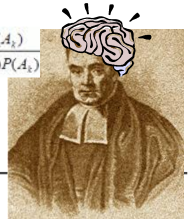
$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{\sum_{k=1}^{n} P(B \mid A_k)P(A_k)}$$

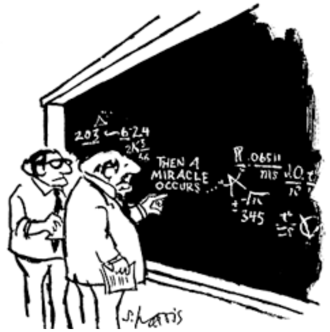


The `Bayesian' approach to perception, cognition and disease

Peggy Seriès,
IANC, University of Edinburgh

The challenge faced by the brain: uncertainty

do i know this person?



Is my model correct?

which is the path they said i should take

Uncertainty everywhere

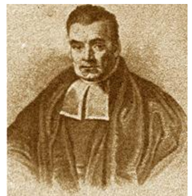
- Humans & animals operate in a world of sensory uncertainty:
  - e.g. mapping of 3D objects to 2D image
  - intrinsic limitations of the sensory systems (e.g. number and quality of receptors in the retina)
  - neural noise

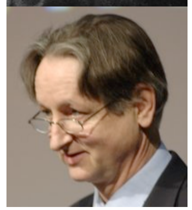
--> multiple interpretations about the world are possible;

- The brain must deal with this uncertainty to generate perceptual representations and guide actions.

- Perception must work *backwards to extract underlying cause of noisy inputs* : unconscious, probabilistic inference
- The brain as a guessing machine.

The Uncertain History of the Bayesian Brain

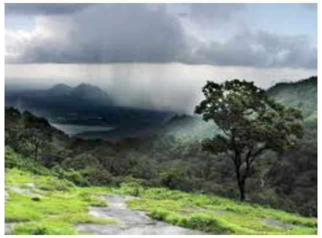




- Bayesian Statistics (mathematics): Thomas Bayes (1702-1761), Pierre-Simon Laplace (1749-1827), Harold Jeffreys (1891-1989), Richard Cox (1898-1991), Edwin Jaynes (1922-1998)
- 1860s: Helmholtz : perception as unconscious inference, making assumptions and conclusions from incomplete data, based on previous experiences.
- 1990s : Geoff Hinton, Peter Dayan - Helmholtz machine -- brain as generative model.
- 2000s --> enters experimental (psychophysics) world, spreads in theoretical world, now physiology?

What is Bayes' theorem about ?

- What is the chance that it will rain today?



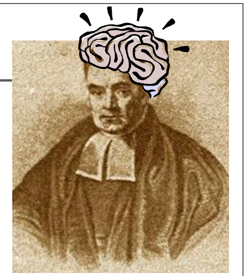
you want to compute P(hle) :

- probability that it is going to rain given the evidence (e.g. the clouds look dark)

you use

- P(elh) : probability of the evidence (that the clouds look dark) when it is actually going to rain (from previous measurements - model of the world).
- P(h): prior knowledge or bias about the probability of rain (before observing any data)

A Bayesian theory of the Brain



Reverend Thomas Bayes, 1702- 1761

- 1990s- Purpose of the brain: infer state of the world from noisy and incomplete data [G. Hinton, P. Dayan, A. Pouget, R. Zemel, R. Rao, etc..]

- Perception often modelled using the framework of Bayesian Inference

$$P(h_1|e) = \frac{P(e|h_1)P(h_1)}{P(e)}$$

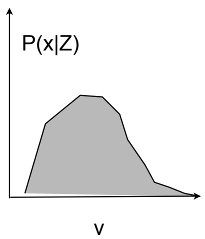
$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{normalizing constant}}$$

manipulating probabilities -- degree of belief.

"Instead of trying to come up with an answer to a question, the brain tries to come up with a probability that a particular answer is correct," Alex Pouget.

