Perception statistique de séquences aléatoires

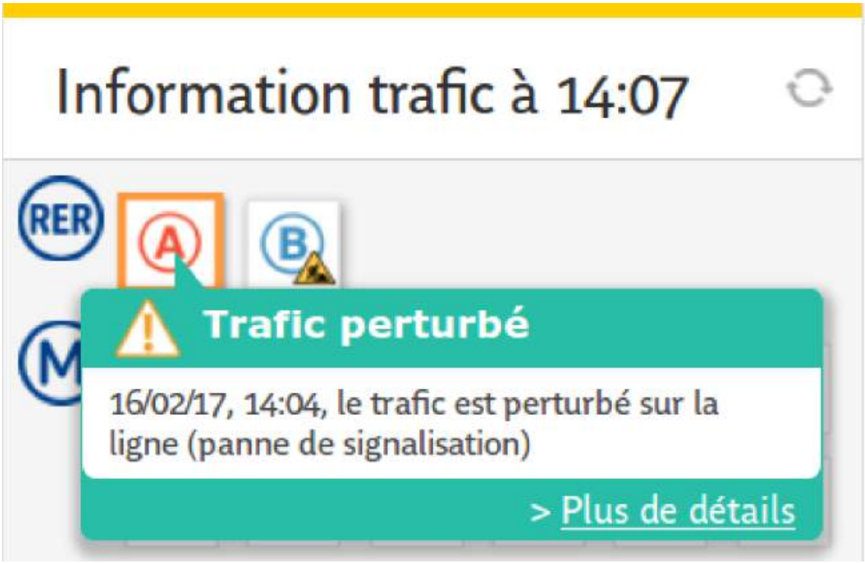
Florent MEYNIEL

Neurospin, CEA, France

The observations that we receive from the world are stochastic and unfold across time

The Old Faithful geyser erupts every 45 – 125 min.

- Events in our world are fraught with uncertainty.
  - Because we don't know all the hidden causes.
  - Because those hidden mechanisms are intractable.
  - Because there is inherent uncertainty.



RER trains are often late.

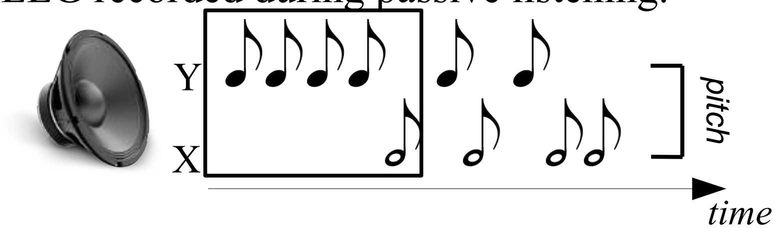
- Events nevertheless often show some regularity.
- Detecting those regularities can be advantageous to adapt our behavior.

Our world is not only stochastic, it is also changing, making prediction difficult

- Unexpected and sudden changes can occur, making previous estimates no longer informative.

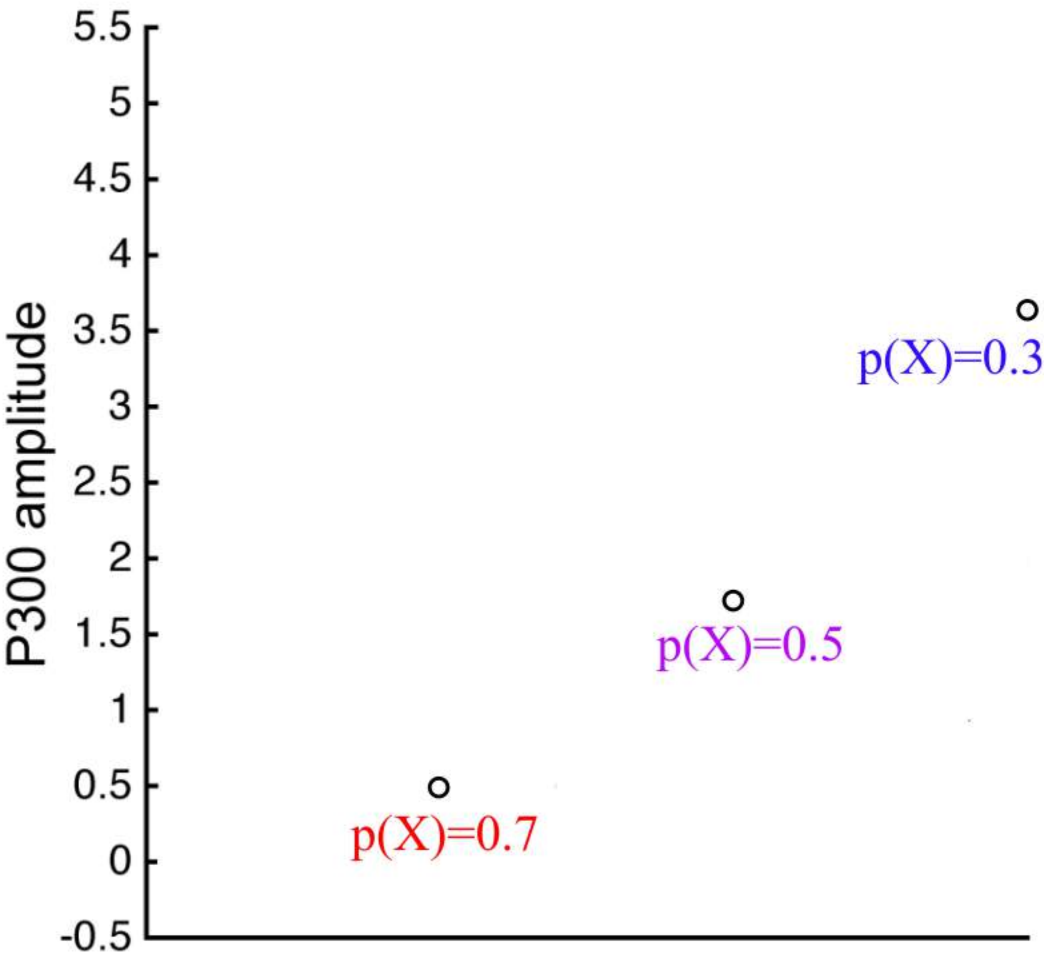
A seminal experiment by Squires et al. 1976 suggests that the brain constantly tracks the statistics of stimuli

EEG recorded during passive listening.



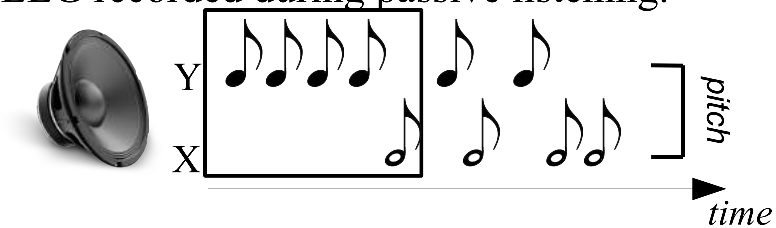
The P300 amplitude relates to the improbability of the current sound given the previous ones, suggesting a tracking of:

→ *The global item frequency*
in the entire sequence.
e.g. $p(\mathrm{X}) = 0.7$ vs. $p(\mathrm{X}) = 0.3$.



A seminal experiment by Squires et al. 1976 suggests that the brain constantly tracks the statistics of stimuli

EEG recorded during passive listening.



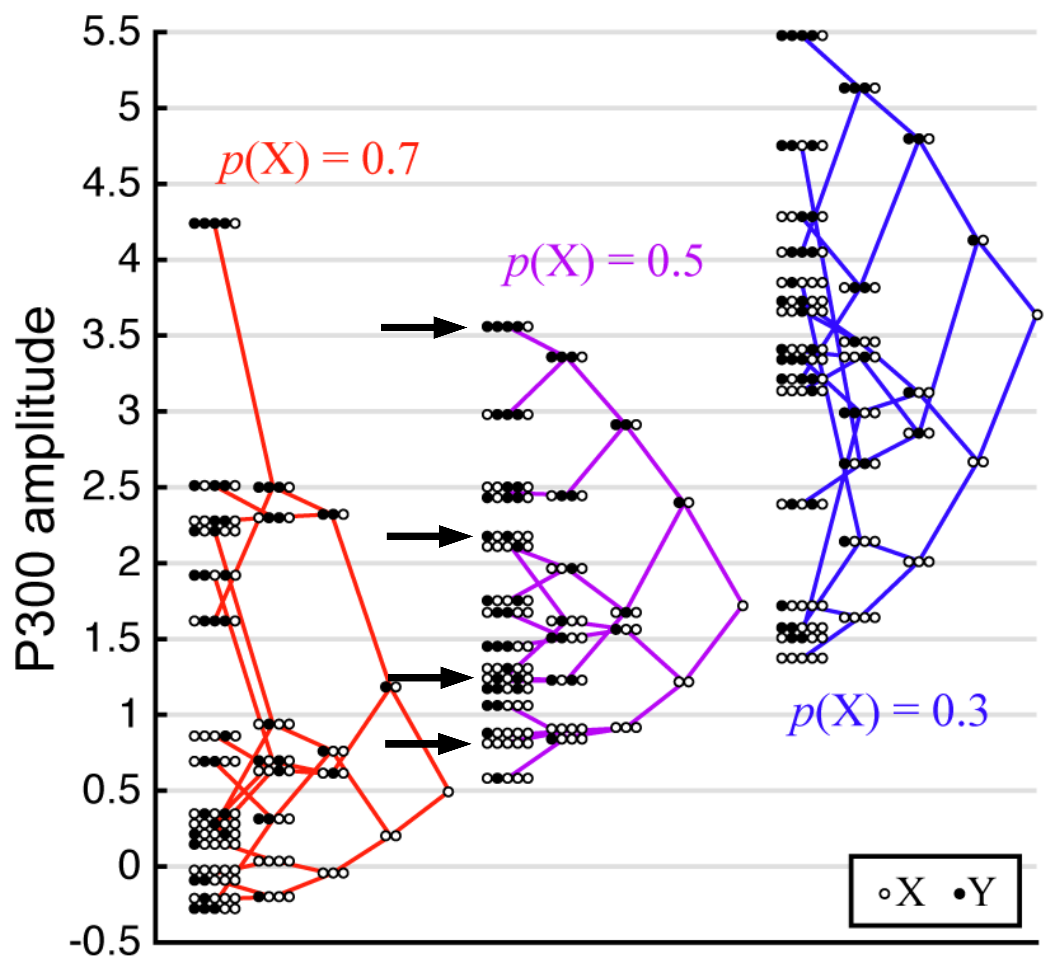
The P300 amplitude relates to the improbability of the current sound given the previous ones, suggesting a tracking of:

→ *The global item frequency*
in the entire sequence.
e.g. p(X) = 0.7 vs. p(X) = 0.3.

