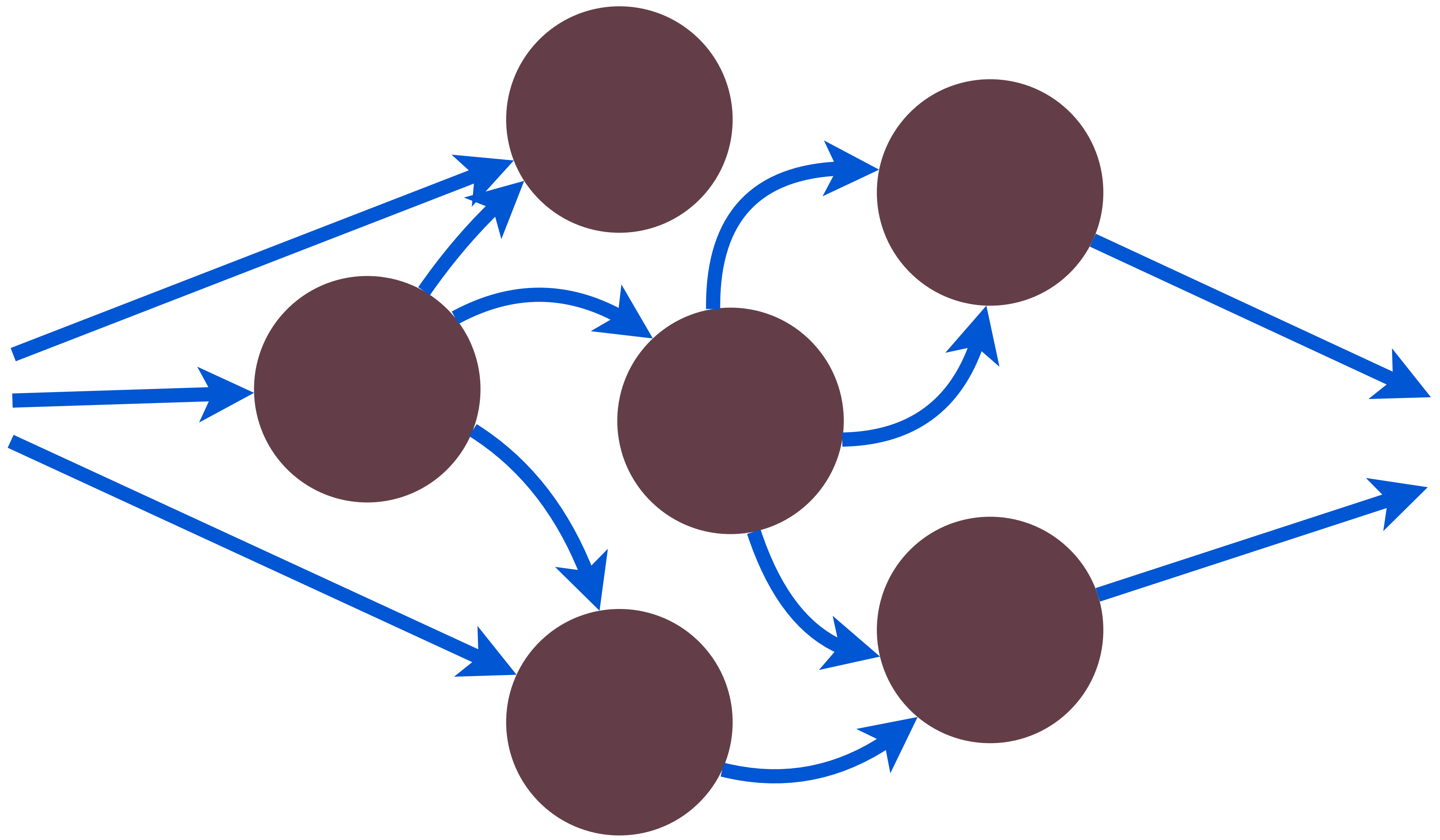
?



$$x \longrightarrow y \longrightarrow P(x \mid y, I)$$

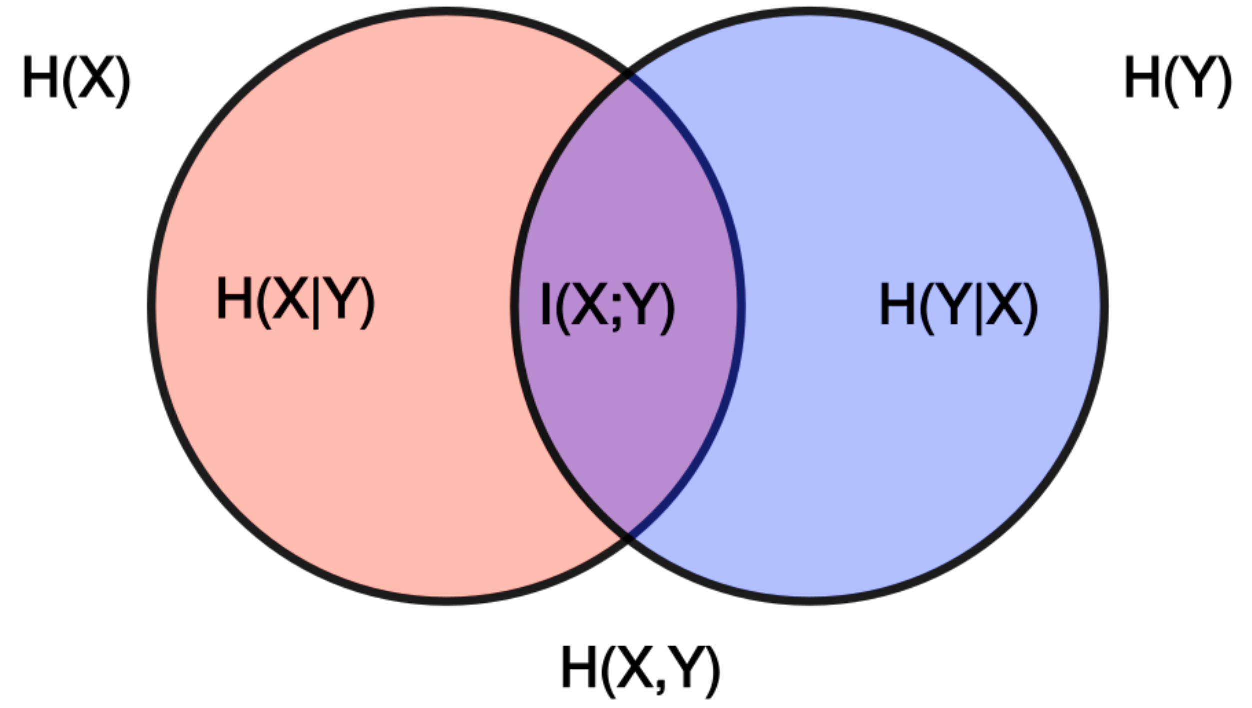
Self-Supervised learning





y $\qquad q(y, \phi) = \hat{x}$ $\qquad \hat{x}$

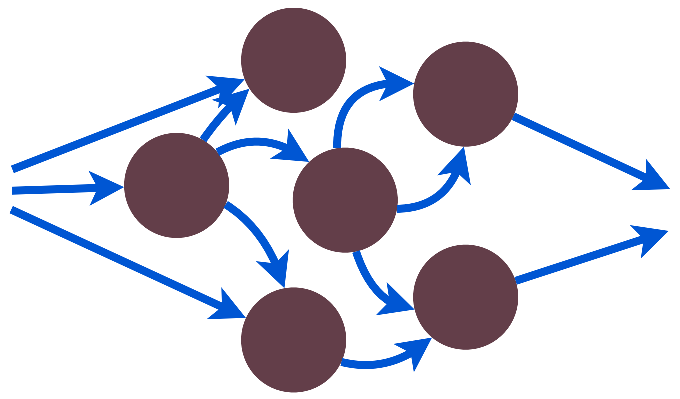
Self-supervised learning: infomax



Problem: x **is generally not observable**

Solution: Generate a representation $\hat{x}$ that

- preserves as much information of y as possible
- is as efficient as possible



maximise $I(y, \hat{x}) = I\left(y, q(y, \phi)\right)$

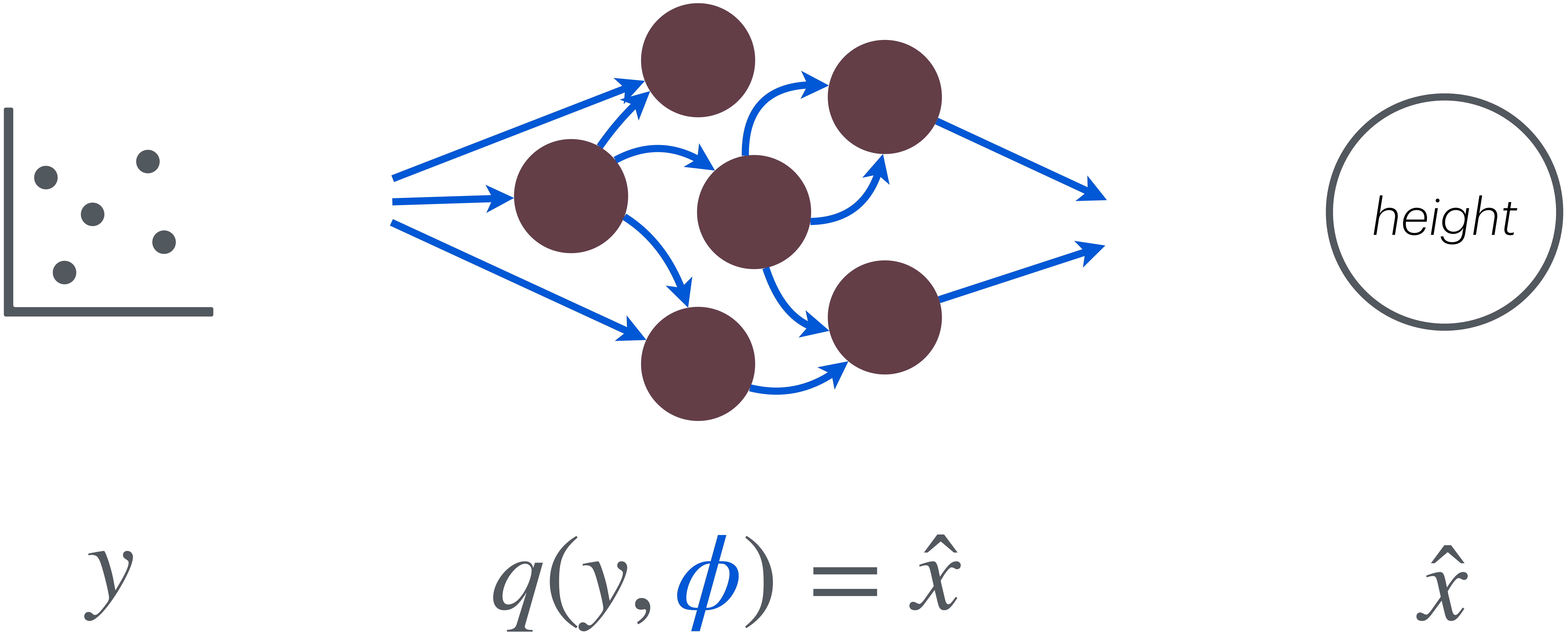
minimise $I\left(q_i(y, \phi), q_j(y, \phi)\right)$