Bayesian coding hypothesis (1)

- Hypothesis: information manipulated by sensory systems has the form of a conditional probability density function
- e.g. the position of an object is represented not a single number, x, but P(xlZ), where Z is the available data



- = stores likelihoods = 'generative models', or 'forward model' of the world, P(Zlx), and prior knowledge / state of the world, P(x).
- Given new data Z, the brain computes P(xlZ)

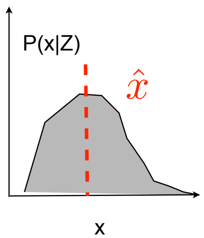
$$P(x|Z) = \frac{P(x, Z)}{P(Z)} = \frac{P(Z|x)P(x)}{P(Z)}$$

Bayes theorem

Bayesian coding hypothesis (2)

- Benefits:

- integrate information efficiently over space & time
- integrate information efficiently from different sensory cues and modalities
- propagate information without committing too early to particular interpretations.



- Commit as late as possible, then collapsing the probability distribution into a single number = decision, or action taken.
- e.g. take the max of the posterior

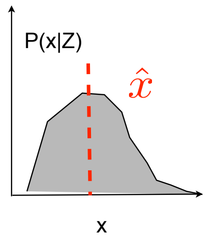
Estimators and cost functions

• How to do that depends on cost function :

• one option is to take the max of the posterior

$$\hat{x} = \text{argmax}_x P(x|Z)$$

this is known to optimize a cost function that is 0 when $\hat{x} = x$ and e=cst otherwise.



max of the posterior

• another option is to take the mean of the posterior

$$\hat{s} = \int x p(x|Z) dx$$

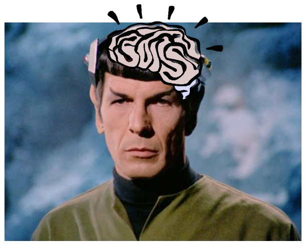
which minimizes the mean squared error $(\hat{x} - x)^2$

• another option is that the brain could use samples from the posterior

The Behavioural Level:

Do People behave as Bayesian Observers?

• Bayesian hypothesis as a benchmark for performance.



Is the Human Brain "Bayesian-optimal"?

• Humans not optimal / achieving the level of performance afforded by the uncertainty in the physical stimulus (e.g. movies)

• The question is:

1 - Do neural computations take into account the uncertainty of measurements at each stage of processing?

2 - Combine it optimally with previous experience?

• testable predictions at the behavioural level

1 - Do brains take into account measurement uncertainty when combining different (simultaneous) information? Combine different sources optimally?

Example: integrating vision and audition

- example: McGurk effect,

Ventriloquism

Why do we get tricked?

http://www.youtube.com/watch?v=G-IN8vWm3m0

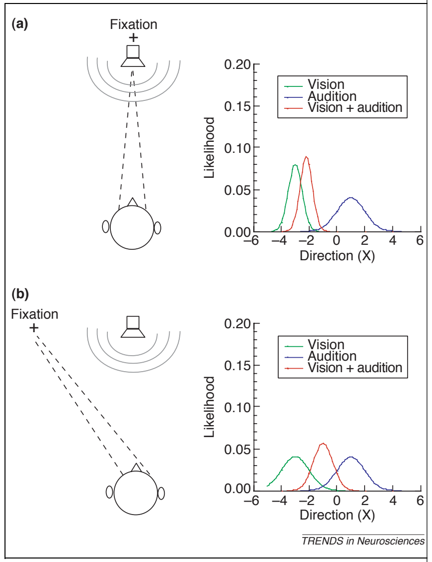
http://www.youtube.com/watch?v=rfNCoSE61w8

Cue Integration (1) : qualitative predictions

- e.g. integration between visual and auditive information

- prediction 1 (position): if visual cue is more reliable, then final estimate is shifted towards visual cue.

- prediction 2 (variance or discrimination threshold): Final discrimination threshold lower than that for each modality ; varies if reliability of one modality varies.



Cue Integration (2):Theory

- Theory tells us how posterior depends on individual likelihoods:

$$\hat{x} = \text{argmax}_x P(x|d_1, d_2)$$

$$P(x|d_1, d_2) = \frac{P(d_1,d_2|x)P(x)}{P(d_1, d_2)} \propto P(d_1|x)P(d_2|x)P(x)$$

- Assuming that the likelihood are gaussian, i.e.

$$P(d_1|x) \propto \exp(-\frac{(d_1 - x)^2}{2\sigma_1^2})$$

- We can determine mean and width of posterior (gaussian):

$$P(d_1|x)P(d_2|x) \propto \exp(-\frac{(d_1 - x)^2}{2\sigma_1^2} - \frac{(d_2 - x)^2}{2\sigma_2^2}) \propto \exp\frac{\left[x - \left(\frac{\sigma_2^2 d_1 + \sigma_1^2 d_2}{\sigma_1^2 + \sigma_2^2}\right)\right]^2}{2\sigma_1^2\sigma_2^2/(\sigma_1^2+\sigma_2^2)}$$