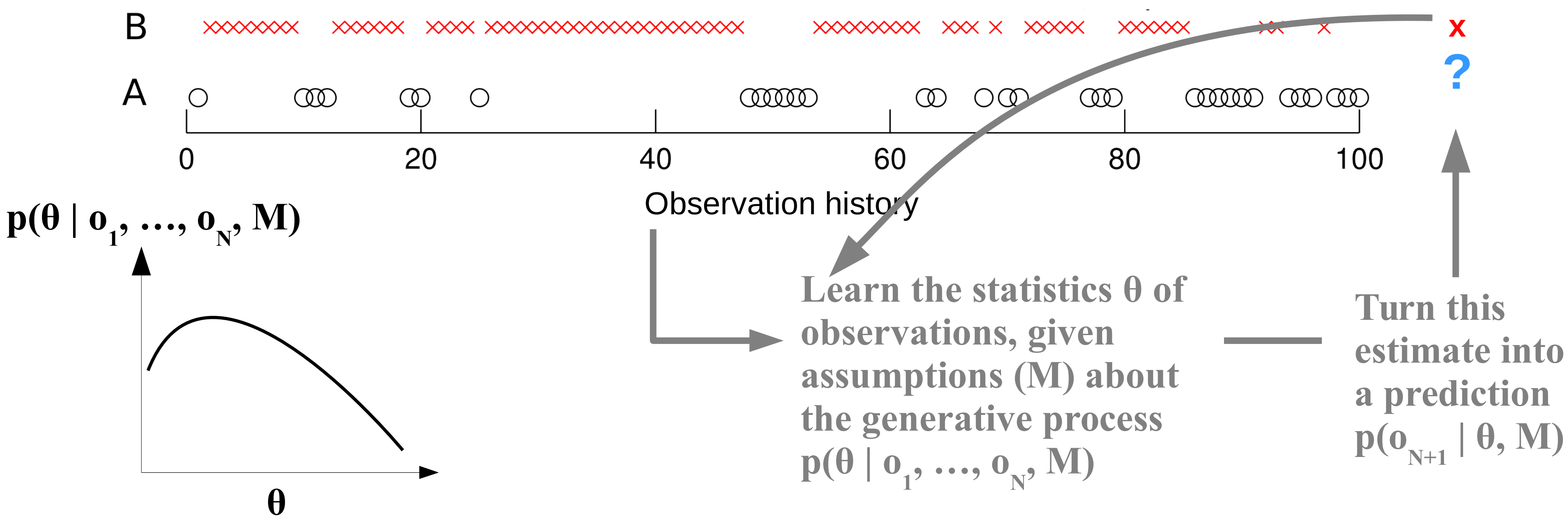
→ *The local item frequency*
in the recent history.
e.g. XXXXX vs. YYYYX.

→ *The local alternation frequency*
Whether items were repeated in the recent history
e.g. YYXXX vs. YXYXX.

Replication and further computational refinements: *Mars 2008, Kolossa 2013, Lieder 2013, Maheu 2017...*



Learning and predicting with Bayesian inference

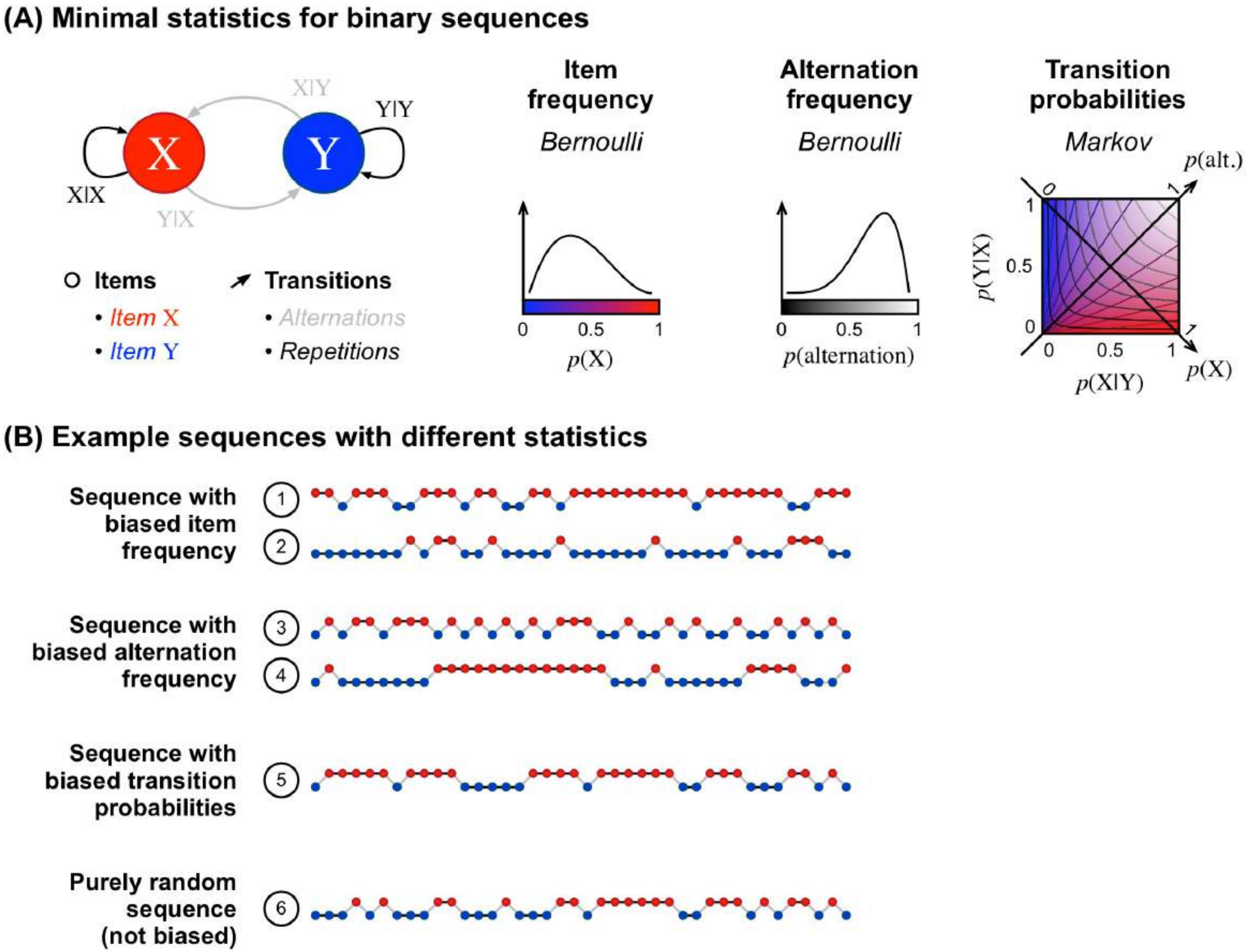


- The Bayesien inference computes with **conditional probabilities** p(- | -).
- Bayesian inference provides an **optimal prediction about future** observations given the previous ones, and given particular assumptions about the generative process, which is called an *ideal observer model*.
- The inference proceeds with **iterations by summarizing observations** (summary statistics or full distributions) *(Gelman, Bishop, Sutton & Barto)*. Tracking improbable (i.e. surprising) events allows a Bayesien learner to revise its estimates *(Friston 2005)*.

Learning and predicting with Bayesian inference

- To apply the Bayesian framework to brain processes, one must identify the properties of the inference:
  - **What is learned**? Which statistics are computed?
  - **Which observations** inform the current estimate?

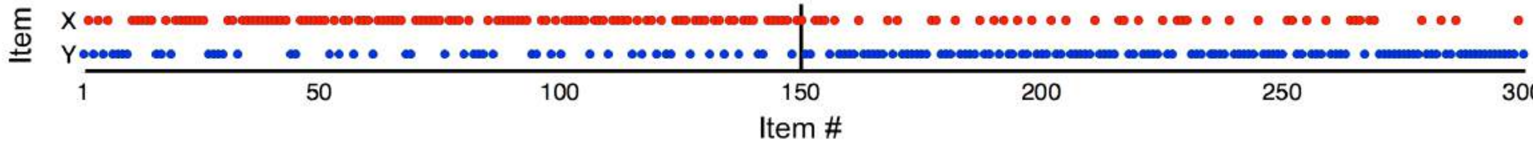
What has been learned? Sequences can be characterized by a hierarchy of statistics



More complex statistics and even rules could be envisaged *(Dehaene, Meyniel et al Neuron 2015)*, but here we start simple.