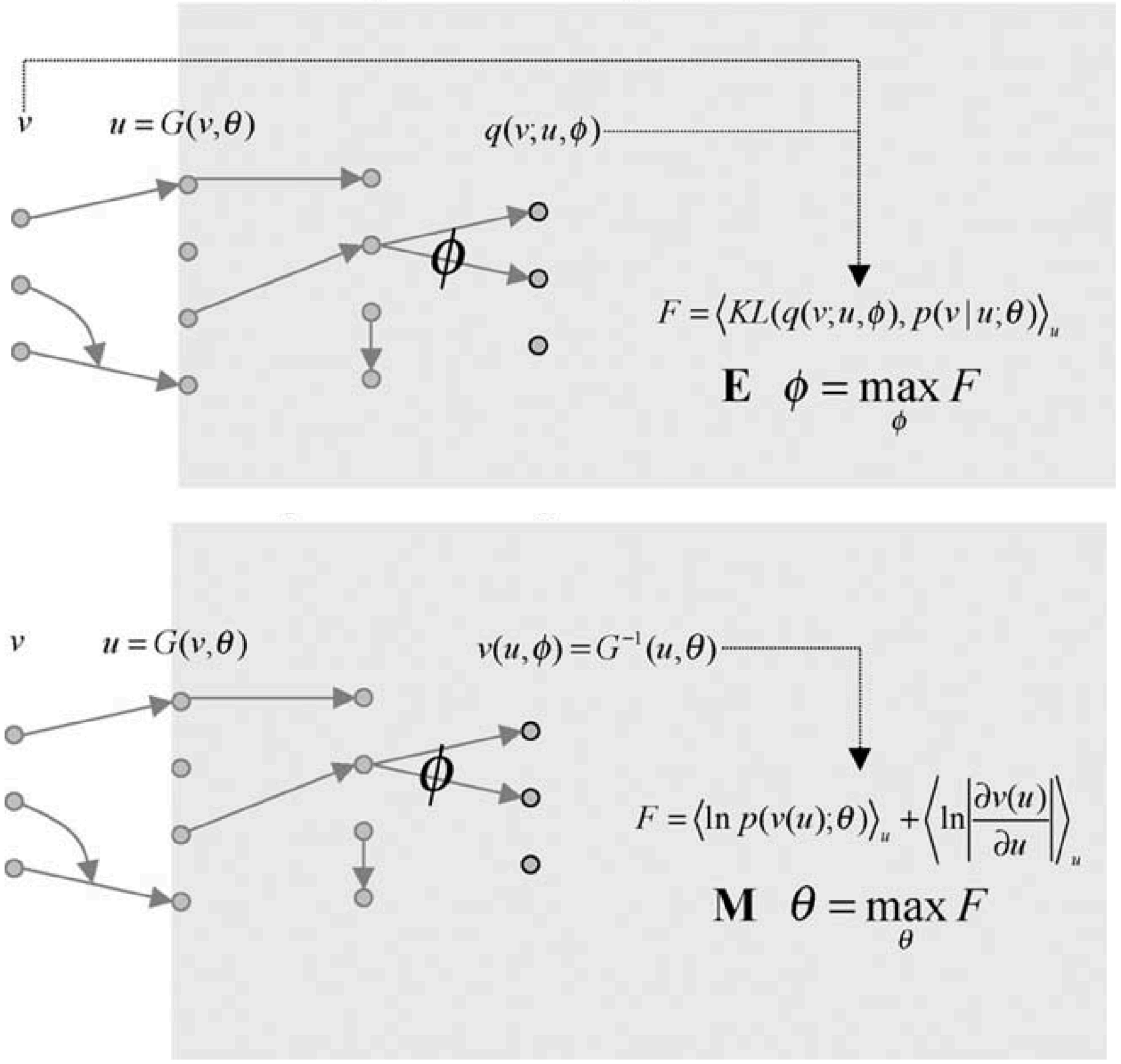
$\phi = \text{argmax}_\phi \mathcal{F}$ $\qquad$ $\mathcal{F} = -I\left(q_i(y, \phi), q_j(y, \phi)\right) + I\left(y, q(y, \phi)\right)$

Self-supervised learning: infomax

$\phi = \text{argmax}_{\phi} \mathcal{F}$

$$\mathcal{F} = -I\Big(q_i(y,\phi), q_j(y,\phi)\Big) + I\big(y, q(y,\phi)\big)$$





v
$u = G(v,\theta)$
$q(v;u,\phi)$
ϕ
$F = \langle KL(q(v;u,\phi), p(v \mid u;\theta) \rangle_u$
E $\phi = \max_{\phi} F$
v
$u = G(v,\theta)$
$v(u,\phi) = G^{-1}(u,\theta)$
ϕ
$F = \langle \ln p(v(u);\theta) \rangle_u + \left\langle \ln \left| \frac{\partial v(u)}{\partial u} \right| \right\rangle_u$
M $\theta = \max_{\theta} F$

Both models are very limited in their scope

Supervised learning:

- Requires having access to the ground truth

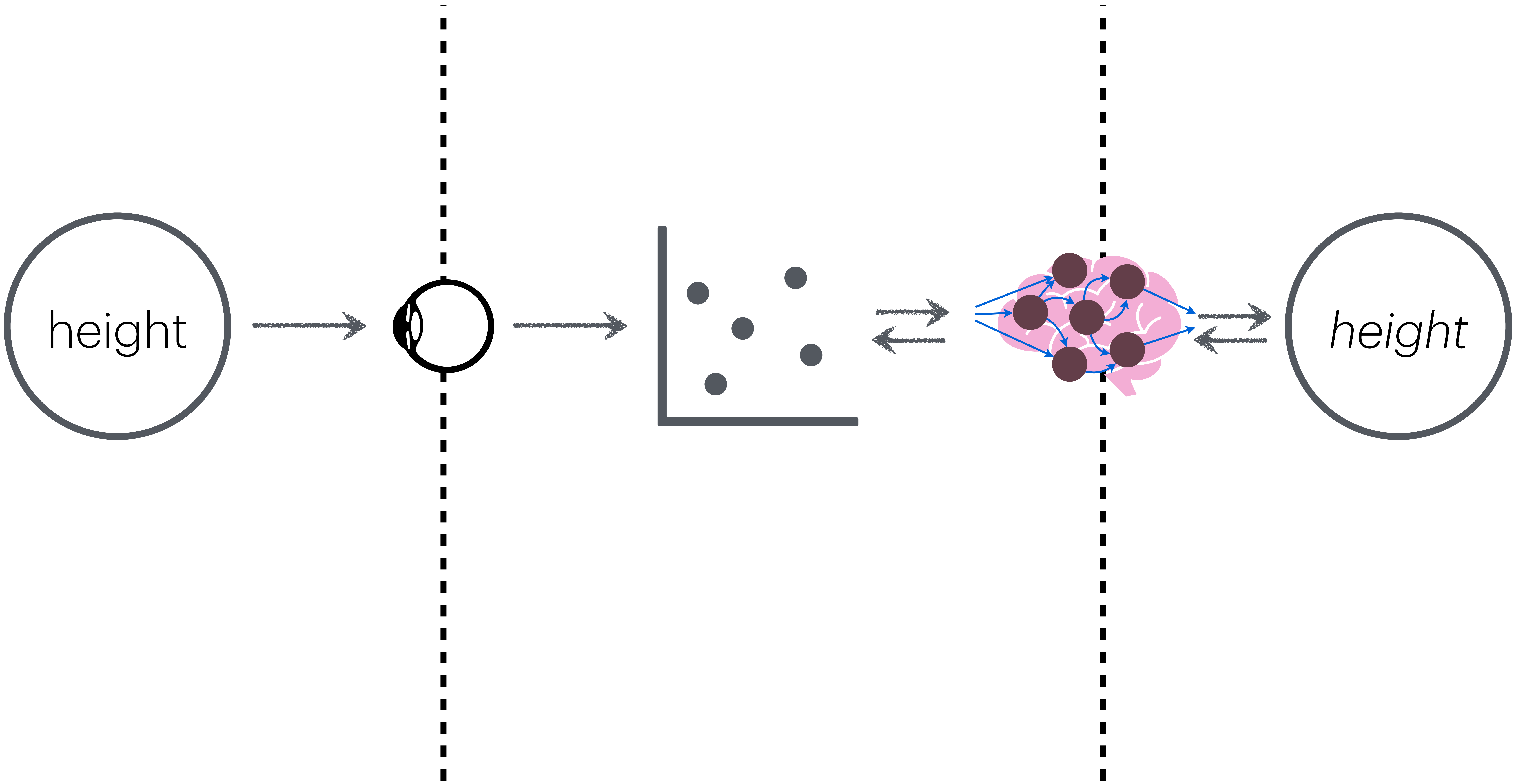
Infomax:

- If the mapping $x \longrightarrow y$ is invertible, $q(\phi, y)$ maximises estimation accuracy
- If the mapping $x \longrightarrow y$ is not invertible, infomax cannot know which x caused y

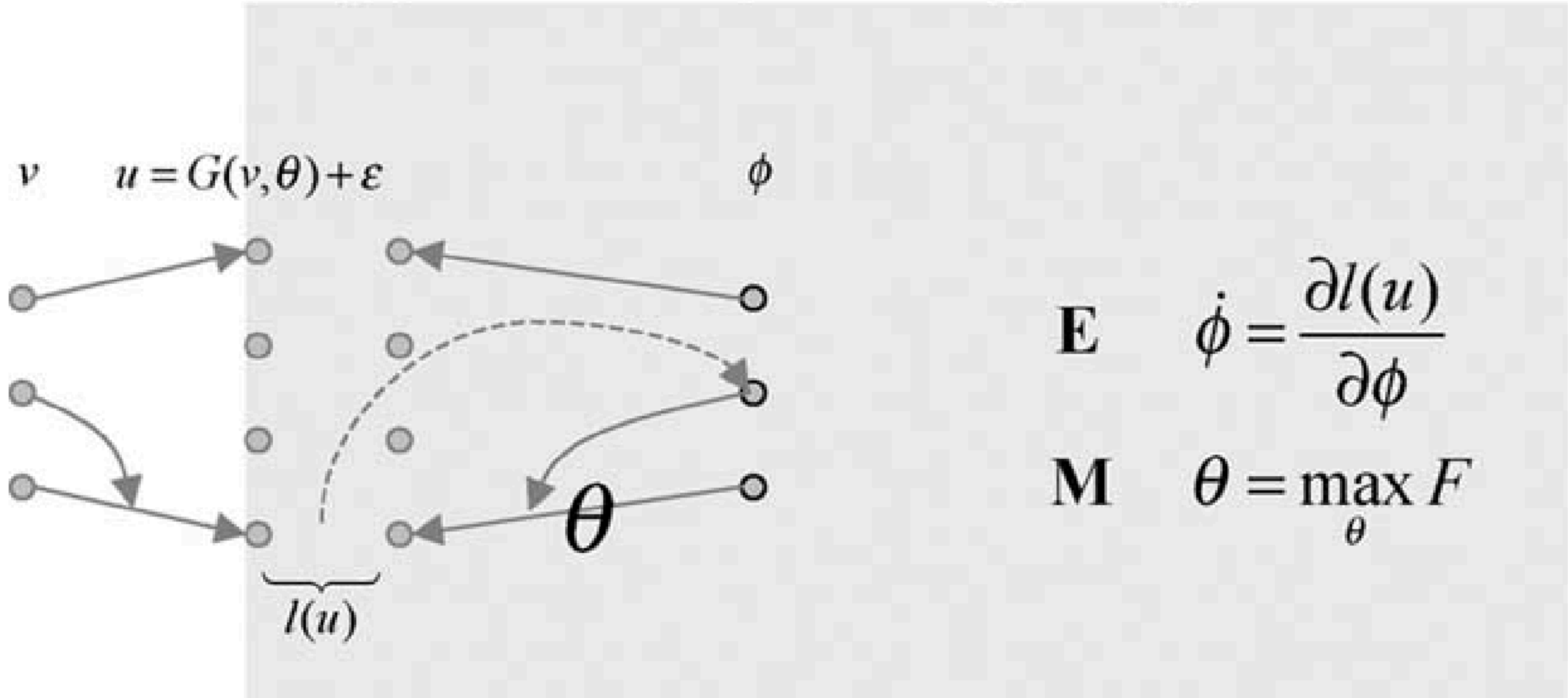
Representational learning and Predictive Coding

1. Representational learning
2. **Predictive coding as representational learning**
3. Perceptual inference
4. Predictive coding as variational perceptual inference
5. Empirical evidence for and controversies of predictive coding
6. The sampling hypothesis
7. Conclusion

Predictive coding



Predictive coding



$$\begin{array}{ccccc} & y & \longrightarrow & q(y, \phi) = \hat{x} & \searrow \\ x \nearrow & & & & \hat{x} \\ & \hat{y} & \longleftarrow & g(\hat{x}, \theta) = \hat{y} & \swarrow \end{array}$$