Cue Integration (3):Theory

- If we know mean estimate and variance for each modality in isolation, we can deduce mean of bimodal estimate:

$$\mu = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} d_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} d_2$$

pushed towards more reliable cue

- and discrimination threshold

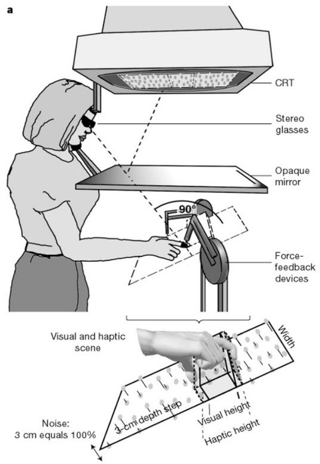
$$T_{1,2}^2 \propto \sigma_{1,2}^2 = \sigma_1^2\sigma_2^2/(\sigma_1^2+\sigma_2^2)$$

smaller than 1 or 2 alone

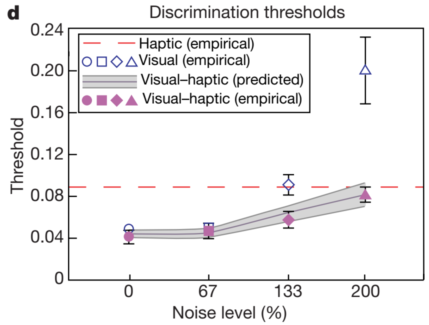
Cue Integration (4): Ernst and Banks, *Nature*, 2002

- visual + haptic cues
- vary noise level / visual cue
- compute discrimination threshold for each cue alone, or when both are present.

$$T^2_{1,2} \propto \sigma^2_{1,2} = \sigma^2_1\sigma^2_2/(\sigma^2_1+\sigma^2_2)$$



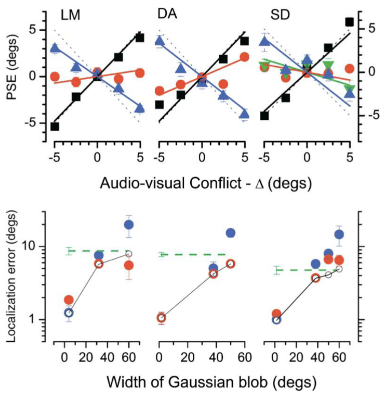
Cue Integration (5): Ernst and Banks, *Nature*, 2002



- height judgment follows optimal integration of visual and haptic cues.
- 'visual capture' for low visual noise, 'haptic capture' for high visual noise
- instantaneous 'switch'
- numerous studies replicate this result in a variety of paradigms (e.g. Alais & Burr, 2004).

Cue Integration (6):Ventriloquist effect

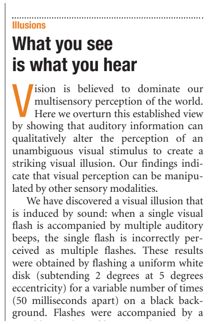
- Alais & Burr, *Curr Biol*, 2004
- visual blob of various size + auditive 'click', possibly in conflict.
- measure both estimate of position (mean), and discrimination threshold
- near optimal integration
- visual capture for small blobs
- auditive capture for large blobs



---: auditory alone
o : visual alone
o : prediction

Cue Integration (7)

- capture of vision by sound
- Shams et al, *Nature*, 2000.

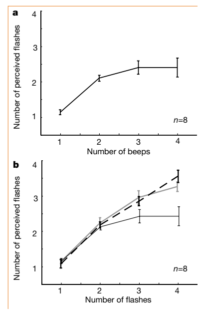


Illusions

What you see is what you hear

Vision is believed to dominate our multisensory perception of the world. Here we overturn this established view by showing that auditory information can qualitatively alter the perception of an unambiguous visual stimulus to create a striking visual illusion. Our findings indicate that visual perception can be manipulated by other sensory modalities.