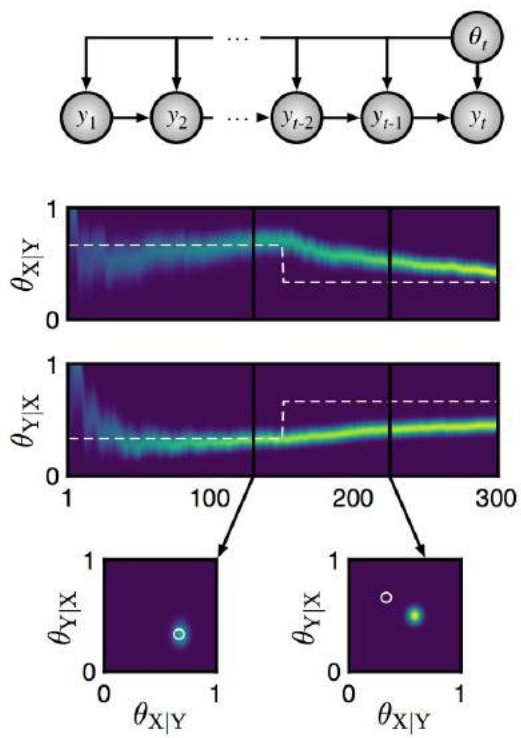
Meyniel, Maheu & Dehaene, *Plos Computational Biology* 2016

Which observations inform the current estimate? Different inference styles

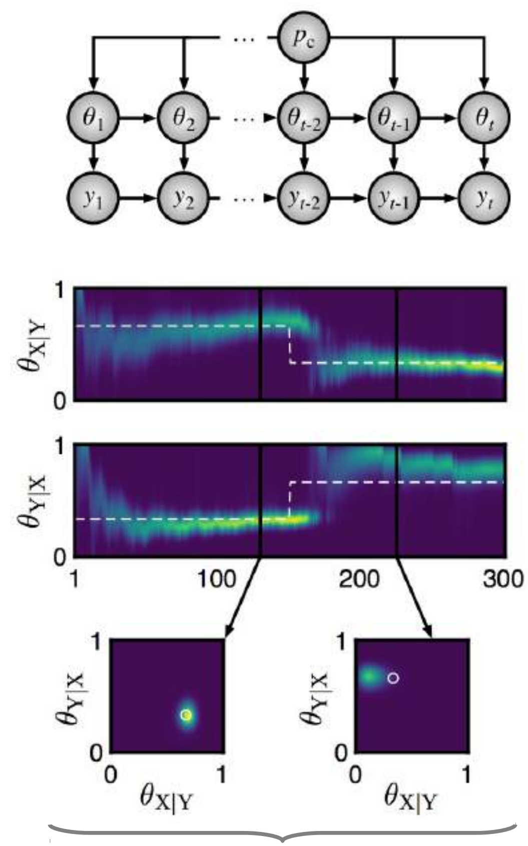


Models differ in their assumptions about changes in the generative process and they weight observations differently

Fixed belief model
With a perfect integration



Dynamic belief model
With a perfect integration



Local integration

Meyniel, Maheu & Dehaene, *Plos Computational Biology* 2016

Our proposal:
The brain entertains at a minimum a "local transition probability model"

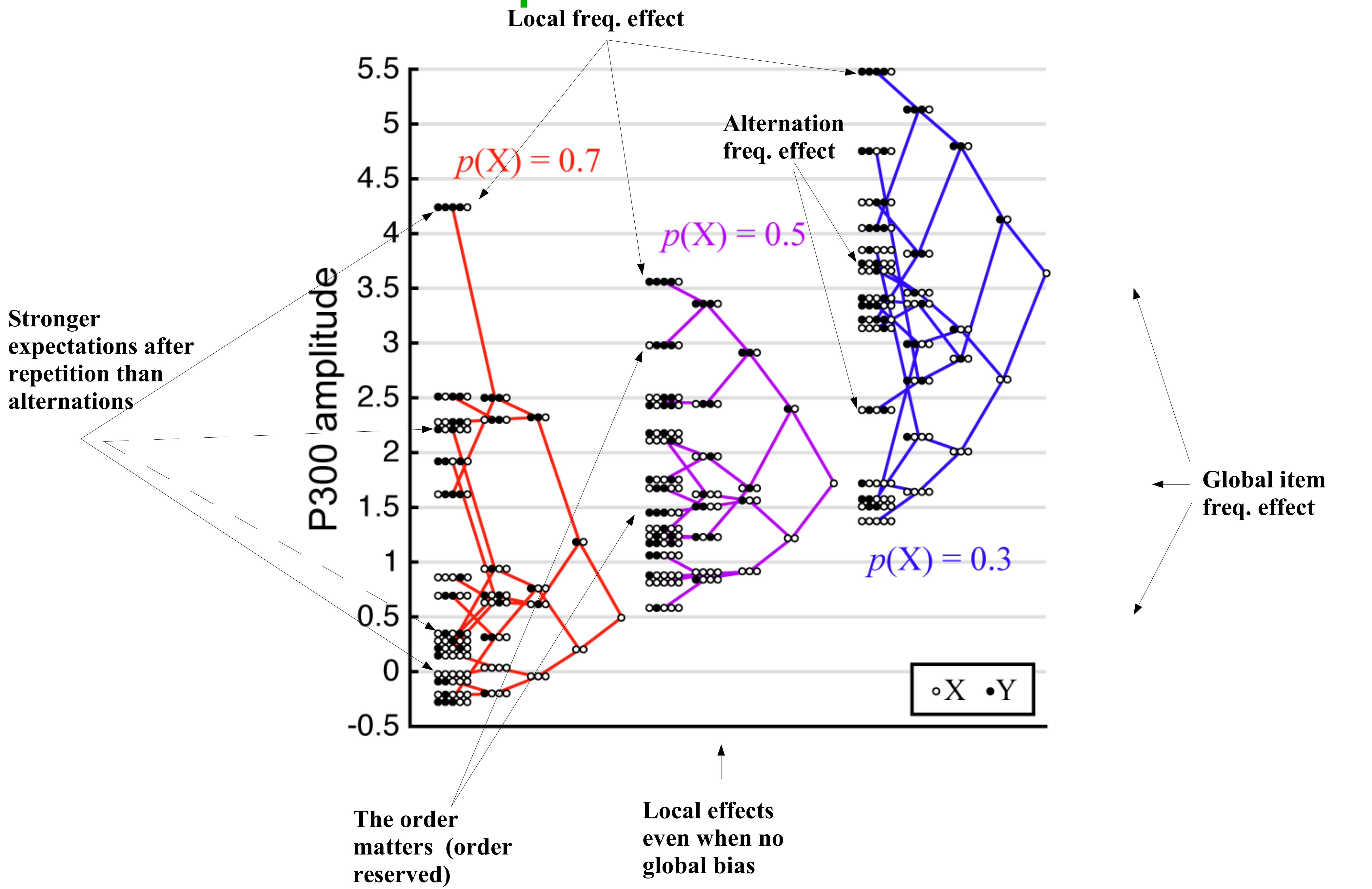
The brain constantly evaluates the **(un)likelihood** of the current observation based on:

- A tracking of **transition probabilities** between successive items
- A **local integration**: estimates are constantly revised based on the most recent observations.

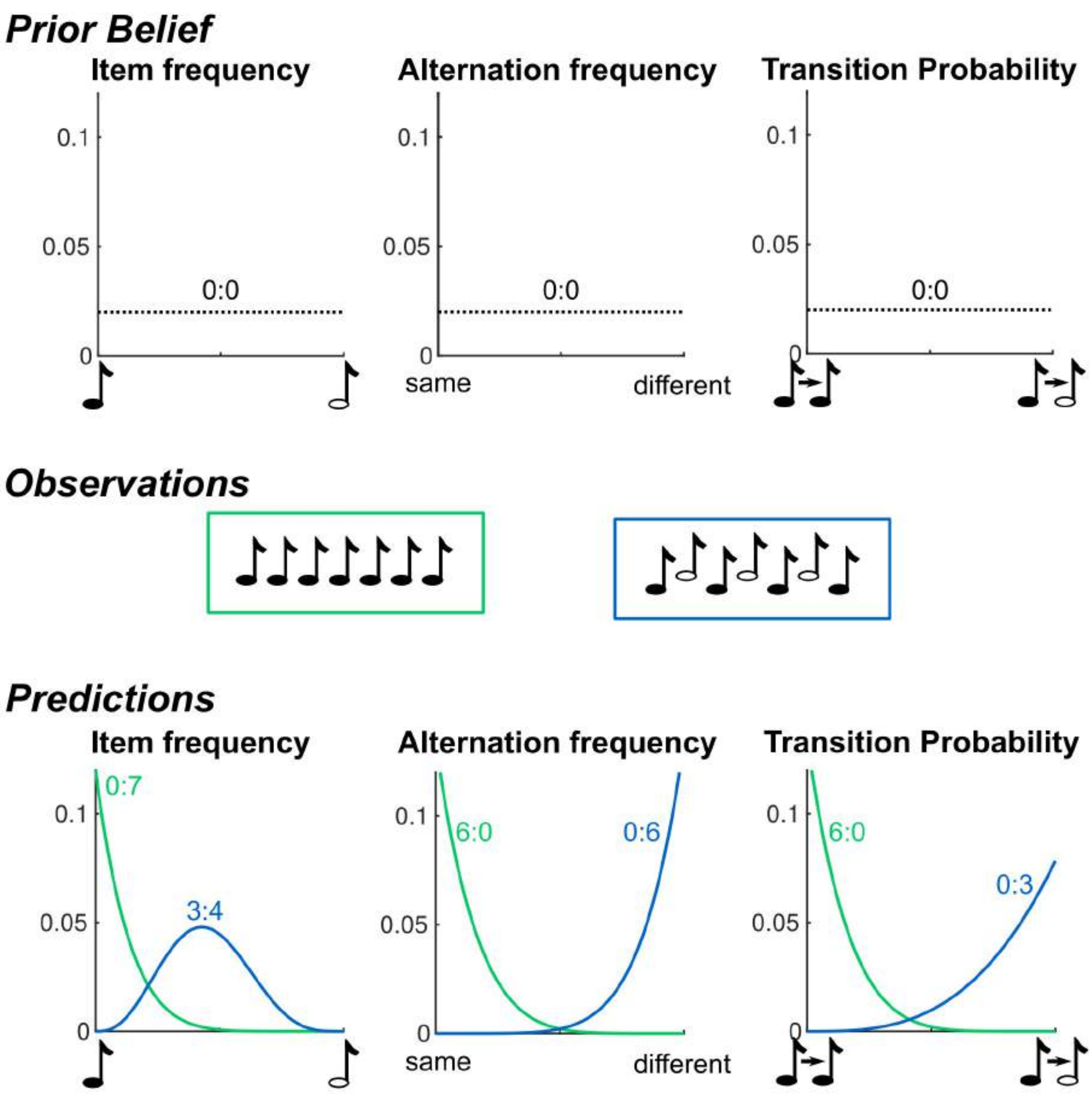
The mathematical notion of surprise quantifies the deviation from expectation *(Shannon 1948)*:

$$surprise = -log_2(P(actual\ observation))$$

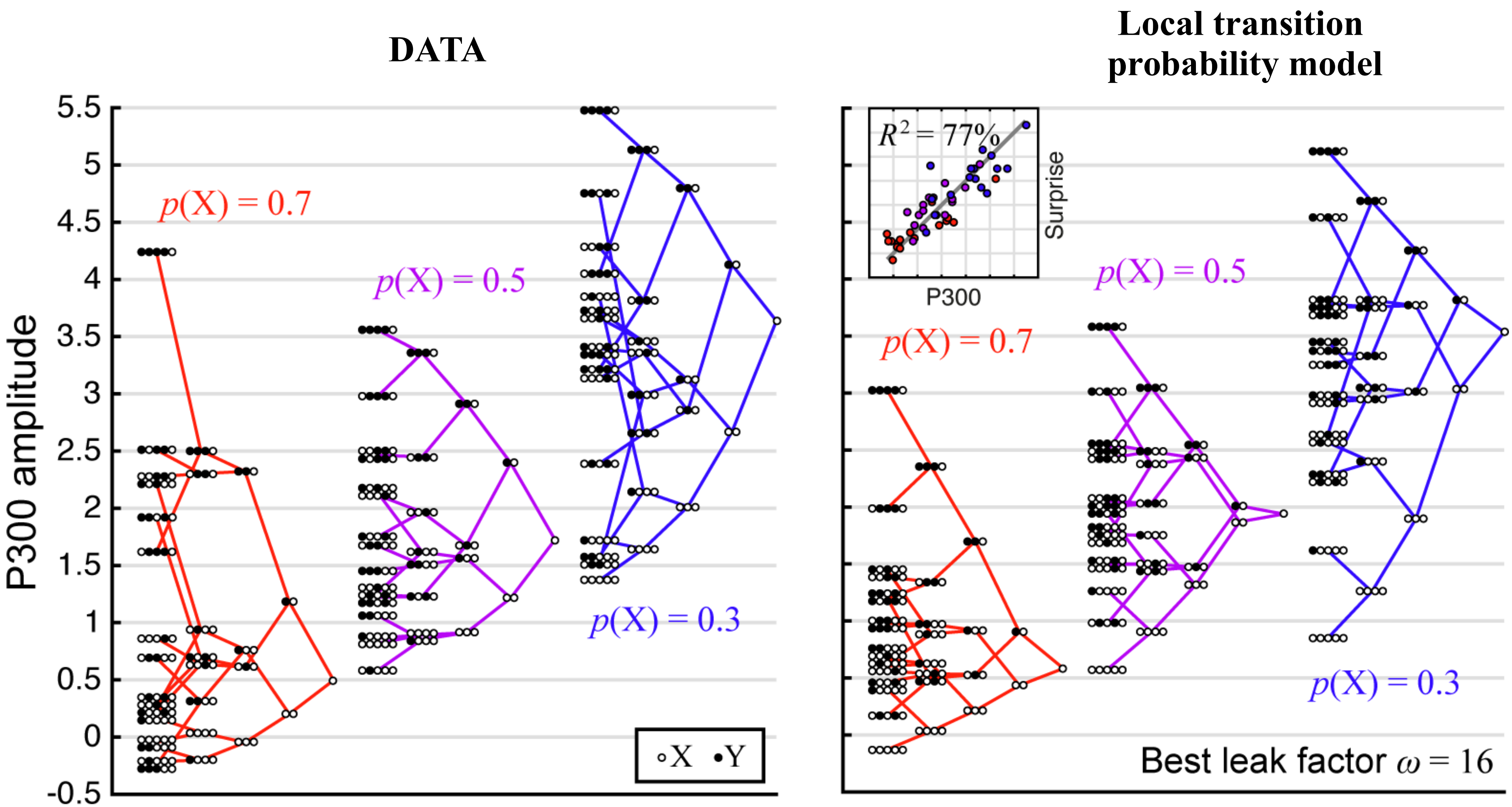
A qualitative agreement with the P300 data by Squires et al. 1976



Expectations emerge more rapidly from repetitions than alternations

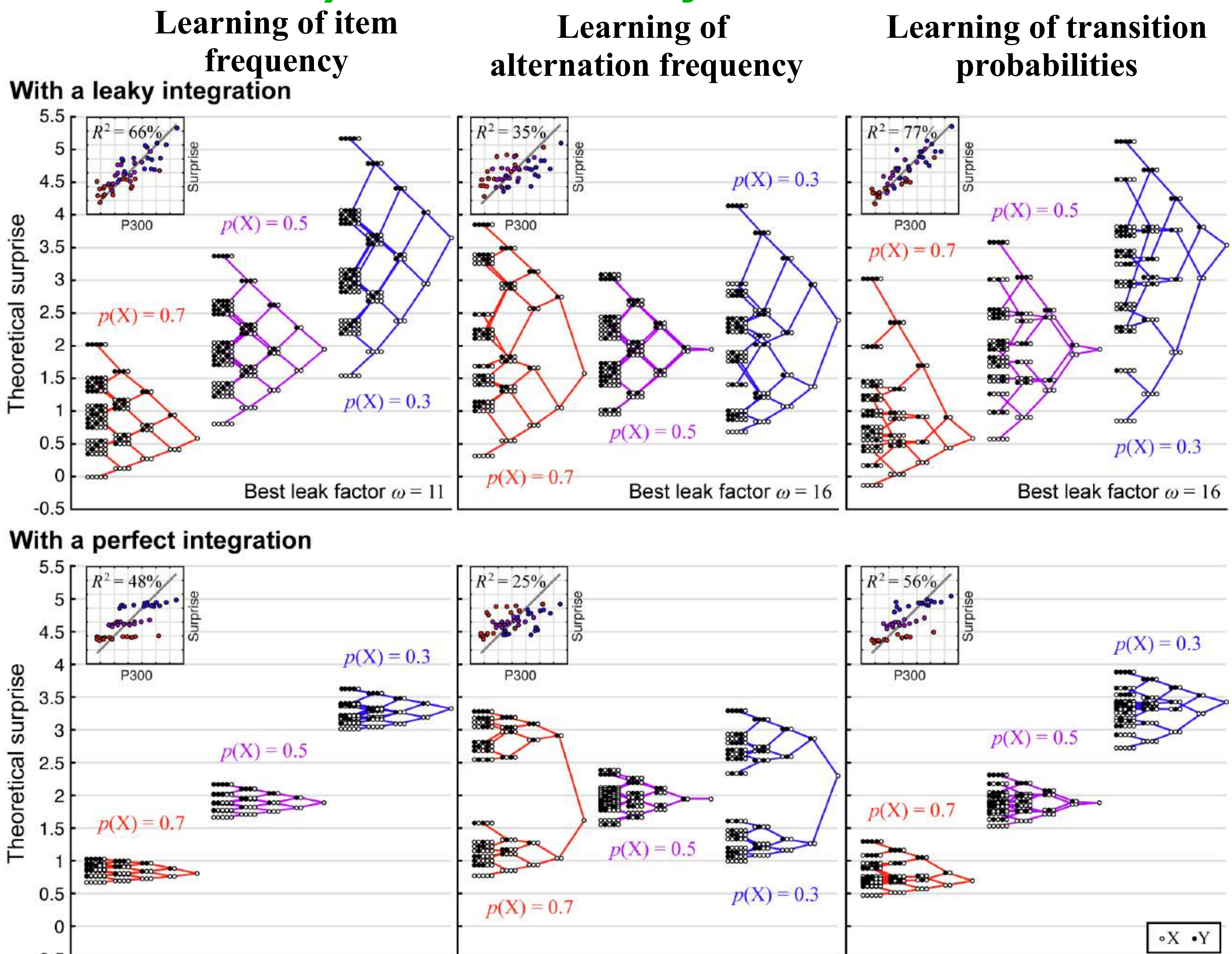


A quantitative agreement with the P300 data by Squires et al. 1976



Meyniel, Maheu & Dehaene, *Plos Computational Biology* 2016

The assumptions of our model (transition probability + local estimate) are necessary to account for the data

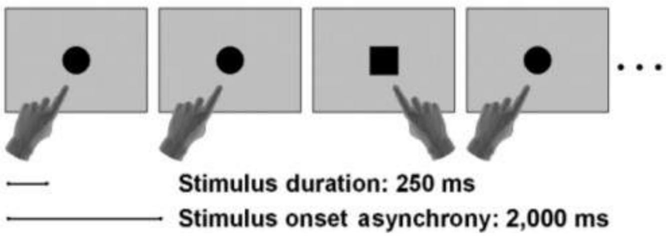


Meyniel, Maheu & Dehaene, *Plos Computational Biology* 2016

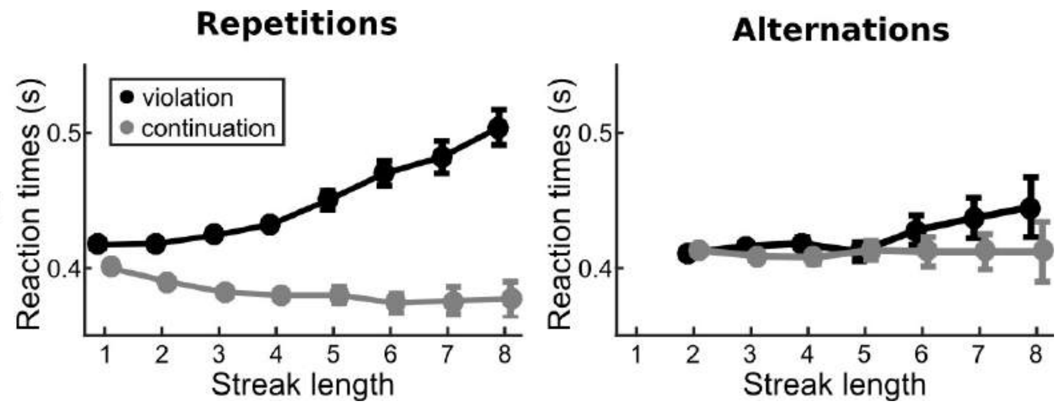
The local transition probability model accounts for classic “sequential effects” in reaction times

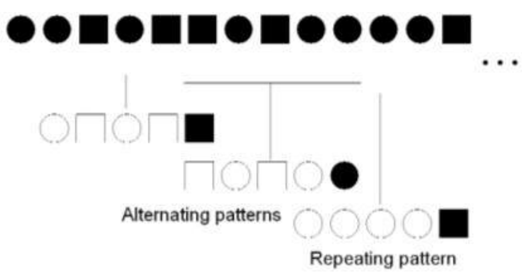
A typical reaction time task: Huettel et al., 2002

(predecessors: Hyman 1953, Bertelson 1961; Kirby 1976; Soetens, Boer & Hueting 1985, Sommer, Leuthold & Soe... 1999; Cho & Cohen 2002; ...)