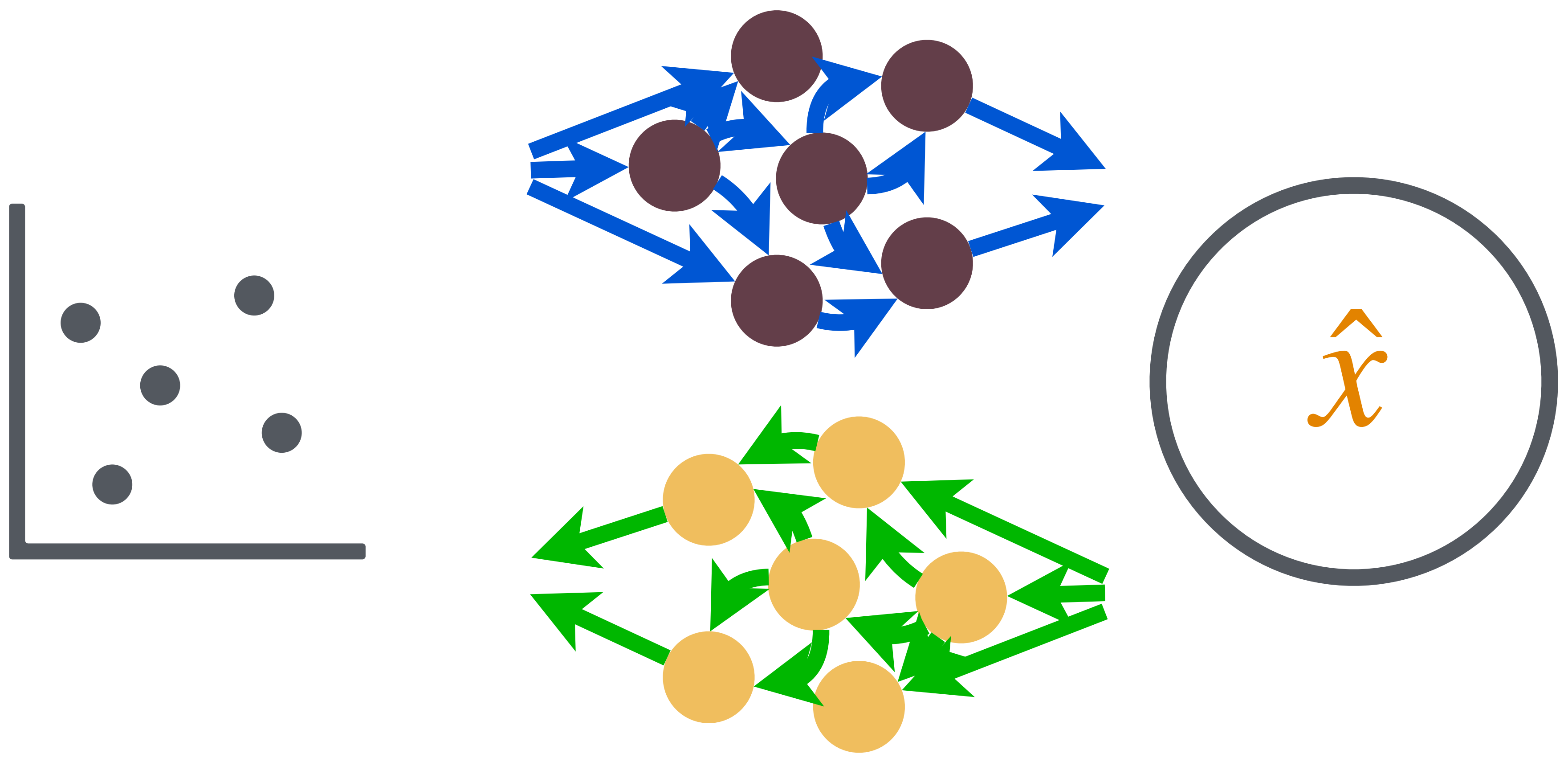
Predictive coding

$$\varepsilon^2(y, \theta, \phi) = \left(y - \hat{y}\right)^2$$

$$\theta = \operatorname{argmin}_{\theta}\ \varepsilon^2(y, \theta, \phi)$$

$$\phi = \operatorname{argmin}_{\phi}\ \varepsilon^2(y, \theta, \phi)$$

$$\hat{x} = \operatorname{argmin}_{\hat{x}}\ \varepsilon^2(\hat{x}, \theta)$$

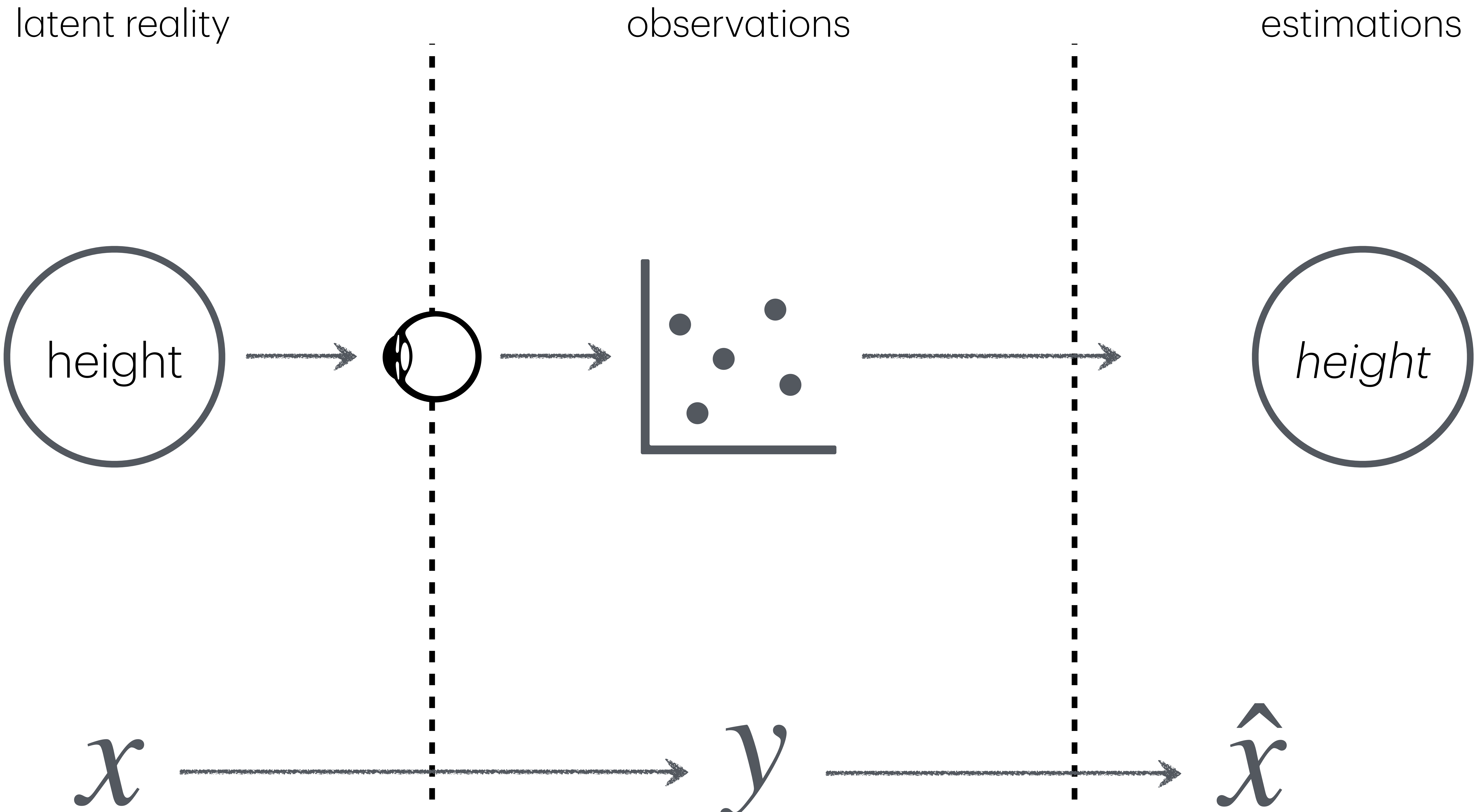


$$y \longrightarrow q(y, \phi) = \hat{x}$$
$$\hat{y} \longleftarrow g(\hat{x}, \theta) = \hat{y}$$

Representational learning and Predictive Coding

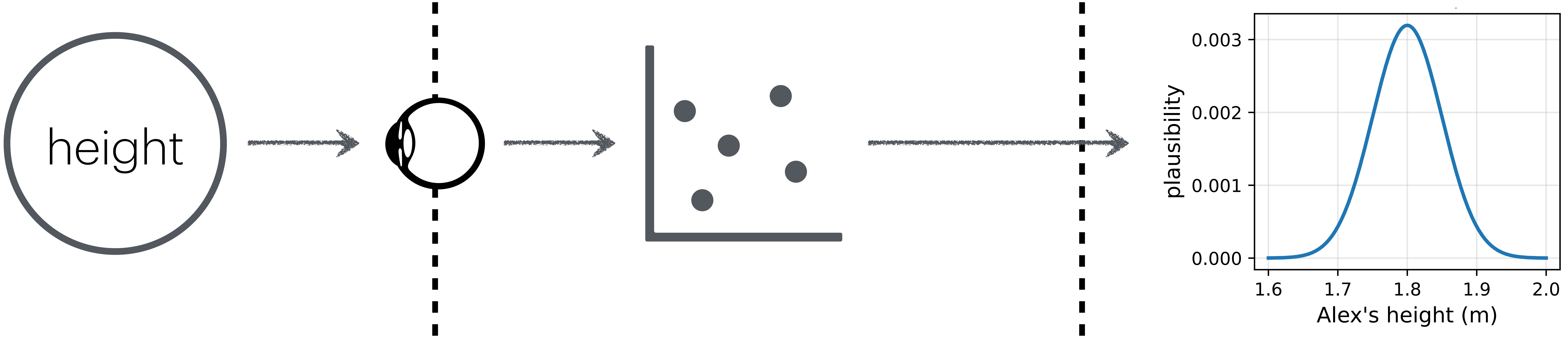
1. Representational learning
2. Predictive coding as representational learning
3. **Perceptual inference**
4. Predictive coding as variational perceptual inference
5. Empirical evidence for and controversies of predictive coding
6. The sampling hypothesis
7. Conclusion

Perception = estimate hidden causes



Perception = estimate hidden causes

latent reality | observations | estimations



$$x \longrightarrow y \longrightarrow P(x \mid y, I)$$

The Bayesian Brain Hypothesis

1. **First tenet: internal representations are probabilistic**
2. **Second tenet: cognition rests on priors**
3. **Third tenet: the brain performs Bayesian inference**
4. **Zeroth tenet: many aspects of cognition are optimal**

Third tenet

The brain performs Bayesian inference to infer the latent causes of the observations

Bayesian perception takes generation into account

If we know how the observations are generated, this should improve our estimates

- **Optimal perception takes into account all available information**

The generation of the observations has two parts:

- Process through which latent variables are generated
- Process through which the latent variables are transformed into observations

Bayesian perception takes generation into account

Example: spoken language

Latent variable: the word that the person wants to communicate. Depends on:

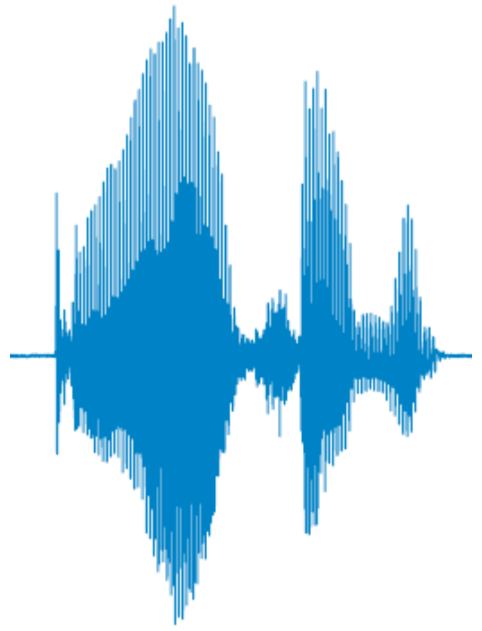
- Context
- What was said before
- Intentions
- ...