We have discovered a visual illusion that is induced by sound: when a single visual flash is accompanied by multiple auditory beeps, the single flash is incorrectly perceived as multiple flashes. These results were obtained by flashing a uniform white disk (subtending 2 degrees at 5 degrees eccentricity) for a variable number of times (50 milliseconds apart) on a black background. Flashes were accompanied by a



http://shamslab.psych.ucla.edu/demos/

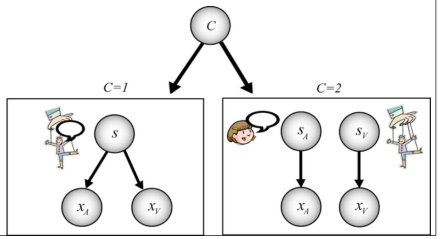
Cue Integration (8): when not to integrate?

• if spatial disparity between the 2 cues is too large: integration is not appropriate anymore --> segmentation.

• problem = not only to infer source location of 2 sensory signals but also whether the signals have a common cause (C)

• Körding et al 2007 : ideal-observer model that infers whether 2 sensory cues originate from same location and also estimates their location(s) accurately predicts nonlinear integration of cues in 2 auditory-visual localization tasks.

$$p(C|x_V,x_A)=\frac{p(x_V,x_A|C)p(C)}{p(x_V,x_A)}$$



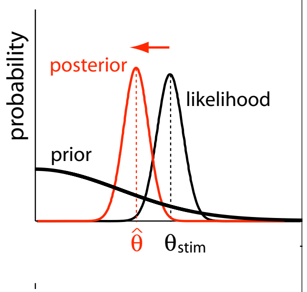
2 - Do brains form a representation of the past statistics of the environment (priors) and combine it optimally with current information?

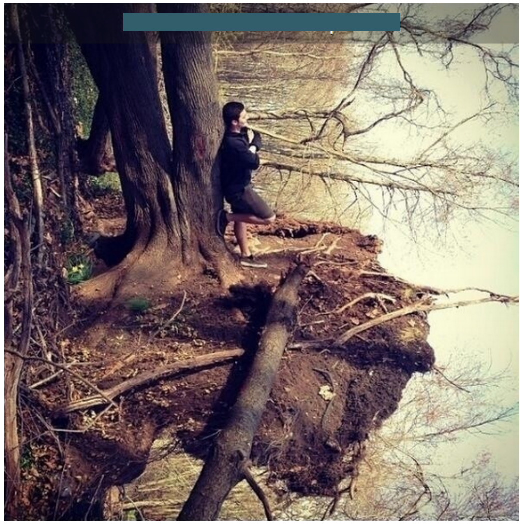
A Bayesian theory of the Brain: Priors

• How is the brain making use of previous knowledge? what priors?

• Prediction 1: the more uncertain the data, the more prior information should influence the interpretation.

• Prediction 2: The priors should reflect the statistics of the sensory world (on which time-scale?).



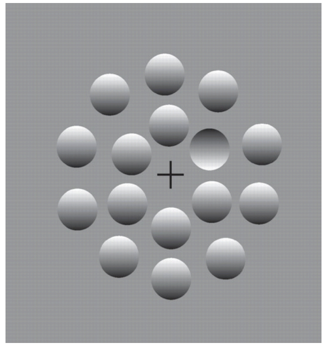


Long-term “structural” priors

Visual illusions : insight into what sort of assumptions the visual system makes.

- Light comes from above
- Cardinal orientations are more frequent [Gershick et al 2011]
- smoothness [Geisler et al 2001]
- symmetry [Knill 2007]
- Objects don’t move or only slowly

[Weiss et al 2001; stocker & Simoncelli 2006]

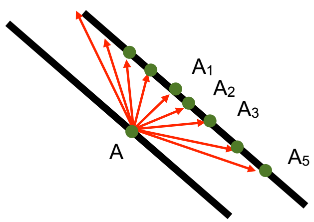


... recently formalized in Bayesian terms

[T. Adelson, E. Simoncelli, O. Schwartz, Y. Weiss]

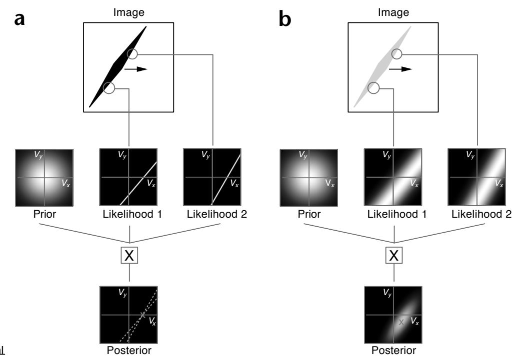
Interpreting motion : A Prior on Low Speeds (1)

- Motion shown in an aperture is fundamentally ambiguous; it can be interpreted in an infinite number of ways
- which one is chosen? why?