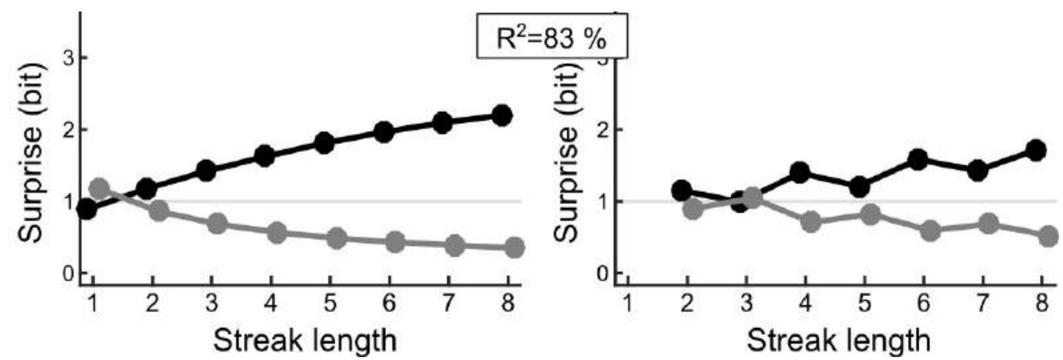


Experimental data





Simulation with a learning of transition probabilities

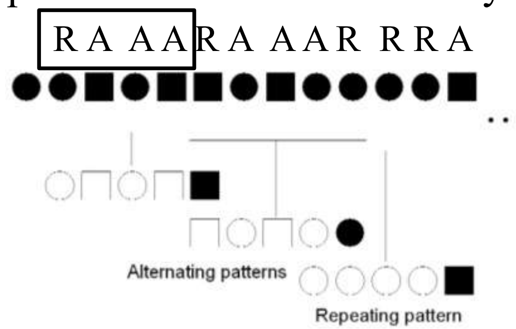


→ Our model reproduces the **gradual build up of expectations** with increasing streak length.
→ It also reproduces the **asymmetry between repetitions and alternations**.
→ The learning of stimulus frequency entails no expectation about alternations.
→ Learning the frequency of alternations (or of repetitions) is symmetric for repetitions and alternations.

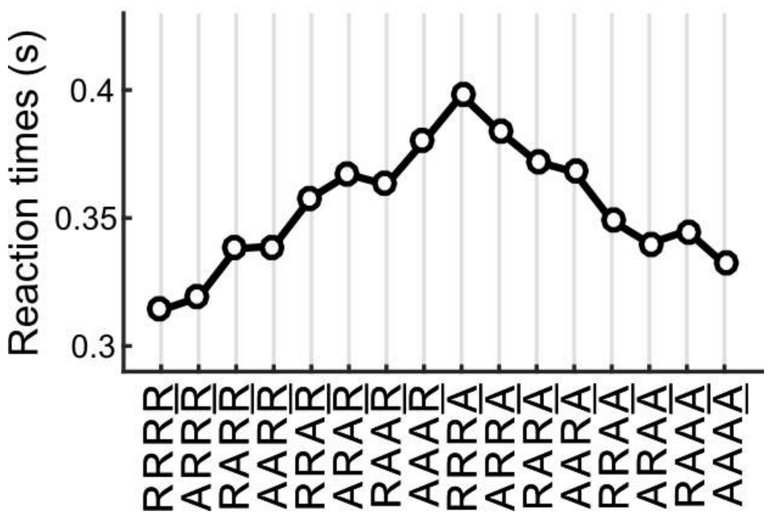
Meyniel, Maheu & Dehaene, *Plos Computational Biology* 2016

The local transition probability model accounts for classic "sequential effects" in reaction times

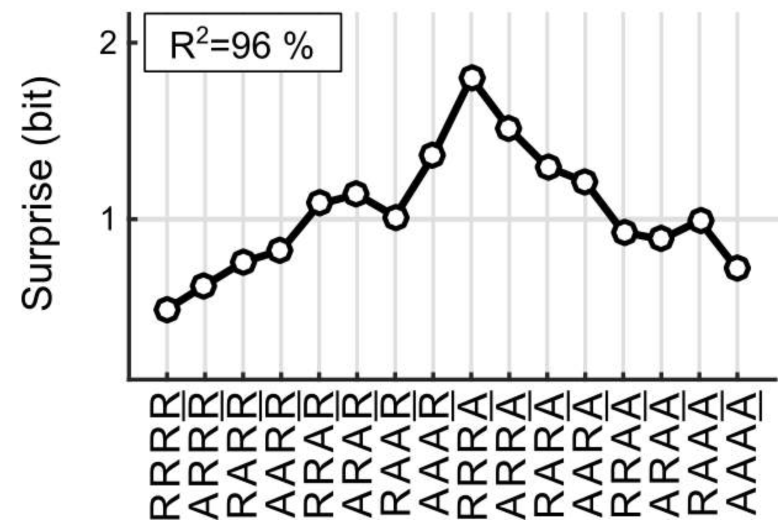
A systematic investigation of all pattern types in a reaction time task by *(Cho et al., 2002)*



Experimental data



Simulation, when learning transition probabilities



→ The local sequential effects survive even after exposure to long, fully unpredictable sequences
→ Our model captures **order effects**, e.g. ARRR < RARR < RRAR.
→ There is (again) an asymmetry between repetitions and alternations
→ The model captures subtle effects in the data, such as **local minima** for RAAR and ARAA
→ The inference based on the stimulus frequency fails to reproduce many aspects of the data.

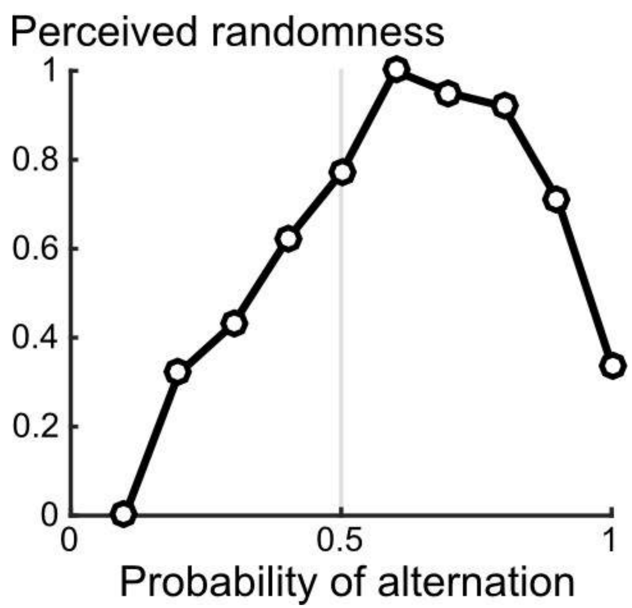
Meyniel, Maheu & Dehaene, *Plos Computational Biology* 2016

The local transition probability model accounts for the asymmetric perception of randomness

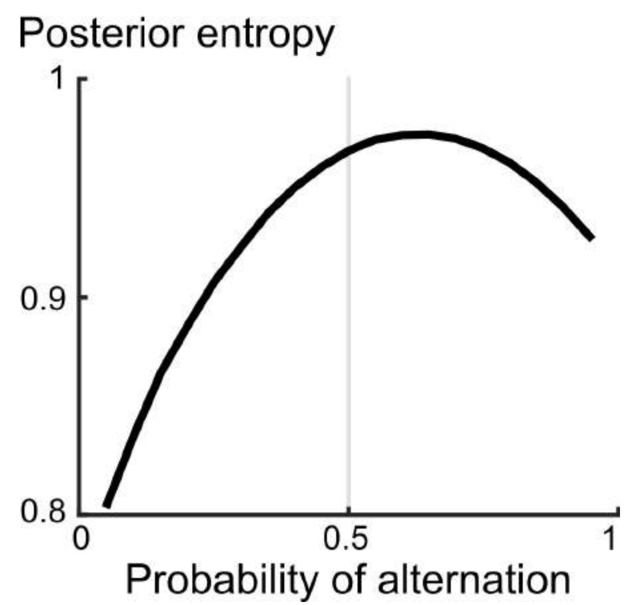
Rating of the perceived randomness of binary sequences. *(Falk, 1975)*

O O X O X X O X O X O O O O X X O X O O O → here, p(alternate) = 12/20

Experimental data
Falk (1975)



Simulation with a learning of transition probabilities



→ Studies of **perceived randomness show a bias for alternations**, max around 0.6. *(Falk, 1975; Falk & Konold, 1997; Bakan, 1960; Budescu, 1987; Rapoport & Budescy, 1992; Kareev, 1992)*

→ The perceived randomness can be formalized as a **posterior entropy**

→ Our model predict an asymmetry of the perceived entropy (that is all the stronger that the integration is local)

→ The asymmetry is specific of our model

Meyniel, Maheu & Dehaene, *Plos Computational Biology* 2016

INTERIM SUMMARY
A local estimation of transition probabilities constitutes a minimum model of human inference about sequences

The model accounts for:

→ Expectations in various types or measurements: brain signals, reaction times and reports of perceived randomness.

→ "Local effects" on expectations (local frequency, local transition probabilities, order of stimuli)

→ Global effects on expectations: the global frequency of stimuli.

→ The asymmetry between repetitions and alternations *(Yu & Cohen 2008 NIPS; Falk & Konold, 1997 Psych Rev)*

The model favors recent observations to form expectations

→ The local integration may not due to be a limitation of processing capabilities but rather of an assumption of non-stationary. *(Yu & Cohen 2008 NIPS; Behrens 2007 Nat Neuro; Meyniel et al 2015 PCB)*

→ Angela Yu's claim: the assumption of non-stationarity is the best default assumption.

The model is principled and parsimonious

→ It is relies on Bayesian inference.