Observation: the speech signal the conversation partner receives. Depends on:

- Accent
- Vocal properties of the speaker
- Intrinsic variability of the signal
- External sources of noise
- ...

Step 1: Generation of the latent reality

latent reality | observations | estimations

"Bayes is the best” → 

height

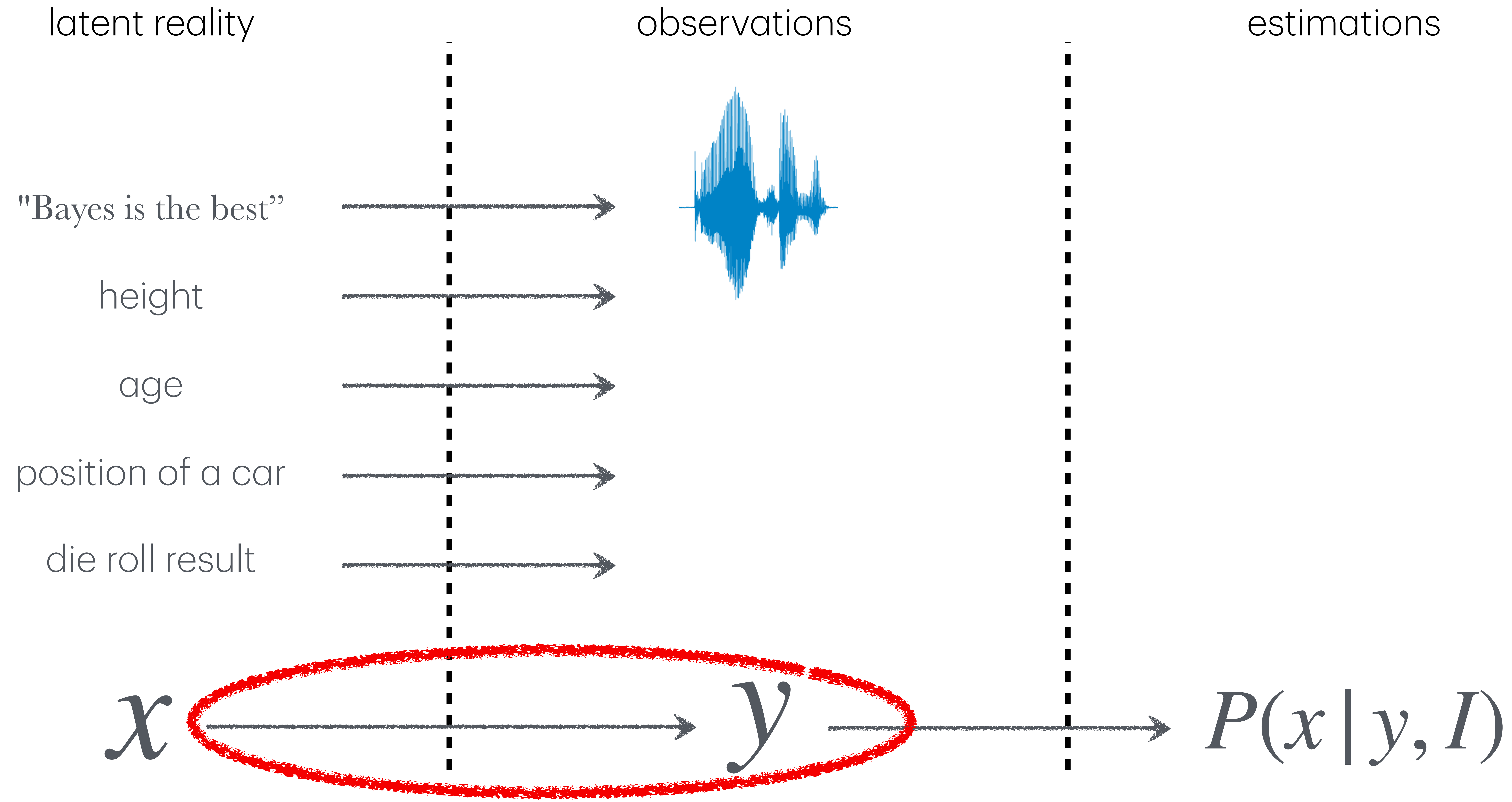
age

position of a car

die roll result

$$x \longrightarrow y \longrightarrow P(x \mid y, I)$$

Step 1: Generation of the latent reality



The brain ignores how exactly reality is generated

If we perfectly knew the process generating x we wouldn’t need perception:

- We would already know the ground truth of everything that has and is to happen.

There is a lot the brain doesn't know.

But we would greatly benefit from encoding what we know:

- That a speaker has a particular accent, timbre, and pitch
- That we are communicating in English
- That we are talking about Bayesian inference

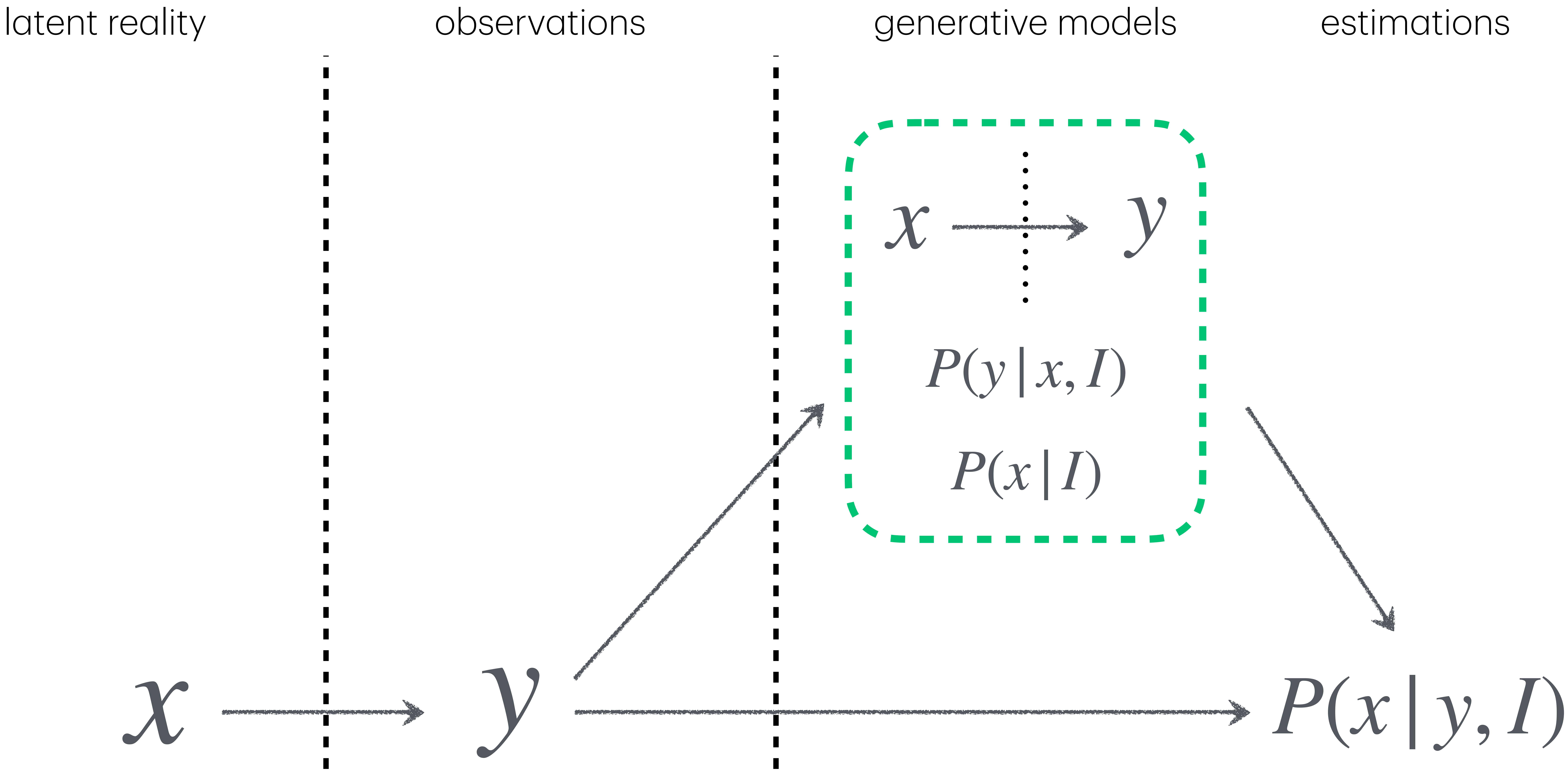
We can encode what we know **and what we don’t know** using models!

We can model our knowledge about how reality is generated

The Bayesian Brain hypothesis of perception:

- **the brain keeps internal generative models (GMs) of how the observations are generated**
- **GMs are themselves estimated** using optimal inference
- GMs contain everything we know about the sensory world
- **GMs explicitly model their uncertainty** about what the do not know
- **the brain perceives by performing optimal inference under its internal GMs**

The internal generative models power perception



Adding uncertainty to the generative models

Most if not all of what we observe is deterministic:

- Stochastic models like $y = x +$ noise are not descriptions of the randomness of the system
- They are descriptions of our uncertainty about factors that influence the data
- None of these systems are really stochastic:
  - die rolls
  - weather
  - classical computing random number generators

Our **generative models use stochasticity to model what we don't know**

Bayes rule (reminder)

Likelihood
how plausible would
y be if latent was x

Posterior
what we know
about x after seeing y

Prior
what we know
about x before seeing y
(knowing I)

$$P(x \mid y) = \frac{P(y \mid x)\, P(x)}{P(y)}$$

Marginal
how plausible is y
(knowing I)

Using GMs to estimate priors

The model of the generation of x allows us to estimate the prior $P(x)$:

- E.g: If know the speaker's language and manners, we can predictions what they will say next

Consider a possible estimation for the prior on the height of a toddler:

- What's your internal generative model for the ground truth?
- Assume a linear model: $h(\text{age}) = h_0 + \text{growth} \times \text{age} \quad \text{with} \quad \text{growth} = 2\,\text{cm/month}, \quad h_0 = 50\,\text{cm}$

We remember the todler was born around 4 months ago, but not the exact day:

- What are our beliefs for the age?

$$P(\text{age}, I) = \mathcal{N}(\text{age}; \mu = 4\,\text{months}, \sigma = 0.2\,\text{months})$$

- What are our prior beliefs for the height?

$$P(h \mid I) = \mathcal{N}(h; \mu = 50 + \text{growth} \times \mu_{\text{age}}, \sigma = \text{growth} \times \sigma_{\text{age}}) = \mathcal{N}(h; \mu = 62\,\text{cm}, \sigma = 0.8\,\text{cm})$$

Using GMs to estimate priors

Generative models also take into account uncertainty on their parameters:

- What's your internal generative model for the ground truth?
- Assume a linear model: $h(\text{age}) = h_0 + \text{growth} \times \text{age}$
- Assume we have uncertainty in $P(h_0 \mid I) = \mathcal{N}(h_0; \mu = 50\,\text{cm}, \sigma = 5\,\text{cm})$

To calculate our prior beliefs for the height we need to sum across all possible h_0:

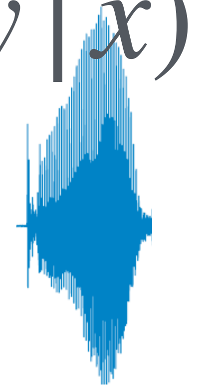
$$P(h \mid I) = \sum_{h_0} P(h_0)\, \mathcal{N}(\mu = h_0 + \text{growth} \times \mu_{\text{age}}, \sigma = \text{growth} \times \sigma_{\text{age}}) = \mathcal{N}(h; \mu = 62\,\text{cm}, \sigma = 5.06\,\text{cm})$$

GMs generally account for uncertainty across all parameters including growth

- Maths get increasingly long and complex with every additional parameter

Using GMs to estimate likelihoods

The model of the mapping $x \rightarrow y$ allows us to estimate the likelihood $P(y \mid x)$:

- E.g: If know the speaker's accent, we can estimate how likely is to hear  if they said "Bayes"

We finally meet the toddler. They look around 75cm tall. How could we estimate our likelihood?