Interpreting motion : A Prior on Low Speeds (2)

- Hypothesis: humans tend to favour slower motions
- Use a (gaussian) prior on low speeds (centred at 0).
- Explain great variety of data -- elegant unifying explanation

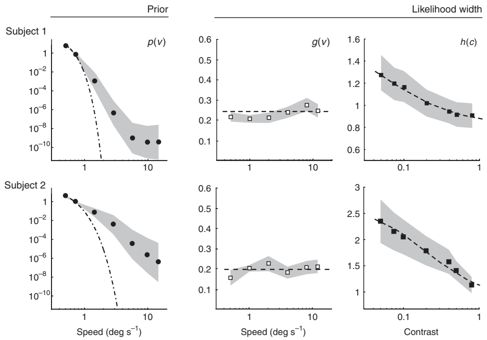


Weiss, Adelson & Simoncelli, *Nat Neuro*, 2002

http://www.cs.huji.ac.il/~yweiss/Rhombus/rhombus.html

Can we measure people's prior experimentally?

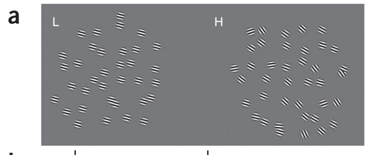
- Method: reverse engineer the shape of the prior from perceptual data
- 2AFC speed discrimination task at different contrast levels -- measure both bias and variability --> recover prior and likelihood



Stocker & Simoncelli, Nat Neuro, 2006

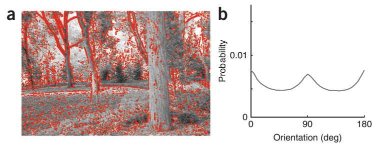
Do such priors correspond to the environment statistics? Cardinal Directions.

- Girshick and Simoncelli, Nat Neuro 2010.



is L stimulus CW or CCW compared to H?

- Orientation judgments are more accurate at cardinal (horizontal and vertical) orientations.
- Biased toward cardinal orientations.
- Prior towards cardinal orientation match orientation distribution measured in photographs.



Learning of priors:
Are we building up new priors constantly?

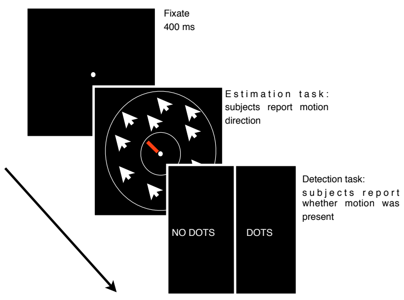
Are priors learned or innate? Do people form new priors for everything? how fast?

[Chalk, Seitz and Seriès, JOV 2010]

Do people form new priors for everything? How fast?

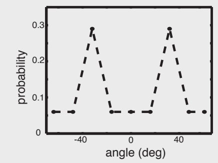
Behavioural Task

- On each trial, participants were presented with either a low contrast random dot motion stimulus (100% coherence) or a blank screen.
- Participants reported direction of motion (estimation), before reporting whether a stimulus was present (detection).



- Two motion directions were presented in a larger number of trials than other directions.

Stimulus distribution

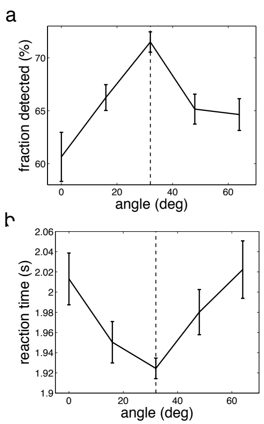


Questions

1. Are participants going to learn implicitly which directions are most likely to be presented?
2. How would these learned expectations bias their perception of subsequently presented motion stimuli?