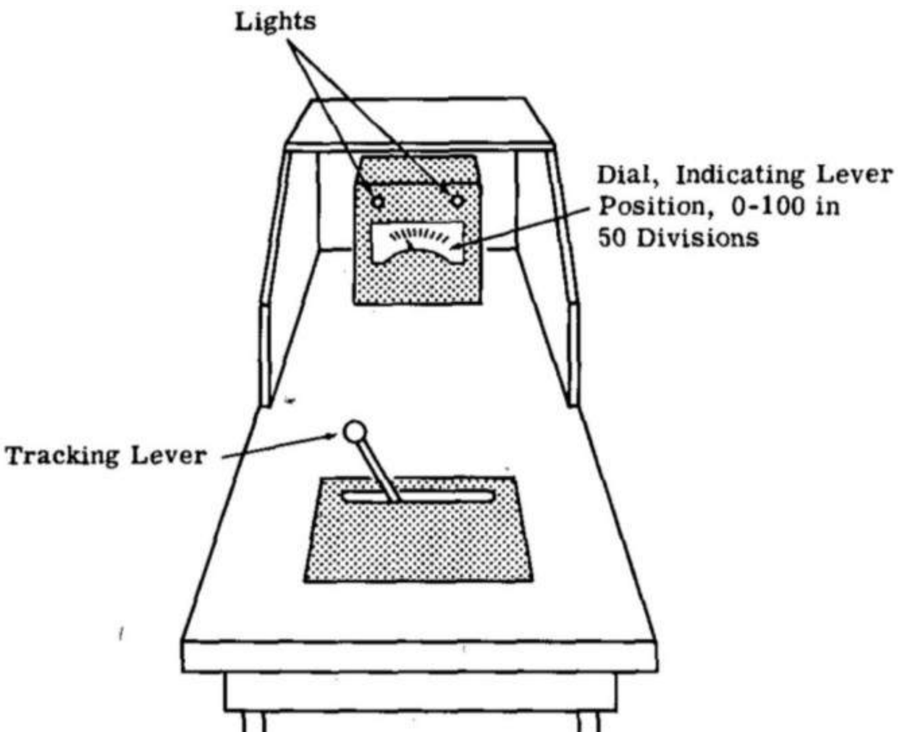
→ It learns transition probabilities, the first building block of sequence knowledge. *(Dehaene Meyniel et al. 2015 Neuron; Wacongne 2012 J Neuro; Strauss 2015 PNAS)*

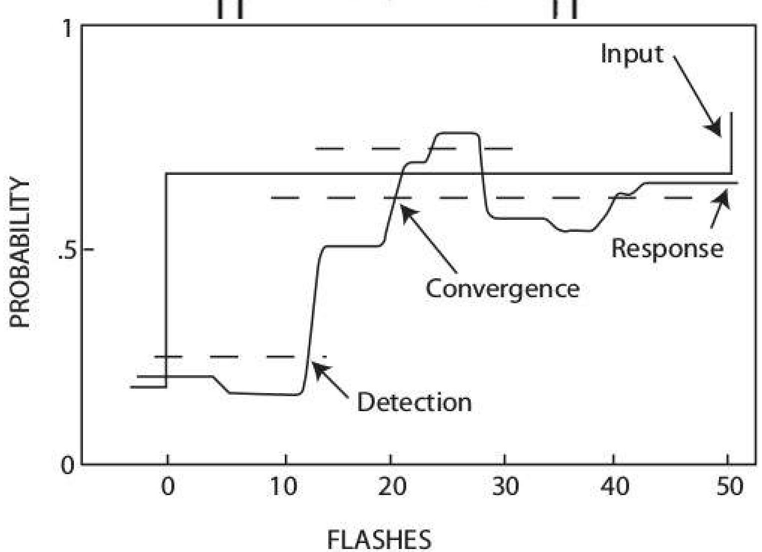
→ It does not need biased priors to account for the asymmetry between repetitions and alternations. *(Yu & Cohen 2008; Falk & Konold, 1997)*

→ It is simple, with only one free parameter controlling the "horizon" of the integration.

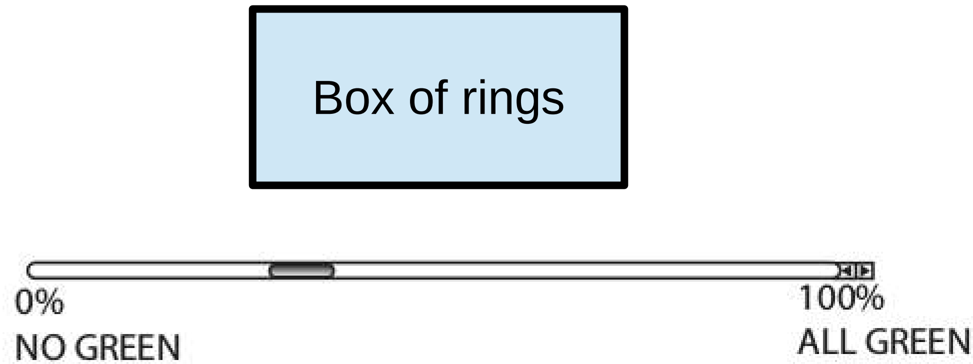
Can human subject report explicitly time-varying probabilities that they infer from a sequence of observation?