- Assume we are unbiased observers: $p(h_{\text{obs}} \mid h_{\text{real}}) = \mathcal{N}(h_{\text{obs}}; \mu = h_{\text{real}}, \sigma = 2\,cm)$

Side note: the maximum likelihood estimator

- Maximum likelihood is the estimator used in classical statistics
- It assumes that the correct solution is the x that maximises $P(y \mid x)$
- In this case, that would be $x = 75$ cm

Using GMs to estimate marginals

The marginal can be computed integrating across all priors and likelihoods.

For the example of the toddler:

$$p(h_{\text{obs}} \,|\, I) = \sum_h p(h_{\text{obs}} \,|\, h = h_{\text{real}})\, p(h = h_{\text{real}}, I)$$

Computing marginals is computationally expensive and, sometimes, impossible:

- We need to consider all potential ground truths
- We need to compute likelihoods and priors for each of them

In practice, however, don’t really need them to make decisions or to estimate our beliefs

Computing perceptual posteriors

We are finally ready to compute our posterior on the toddler's height!

$$p(h_{\text{real}} \,|\, h_{\text{obs}}) = \frac{p(h_{\text{obs}} \,|\, h_{\text{real}})}{p(h_{\text{real}})}$$

If we just want to estimate the real heigh of the toddler, we need to select:

- h equal to the mean of $p(h_{\text{real}} \,|\, h_{\text{obs}})$, to minimise squared error
- h equal to the median of $p(h_{\text{real}} \,|\, h_{\text{obs}})$, to minimise absolute error
- h for which $p(h_{\text{real}} \,|\, h_{\text{obs}})$ is maximum (maximum a posteriori, or MAP) to maximise our chances to be exactly right

Luckily, our posterior is symmetric, which means all these are the same

$$\hat{h} = h_{obs} \times \frac{\sigma_{\text{obs}}}{\sigma_{\text{obs}} + \sigma_{\text{prior}}} + \mu_{prior} \times \frac{\sigma_{\text{prior}}}{\sigma_{\text{obs}} + \sigma_{\text{prior}}} = 65.7 \text{ cm}$$

Computing posteriors as belief updating

$$\hat{h} = h_{obs} \times \frac{\sigma^2_{\text{obs}}}{\sigma^2_{\text{obs}} + \sigma^2_{\text{prior}}} + \mu_{prior} \times \frac{\sigma^2_{\text{prior}}}{\sigma^2_{\text{obs}} + \sigma^2_{\text{prior}}} = 65.7 \text{ cm}$$

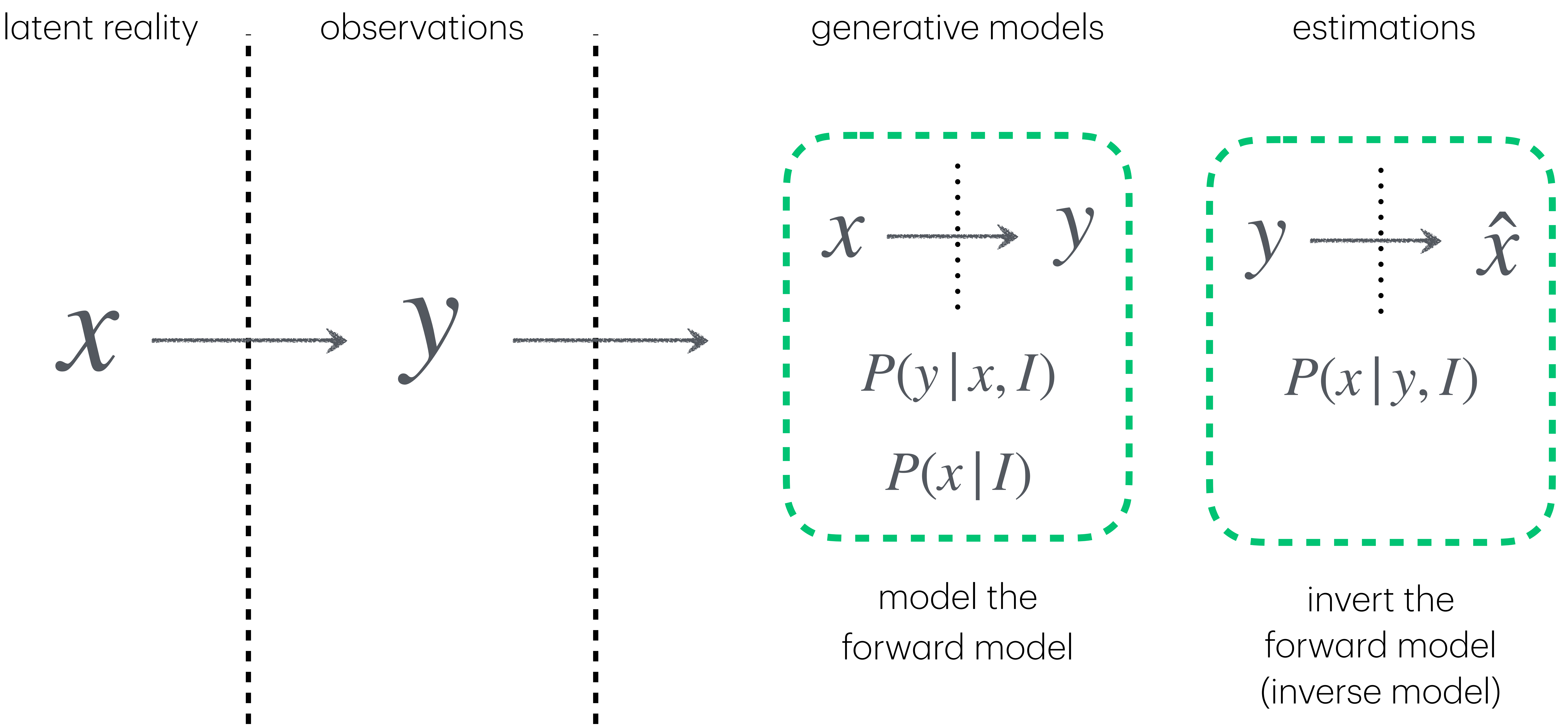
We can rewrite our maximum a posteriori estimate as:

$$\hat{h} = k\, h_{obs} + (1 - k)\, h_{prior} \quad \text{with} \quad k = \frac{\sigma^2_{\text{obs}}}{\sigma^2_{\text{obs}} + \sigma^2_{\text{prior}}}$$

The **optimal estimate is a weighted average between what we believed and what we observed**

- Weights scale with our relative uncertainty in each of them

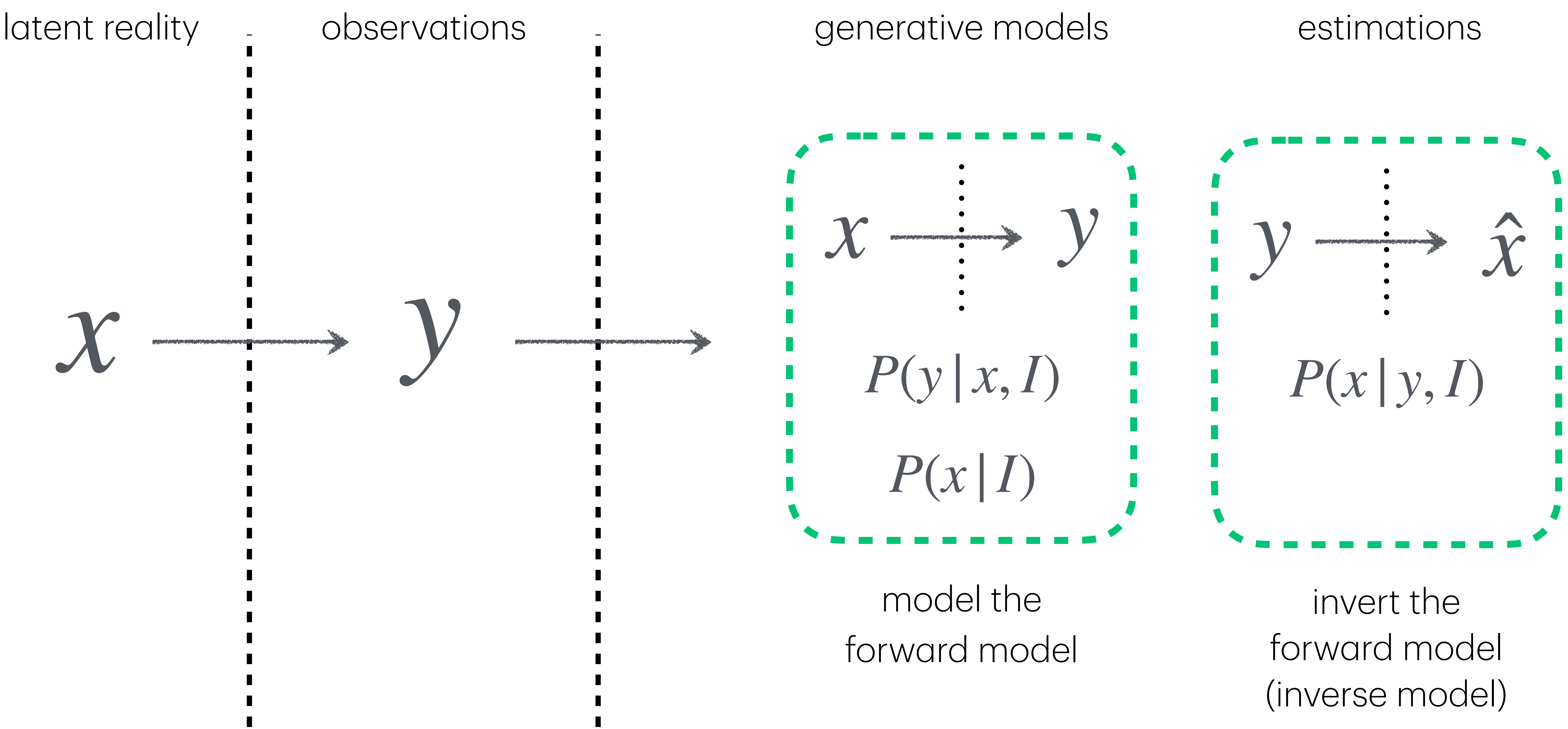
There and back again



Representational learning and Predictive Coding

1. Representational learning
2. Predictive coding as representational learning
3. Perceptual inference
4. **Predictive coding as variational perceptual inference**
5. Empirical evidence for and controversies of predictive coding
6. The sampling hypothesis
7. Conclusion

The Bayesian Brain hypothesis of perception



How does the brain implement optimal inference?

The Bayesian Brain Hypothesis makes no claims on the underlying mechanisms:

- The theory lives in the normative realm: it explains the “why” of the computations, not the “how"

Mechanistic theories are part of the Bayesian framework, but tangential to the hypothesis

The main problem the theories address is that optimal inference is, in practice, impossible:

- High-dimensional latent spaces
- Nonlinear, multimodal likelihoods
- **Exact inference is computationally intractable**

Two major families of mechanistic theories

Mechanistic theories generally **assume that inference is approximate**

The two main families differ in their approach to approximate inference:

- **Variational** approach: **approximate the representation** of complex distributions
- **Sampling** approach: **approximate the computation** of complex operations by sampling

Variational theories

Insight: what makes computation intractable is he complexity of the distributions

Approach: approximate distributions by simpler representations (usually: Gaussians)

- $p(x \mid I) \sim q(x; \theta = \{\mu, \sigma\}) = \mathcal{N}(x; \mu, \sigma^2)$-
- Approximations are derived by optimising θ to minimise the expected divergence between $q(x; \theta)$ and $p(x \mid I)$
- $q(x; \theta)$ is not necessarily a Gaussian, just a simpler approximation of $p(x \mid I)$

Variational Free Energy

Once we have parametrised the problem, we find the parameters that minimise:

$$KL\left[q(x;\theta||p(x|y)\right]$$

where $KL[f||g]$ is a measure of how different f and g are from each other.