Result 1/3: Detection is better and faster for the expected directions

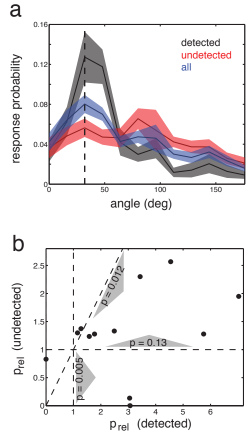
- Detection performance was best for most frequently presented directions
- Reaction times were shorter
- Similar to the effects of selective attention (Posner et al. 1980) - suggesting that subjects were attending to expected directions.
- Knowledge about the statistics of the stimulus was however not conscious.



Result 2/3: Participants 'hallucinate' motion in expected directions

- On trials where no stimulus was presented, but where participants reported seeing a stimulus (in detection task), they were strongly biased to report motion in two most frequently presented directions.
- Did not occur on trials where participants did not report seeing a stimulus, arguing against a 'response bias' explanation.
- This effect was fast to develop, occurring in less than 200 trials / few minutes.

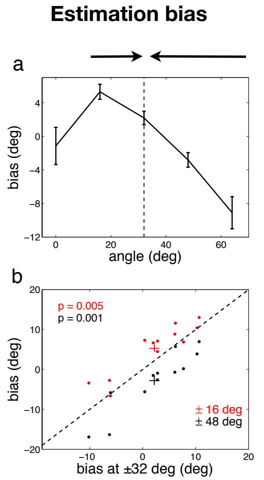
Distribution of estimates when no stimulus displayed



Result 3/3: Expectations bias perception of motion direction

- Estimates of motion direction were biased towards most frequently presented directions:

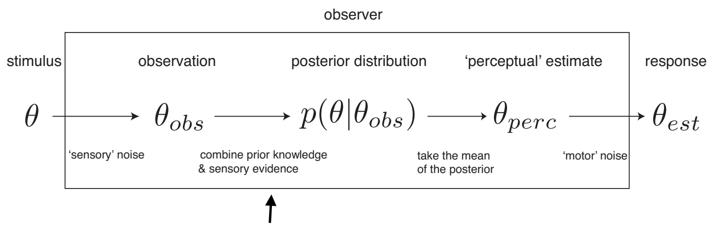
  subjects perceive motion direction to be more similar to expected direction than it really is.



[Chalk, Seitz and Seriès, JOV

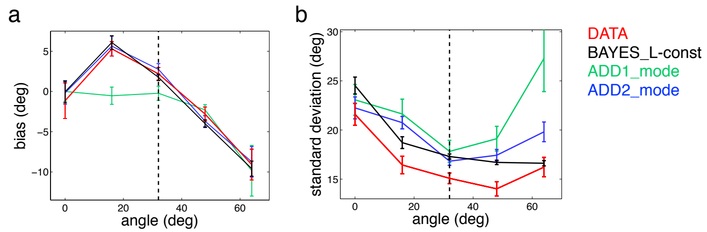
Modelling the observed estimation biases

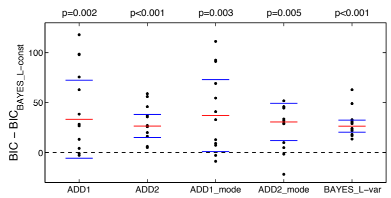
- subjects learn an expected distribution of the stimuli (prior) and combine it with sensory evidence
- 4 free parameters: center and width of prior, width of likelihood, fraction of 'random' trials + motor noise (fixed with high contrast trials)



$$p_{exp}(\theta) = \frac{1}{2}[V(-\theta_{exp}, \kappa_{exp}) + V(\theta_{exp}, \kappa_{exp})]$$

Model comparison





$BIC = -2 \cdot \ln(L) + k \cdot \ln(n)$

- Bayesian model describes the data better than response strategy models.

[Chalk, Seitz and Seriès, JOV 2010]

Conclusions

- Participants rapidly learn multimodal stimulus expectations (< 200 trials).
- These expectations bias their perception of simple motion stimuli, causing them to 'hallucinate' motion in the expected direction, and perceive motion stimuli as closer to the expected directions than they actually are.
- The biases we observed can be explained assuming that participants combine a 'learned prior' about the stimulus statistics with their sensory evidence in a probabilistically optimal way.
- A number of open questions (specificity of prior, time scale, neural implementation - substrate of expectation)
- in particular: can one learn any prior like this ? or are some priors fixed?