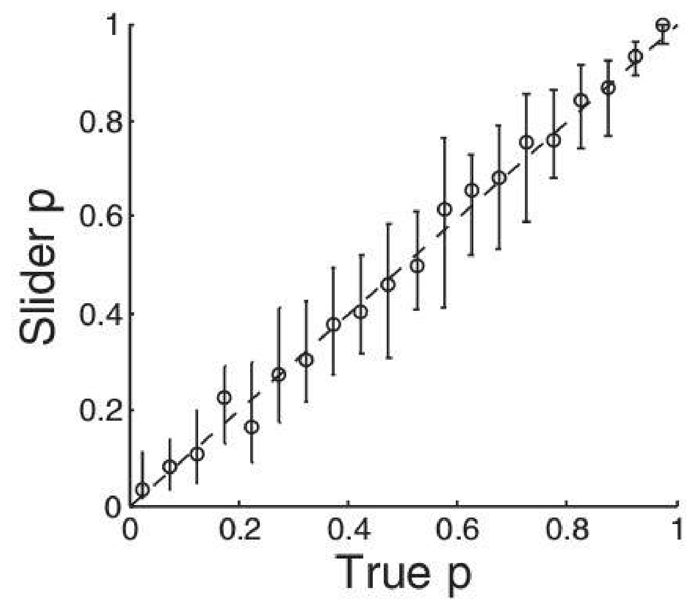
Robinson, *Ergonomics* 1964



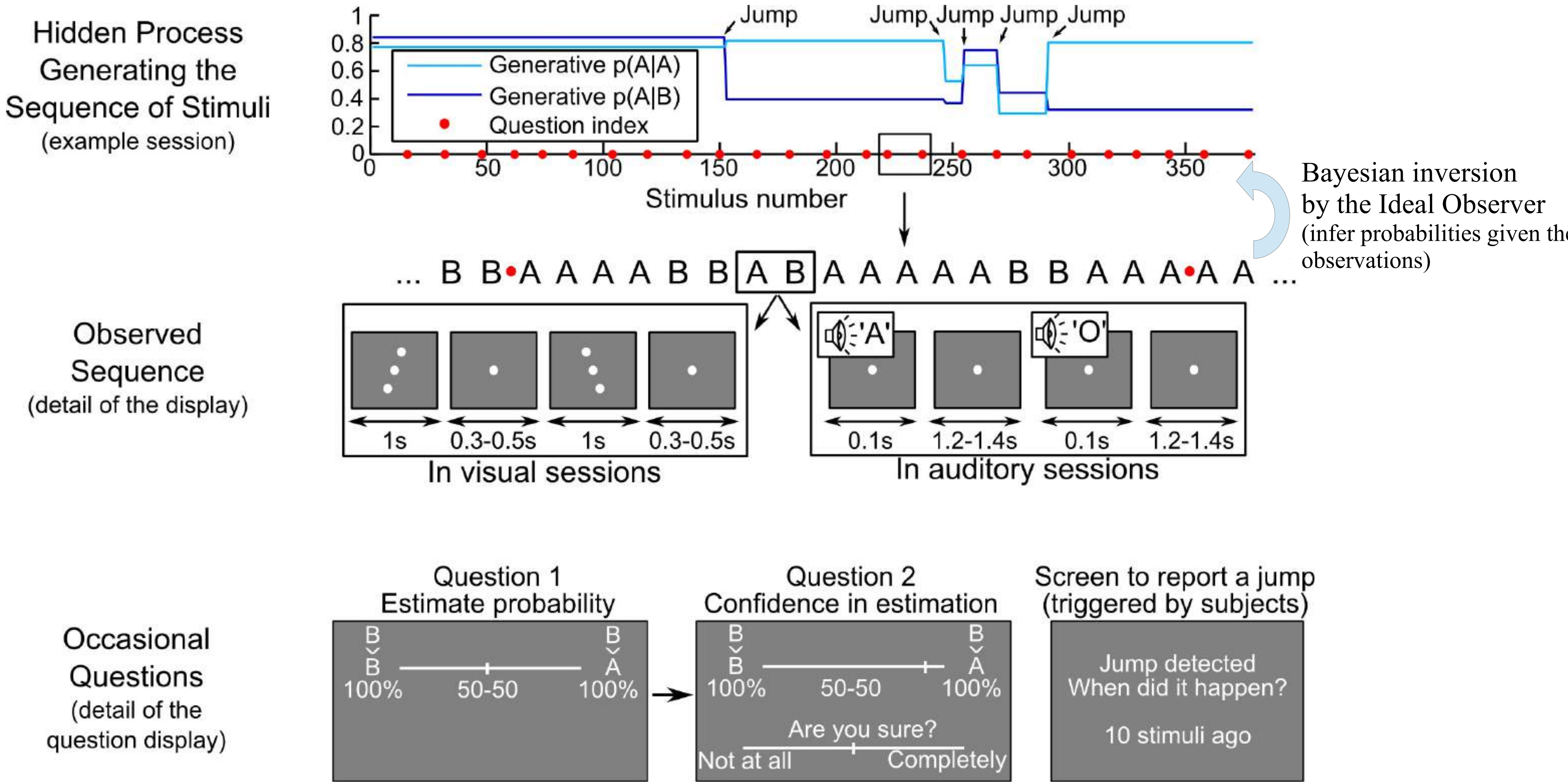


Gallistel et al, *Psych Rev* 2014



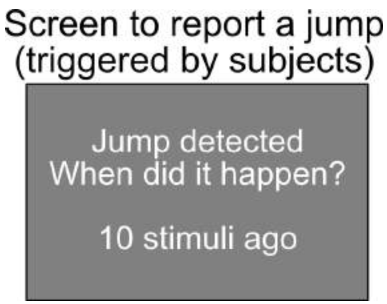


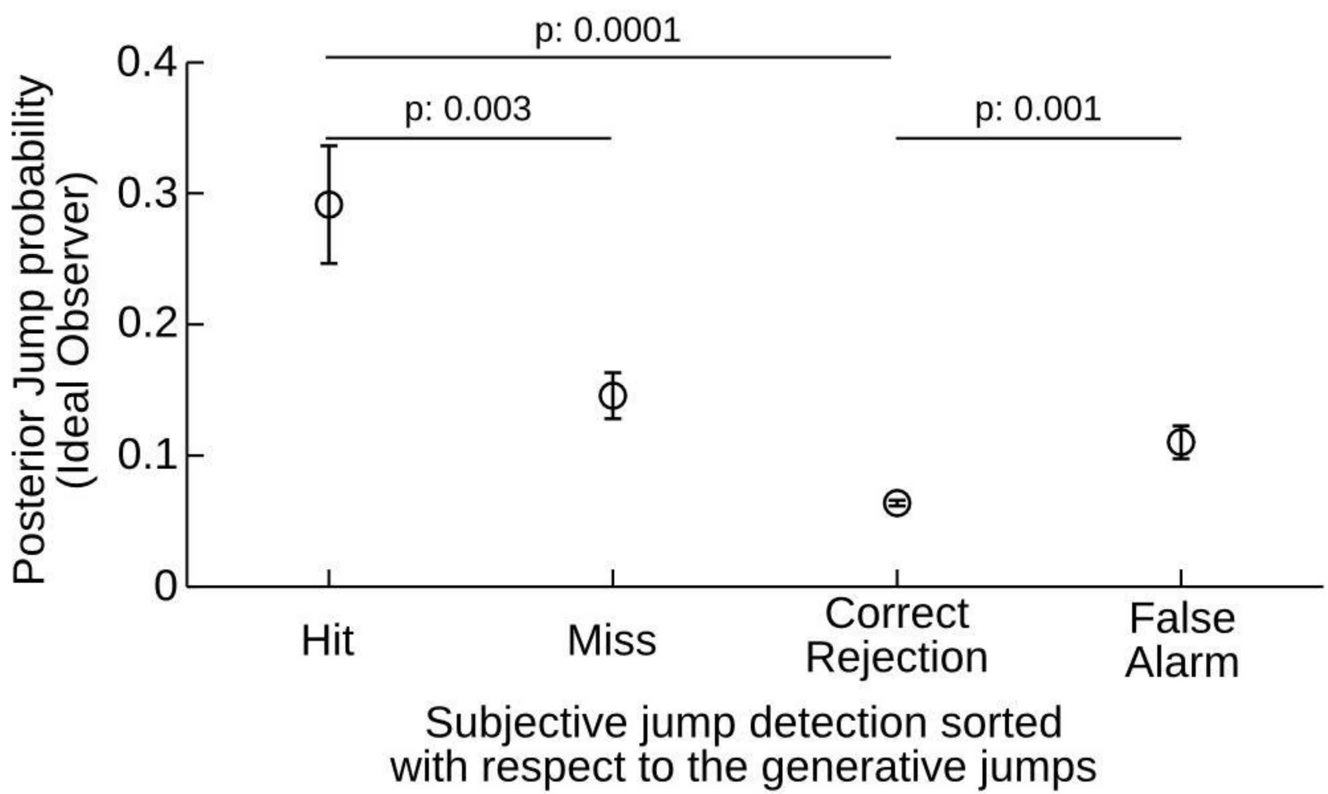
Can human subjects explicitly track time-varying transition probabilities?



Meyniel, Schlunneger & Dehaene, *Plos Computational Biology* 2015

Subjects accurately detect changes in the hidden transition probabilities

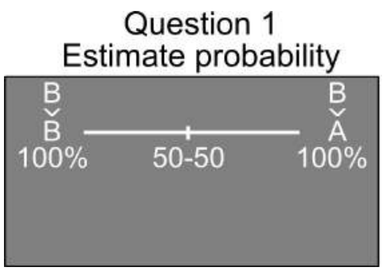


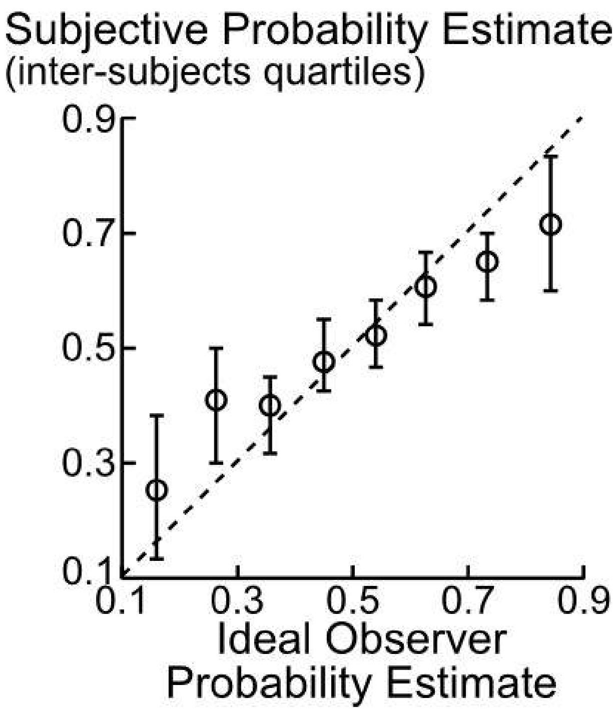


Subjects detected changes given the evidence provided by the observations presented to them.

Meyniel, Schlunneger & Dehaene, *Plos Computational Biology* 2015

Subjective estimates of probabilities are accurate, regardless of the sensory modality

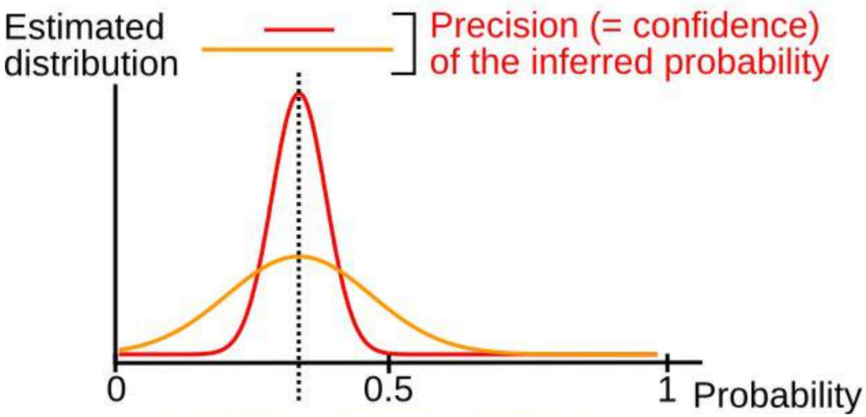




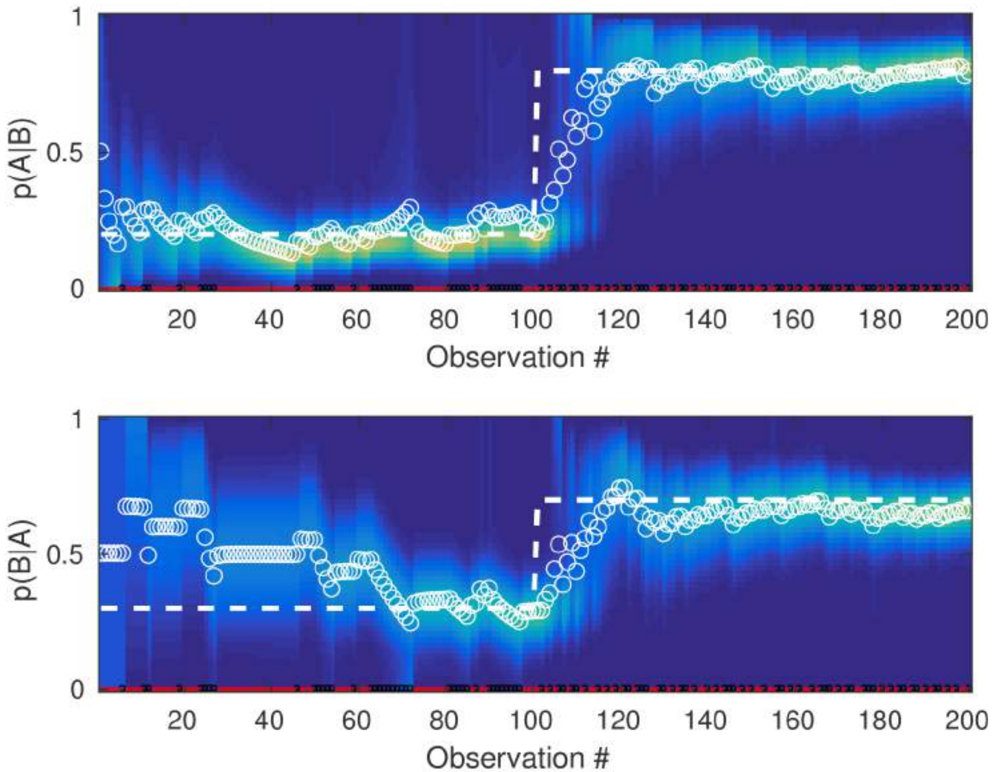
This strong correlation is found when each modality (visual, auditory) are tested separately. Results are also highly correlated between modalities, arguing in favor of a high-level inference system.