Computing the posterior $p(x|y)$ is unfeasible because $p(y,I)$ is untractable, however:

$$KL\left[q(x;\theta||p(x|y)\right] \leq KL\left[q(x;\theta||p(x,y)\right] = KL\left[q(x;\theta||p(y|x)\,p(x)\right] \equiv \mathcal{F}$$

So if we minimise $\mathcal{F}$ we minimise a high-bound for the actual divergence

The Laplace assumption

Predictive coding is a variational theory that approximates:

- all probabilistic representations as Gaussian distributions (Laplace assumption)

If all distributions are Gaussian, then the likelihood is:

$$p(y|x) = \mathcal{N}(y; x, \sigma_r^2) = \frac{1}{\sqrt{2\pi}\sigma_r} e^{-\frac{(x-y)^2}{2\sigma_r^2}}$$

And the prior is

$$p(x) = \mathcal{N}(x; \mu_p, \sigma_p^2) = \frac{1}{\sqrt{2\pi}\sigma_p} e^{-\frac{(x-\mu_p)^2}{2\sigma_p^2}}$$

The Free Energy Principle

Computing the KL between both distributions:

$$\mathcal{F} = KL\left[q(\hat{x};\theta\,||\,p(y\,|\,x)\,p(x)\right] = \frac{1}{\sigma_p^2}\varepsilon_p^2 + \frac{1}{\sigma_r^2}\varepsilon_r^2 - \log 4\pi\sigma_r\sigma_p$$

$$\varepsilon_p = \hat{x} - \mu_p$$

$$\varepsilon_r = y - \hat{x}$$

To maximise accuracy: minimise prediction errors, weighted by precision (inverse variance)

$\mathcal{F}$ is called *Free Energy* and the result is called *the Free Energy Principle*

Computing posteriors as belief updating

Optimal posterior estimate with Gaussian beliefs:

$$\hat{h} = k\, h_{obs} + (1 - k)\, h_{prior} \quad \text{with} \quad k = \frac{\sigma_{\text{obs}}}{\sigma_{\text{obs}} + \sigma_{\text{prior}}}$$

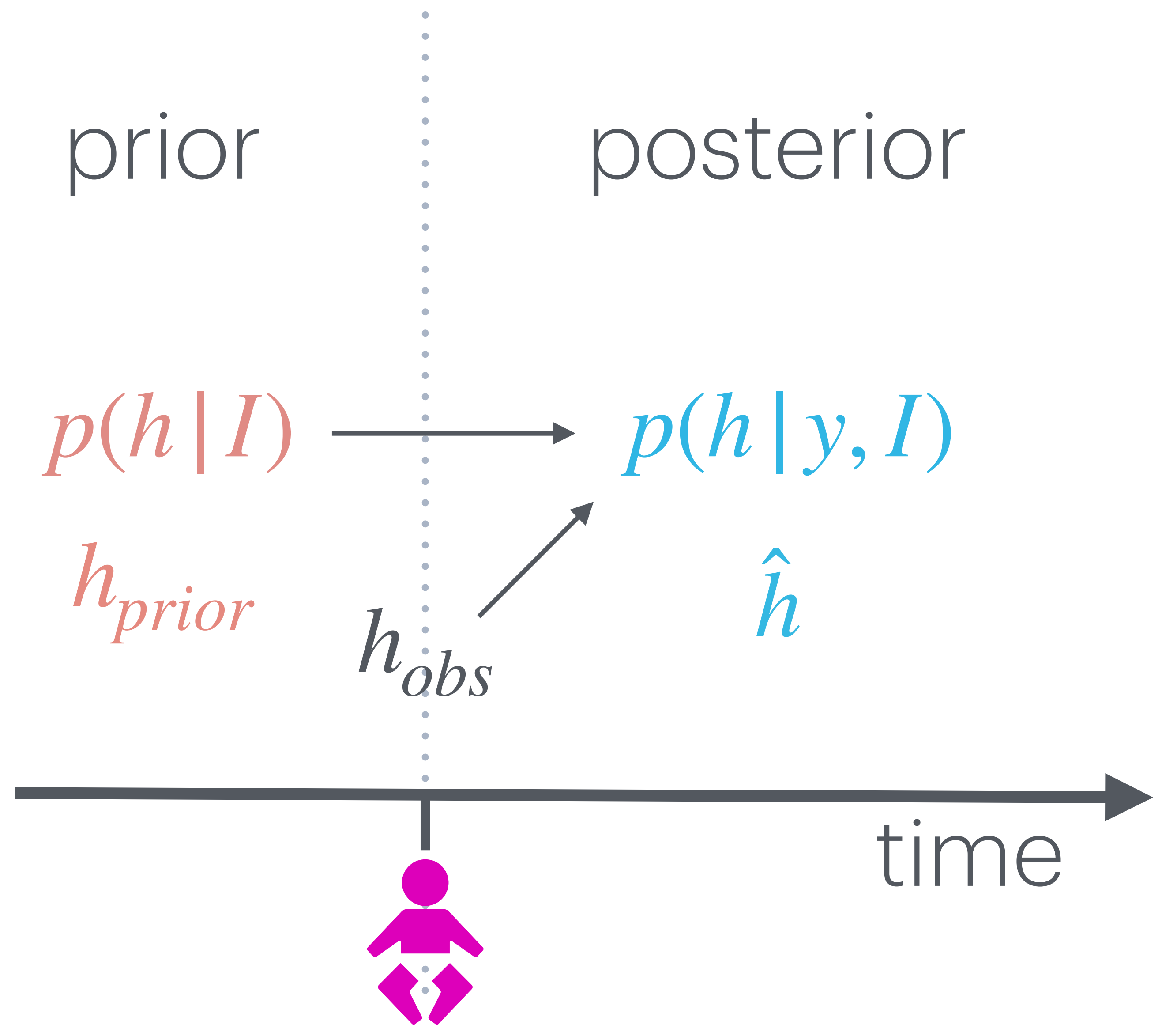
This can be neatly rewritten as:

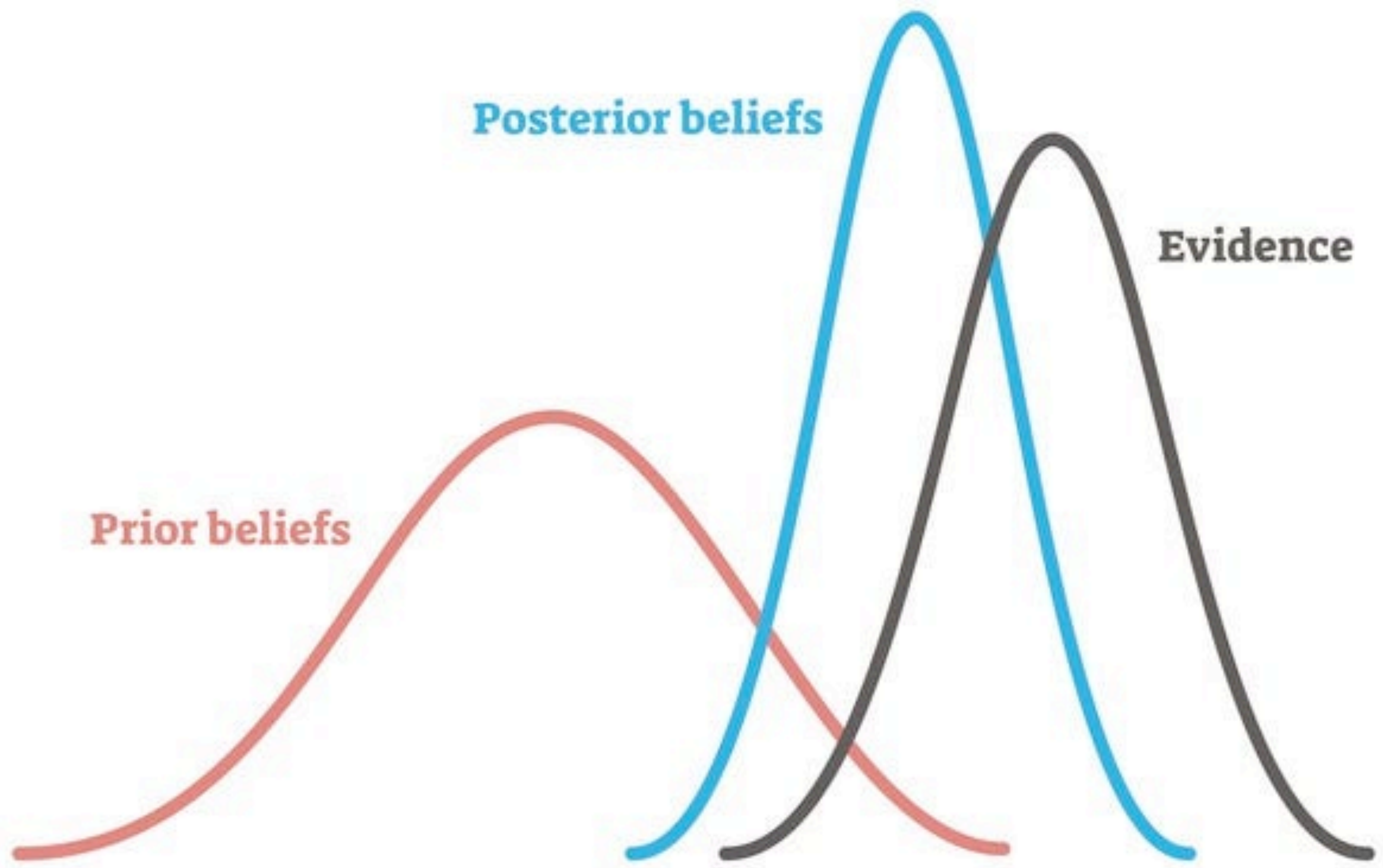
$$\hat{h} = h_{prior} + k\,(h_{obs} - h_{prior}) = h_{prior} + k\,\varepsilon_{obs}$$

Predictive coding is the optimal solution to perceptual inference when all distributions are Gaussian

Computing posteriors as belief updating

Belief updating: $\hat{h} = h_{prior} + k\,(h_{obs} - h_{prior}) = h_{prior} + k\,\varepsilon_{obs}$





Representational learning and Predictive Coding

1. Representational learning
2. Predictive coding as representational learning
3. Perceptual inference
4. Predictive coding as variational perceptual inference
5. **Empirical evidence for and controversies of predictive coding**
6. The sampling hypothesis
7. Conclusion

Empirical evidence for predictive coding

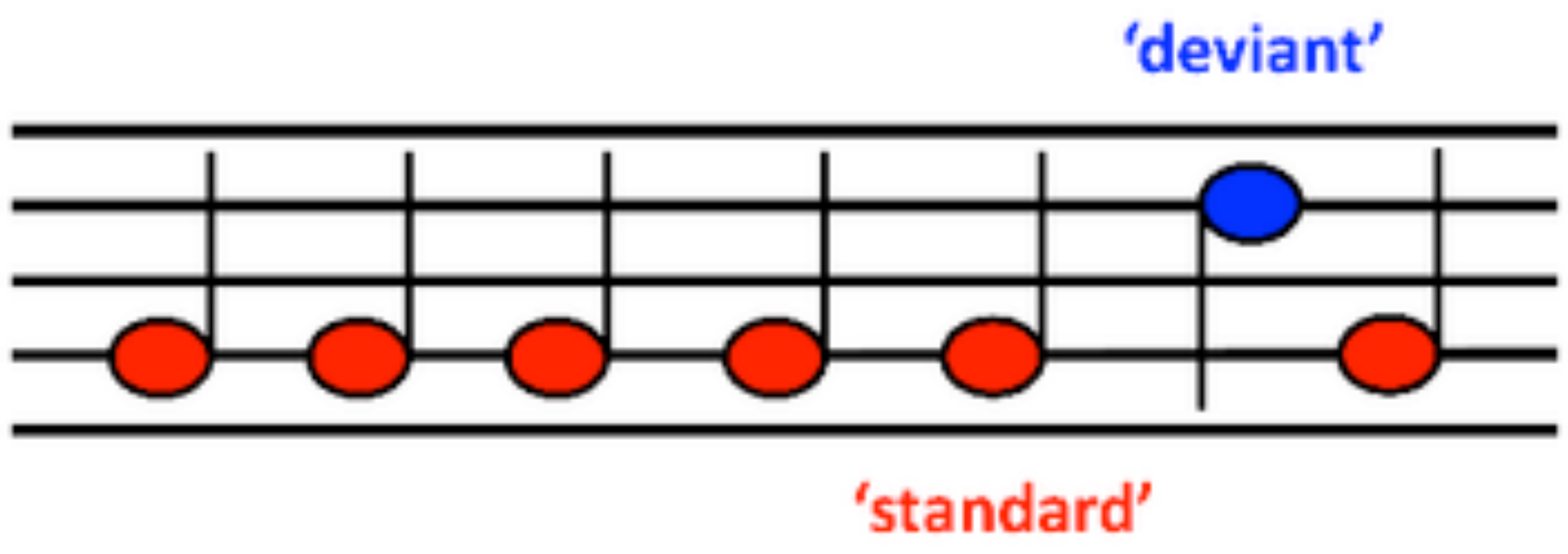
Many independent sources of evidence:

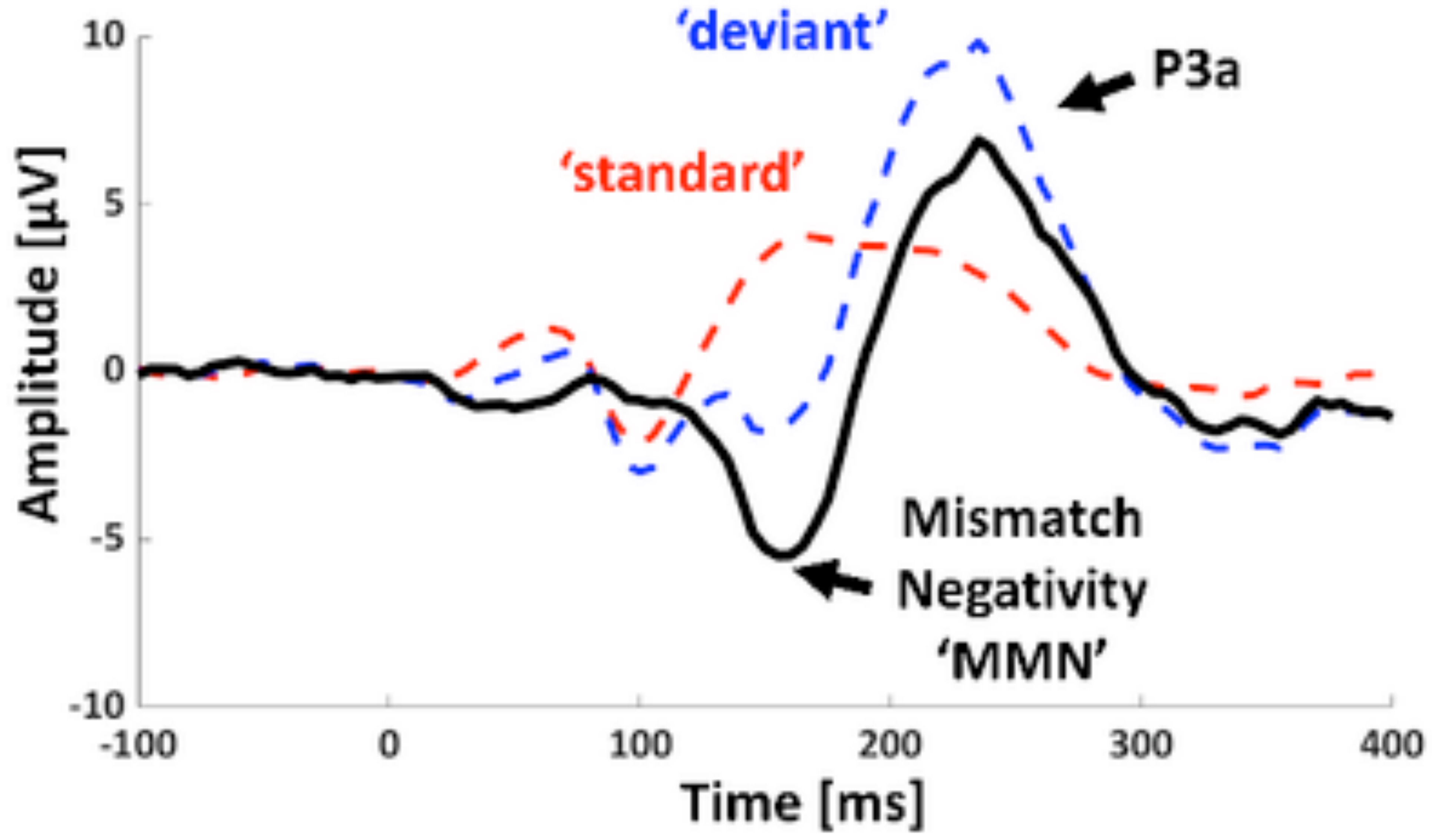
- Human electrophysiology: event-related potentials ~ prediction error
- Human fMRI: correct expectations → lower BOLD responses, sharpened representations
- Animal electrophysiology: neuronal responses ~ prediction error
- Physiology: laminar microcircuits and oscillatory channels match PC requirements

We will see only a few examples

Evidence from human electrophysiology

The strength of the mismatch negativity matches the expected amount of prediction error





doi.org/10.1016/j.clinph.2008.11.029

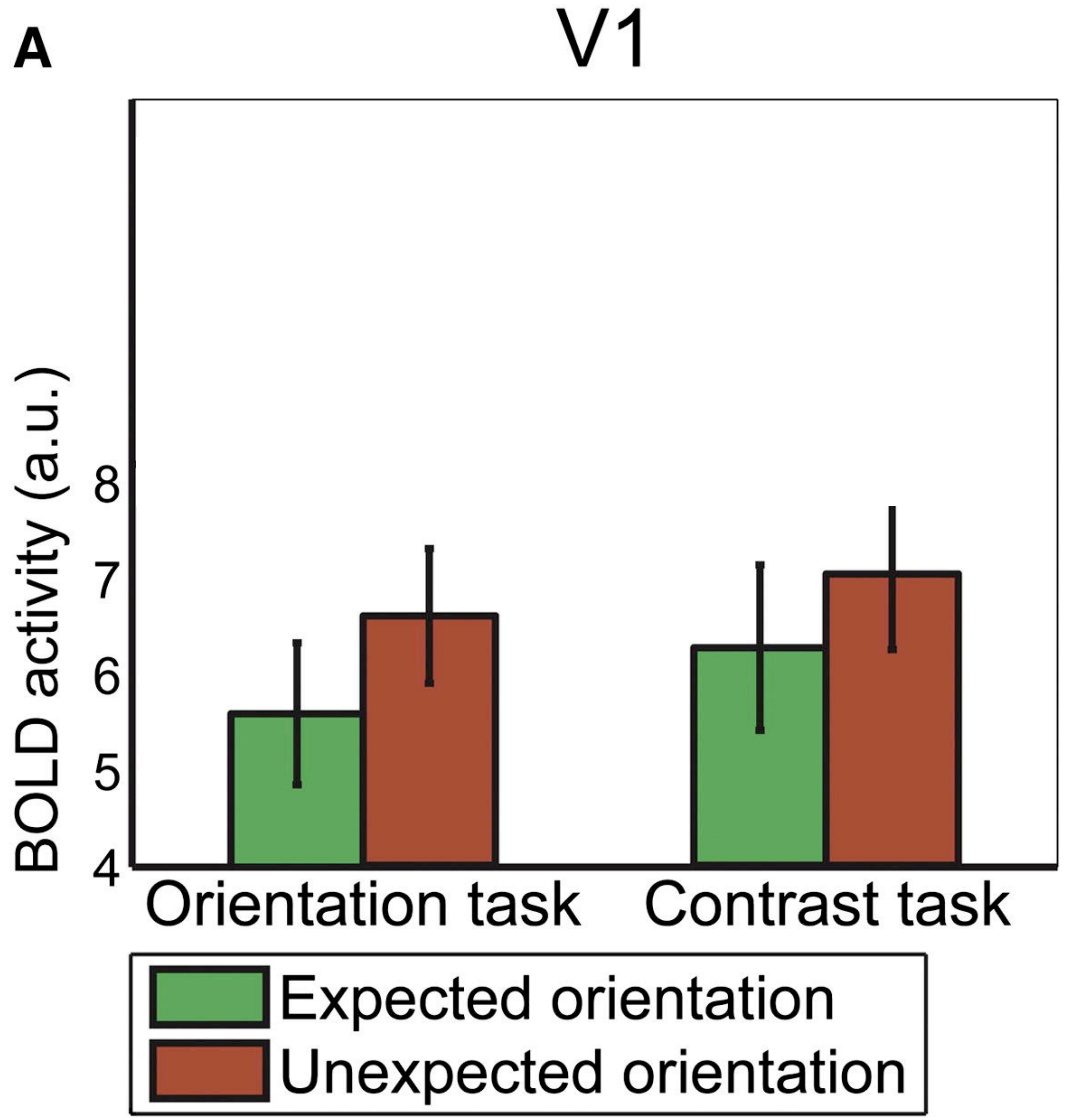
doi.org/10.1523/JNEUROSCI.5003-11.2012

Evidence from human BOLD responses

Correct expectations evoke, in human V1:

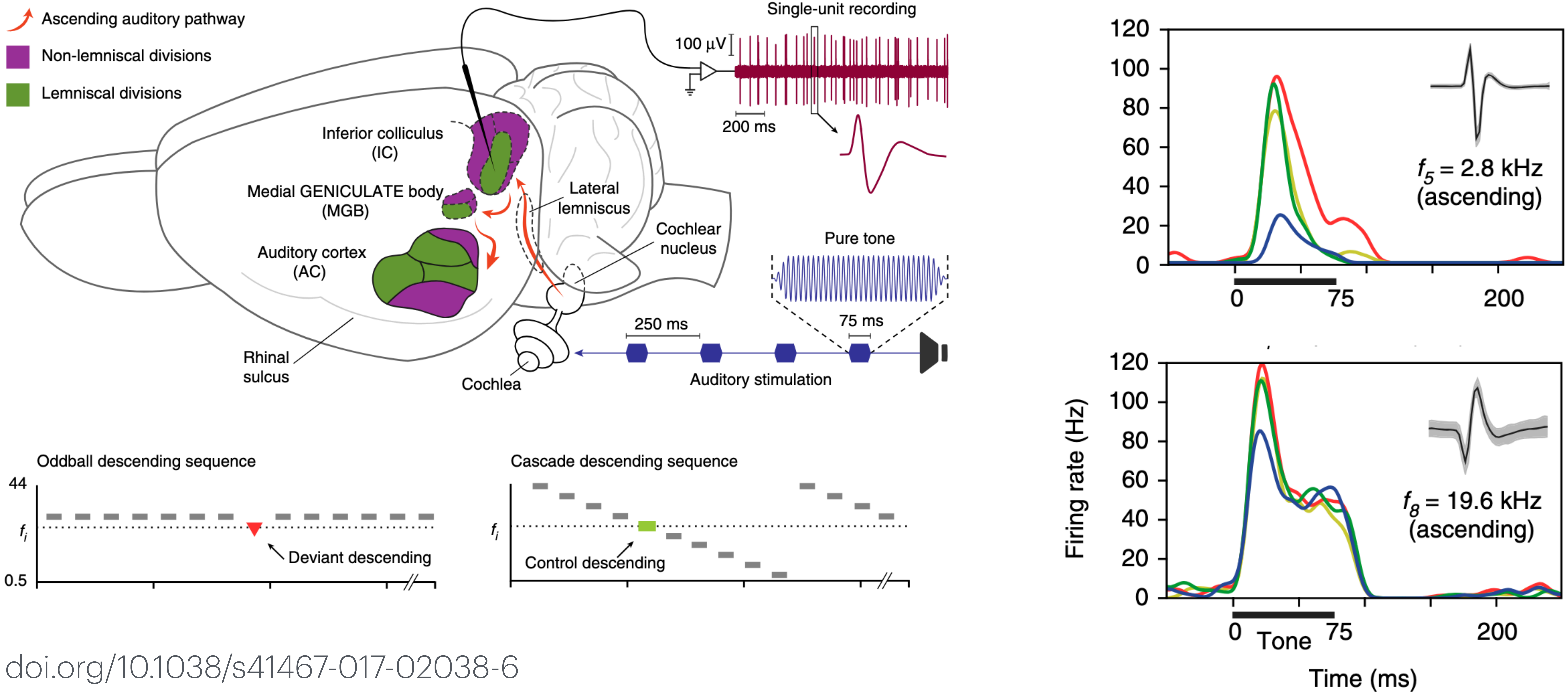
- less activity

- responses from which the stimuli are easier to decode (sharpening)



doi.org/10.1016/j.neuron.2012.04.034

Evidence from animal electrophysiology



doi.org/10.1038/s41467-017-02038-6

Evidence from neuroanatomy

doi.org/10.1016/j.neuron.2012.10.038

Controversies of predictive coding

- Prediction error vs computation: **is** the **correlation** between ε and responses **evidence** for the encoding of ε?
- Results from laminar recordings are **mixed** with respect to the layer microcircuit implementation
- Responses, even mesoscopic in humans, not always correlate with ε. When they don't, authors resort to precision:
  - PC's main tenet is that the brain minimises precession-weighted prediction error
  - Precision originally described as inverse variance
  - Precision later generalised to a **wildcard modulatory signal** (e.g., attention)
- Predictive coding requires Gaussian internal beliefs, but **many of our predictions are muti-modal**
- Predictive coding explains inference of values (x), but it **does not explain how the brain infers events** (A)
- **Is the Free Energy Principle falsifiable**?

Representational learning and Predictive Coding

1. Representational learning
2. Predictive coding as representational learning
3. Perceptual inference
4. Predictive coding as variational perceptual inference
5. Empirical evidence for and controversies of predictive coding
6. **The sampling hypothesis**
7. Conclusion

Inference by sampling: Monte Carlo

Monte Carlo:

- collection of **methods for approximating probabilistic computations**
- **simulates random processes** instead of deriving closed forms
- a simulation of the process is called “a sample”
- accuracy improves with the number of samples we similate
- works even when the distribution is complex, high-dimensional, or has no closed form

Monte Carlo example: expected values

Goal: estimate the expected value of a variable X drawn from an unknown distribution

Procedure: draw N random samples $x_1, x_2, \ldots, x_N$

- Monte Carlo estimate: $E[X] \approx \frac{1}{N} \sum_{i=1}^{N} x_i$

Example: Suppose we sample $N = 5$ values: 2.1, 1.7, 2.4, 1.9, 2.0

- $E[X] \simeq (2.1 + 1.7 + 2.4 + 1.9 + 2.0)/5 = 2.02$

Monte Carlo example: estimating π

Goal: estimate π as the ration between:

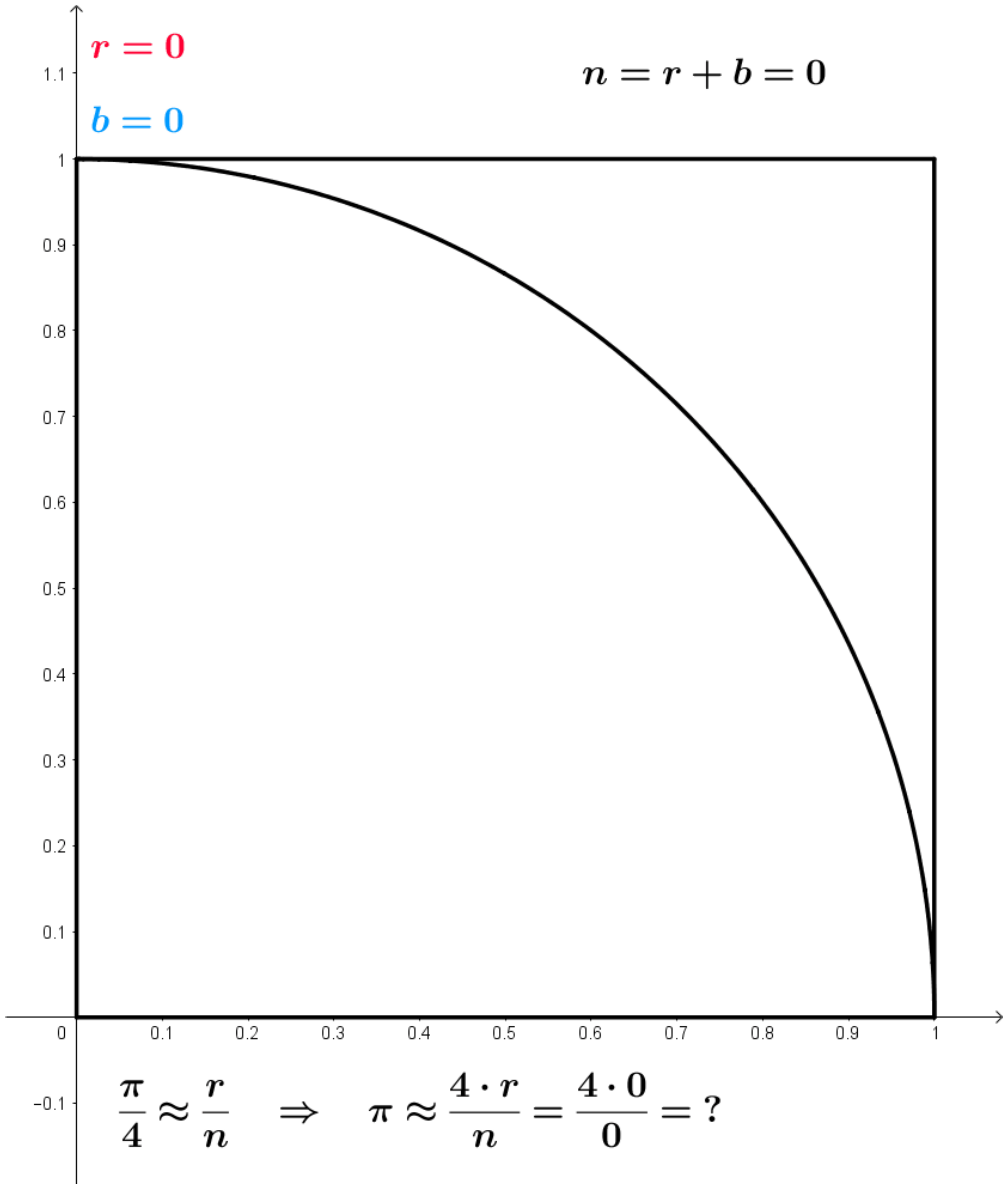
- The area of a square with sides of length 1 $A_\square = 1^2 = 1$
- The area of a quarter of a circle or radius 1 $A = \pi r^2/4 = \pi/4$
- If we know both areas, we can estimate $\pi = 4\,A_\circ/A_\square$

Procedure:

- draw N random pairs of samples (y, x) with $x, y \in [0,1]$
- plot each pair and check if they are inside the circle:

  (they are in the circle if $x^2 + y^2 < 1$)

Approximate $\dfrac{A_\circ}{A_\square} \simeq \dfrac{N_{\text{in circle}}}{N_{\text{total}}}$ and $\pi \simeq 4 \simeq \dfrac{N_{\text{in circle}}}{N_{\text{total}}}$



The sampling hypothesis

Hypothesis: **the brain represent beliefs through the statistics of samples**

Example: compute the prior for the height of a toddler with:

- $h(\text{age}) = h_0 + \text{growth} \times \text{age}$
- $P(\text{age}, I) = \mathcal{N}(\text{age}; \mu = 4 \text{ months}, \sigma = 0.2 \text{ months})$
- $P(h_0 \,|\, I) = \mathcal{N}(h_0; \mu = 50 \text{ cm}, \sigma = 5 \text{ cm})$
- $P(\text{growth} \,|\, I) = \mathcal{N}(\text{growth}; \mu = 2 \text{ cm/month}, \sigma = 0.2 \text{ cm/month})$

Sampling the toddler height problem

Sampling solution:

for sample s in 1, 2, …, N:

  sample age $\sim \mathcal{N}(\mu = 4\,\texttt{months}, \sigma = 0.2\,\texttt{months})$

  sample growth $\sim \mathcal{N}(\mu = 2\,\texttt{cm/month}, \sigma = 0.2\,\texttt{cm/month})$

  sample $h_0 \sim \mathcal{N}(\mu = 50\,\texttt{cm}, \sigma = 5\,\texttt{cm})$

  compute $h_s = h_0 + \texttt{growth} \times \texttt{age}$

compute $\hat{h} = \frac{1}{N} \sum h_s$

Plausibility of the sampling hypothesis

Is the sampling hypothesis plausible?

- neural noise and recurrent dynamics can implement sampling
- It would explain trial-to-trial variability and probabilistic population codes
- sampling and transformation operations much more simpler than the exact case
- approximation better than the variational approach for complex distributions
- to date there is only indirect evidence for the sampling hypothesis

Representational learning and Predictive Coding

1. Representational learning
2. Predictive coding as representational learning
3. Perceptual inference
4. Predictive coding as variational perceptual inference
5. Empirical evidence for and controversies of predictive coding
6. The sampling hypothesis
7. **Conclusion**

Take home messages

Representational learning is hard:

- The brain needs to learn to identify x given y, but it cannot how to do this by seeing x

Predictive coding is a Bayesian-based **mechanistic** solution:

- Offers a **variational approximation** to Bayesian inference
- Proposes the **specific circuits** implementing the approximation
- **Explains a vast amount of existing data** across multiple domains and techniques
- Many of its assumptions and results are **controversial**, some believe that unfalsifiable

Other Bayesian- GM- based solutions exist:

- The sampling hypothesis is feasible and adds no assumptions on the distributions of beliefs

Questions for next session

Questions for next session: Q4.1

Hierarchical predictive coding assumes a multi-layered architecture where each level predicts the level below. Explain this hierarchy. What represents x, $\hat{x}$, ,y, and $\hat{y}$ at each level? Why might the brain favour this organisation over a flat architecture where all features are predicted in parallel?

Questions for next session: Q4.2

What does it mean to say the brain has a "generative model of the world"? Is this an internal simulation, a set of learned regularities, a collection of statistical dependencies, a causal model, or something else entirely? Defend your interpretation and explain what evidence would distinguish between these possibilities.

Questions for next session: Q4.3

Predictive coding assumes the brain's internal model approximates the true generative process that creates sensory data. Find a few examples where the brain's model is systematically wrong. Why would the brain make these errors, if the strategy of using GMs is optimal?

Questions for next session: Q4.4

The McGurk effect demonstrates that visual information about mouth movements can change what we hear. Explain this phenomenon in terms of predictive coding. How does the brain treat different sensory modalities when forming its generative model?

Questions for next session: Q4.5

Many scientists claim the Free Energy Principle cannot be falsified. Research this controversy: what do critics mean by "unfalsifiable"? What is the difference between a framework that cannot be wrong and a framework that is not empirically testable? Give your own position on the Free Energy falsifiability debate.

Questions for next session: Q4.6

Compare and contrast the sampling hypothesis with firing rate coding and predictive coding. Are they mutually exclusive, or could the brain implement predictive coding through sampling? What would neural sampling of error signals look like?

Questions for next session: Q4.7

Binocular rivalry occurs when different images are presented to each eye, causing perception to alternate between them. Research how predictive coding explains this phenomenon. Could the sampling hypothesis provide an alternative explanation?

Questions for next session: Q4.8

Omission responses are neural responses to the unexpected absence of a stimulus. How does predictive coding explain these responses? In an omission, what is the x, what is y, and what is the prediction error? Do omission responses provide strong evidence for predictive coding or the Bayesian Brain hypothesis?

Questions for next session: Q4.9

When predictive coding predictions fail, researchers invoke "precision weighting" to explain why some prediction errors are weighted more than others. If precision can be adjusted post-hoc to fit any data pattern, does this make the theory unfalsifiable? How would you distinguish a legitimate theoretical parameter from an unfalsifiable rescue device?

Questions for next session: Q4.10

Psychedelic drugs may work by disrupting the predictive apparatus supporting perception. Research this hypothesis. What are the therapeutic and ethical implications?

All course materials:

github.com/qtabs/compneuro4cogneuros