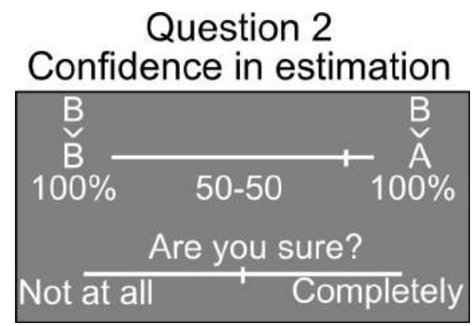
Meyniel, Schlunneger & Dehaene, *Plos Computational Biology* 2015

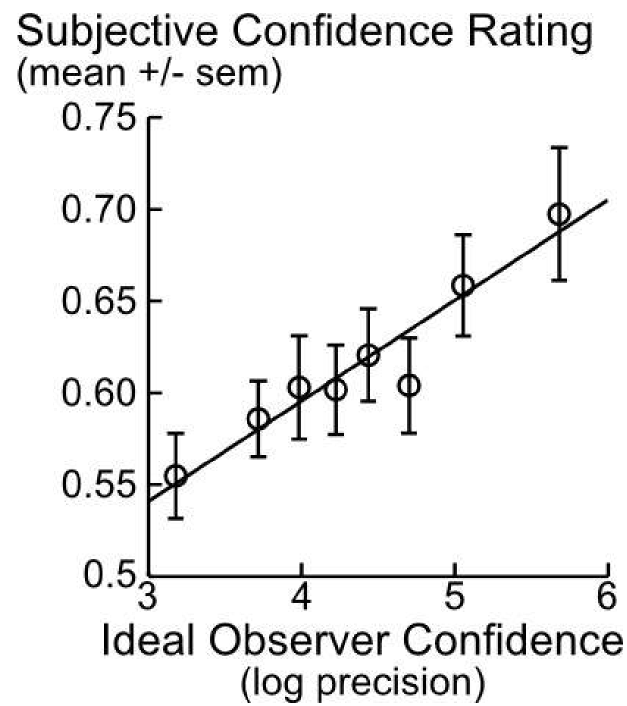
Is the inference probabilistic in the Bayesian sense? Evidence from the human sense of confidence



Example portion of sequence and inferred probabilities

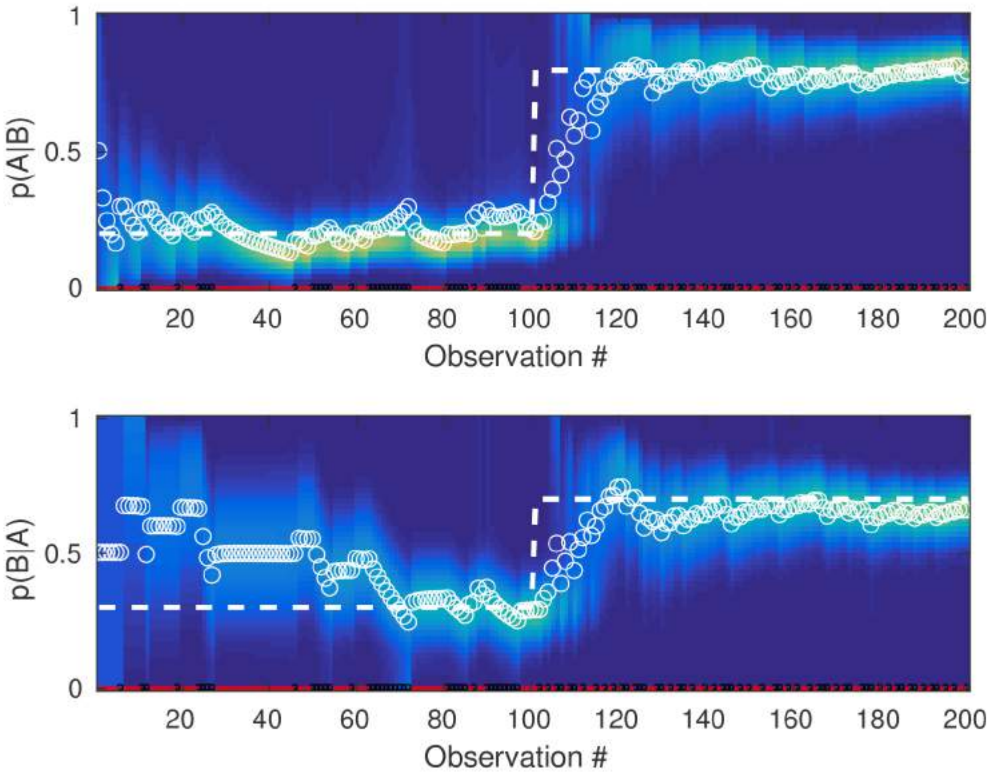


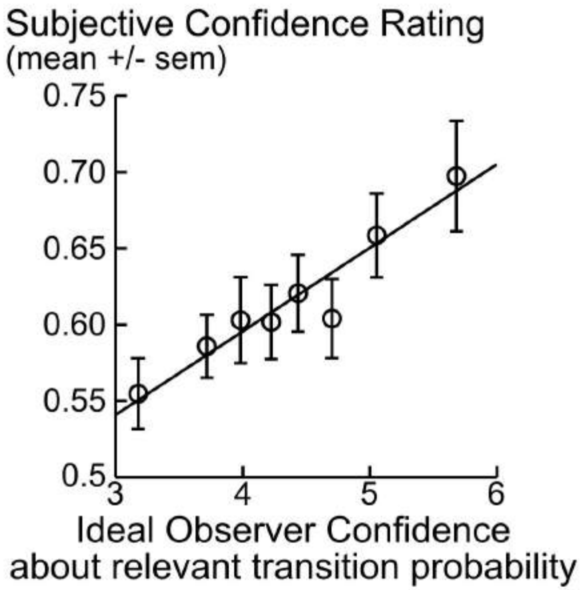


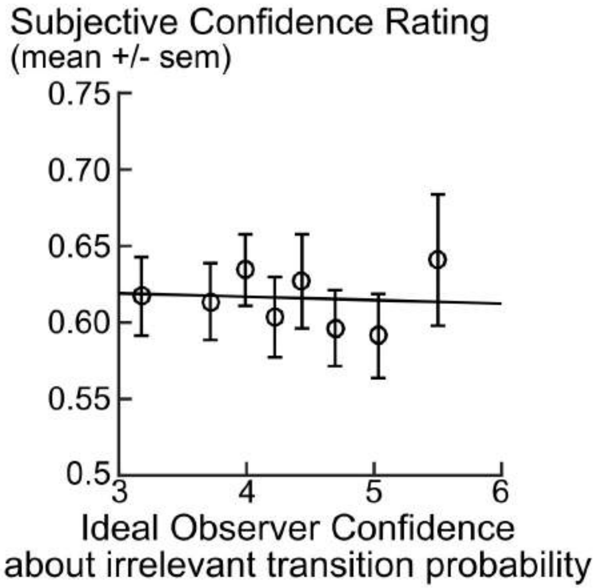


Meyniel, Schlunneger & Dehaene, *Plos Computational Biology* 2015

Subjects keep track of multiple confidence levels attached to the different probabilities they estimate



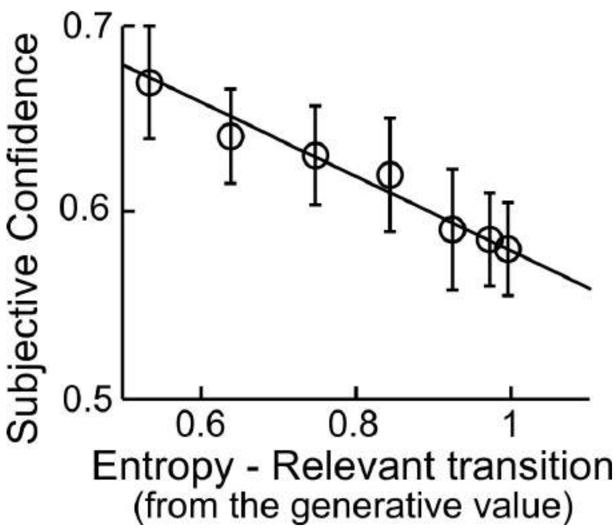




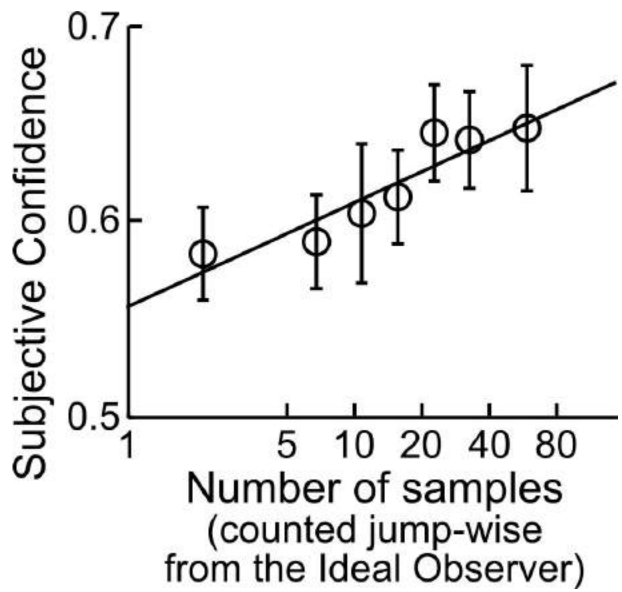
Meyniel, Schlunneger & Dehaene, *Plos Computational Biology* 2015

Human confidence judgments are rational: they are impacted by several factors similarly to the optimal inference

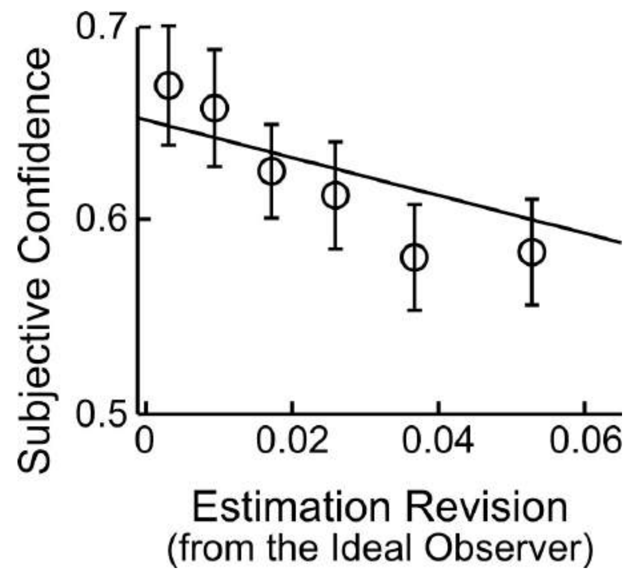
When outcomes are more difficult to predict (low predictability) confidence should be lower.



When more data support the inference, confidence should be higher.



When the current estimates need to be profoundly revised, confidence should be low.



Meyniel, Schlunneger & Dehaene, *Plos Computational Biology* 2015

SUMMARY

A Bayesien inference machinery constantly operates to extract the statistics of the observed sequence of events

- Our brain is equipped with a **powerful machinery for computing statistics** from sequences of observations.
- This inference machinery operates constantly to **predict future observations**.
- A central assumption of this inference process is that **changes may occur**.
- The brain infers, at a minimum, the **transition probabilities** between successive event types.
- This inference **Bayesian** in essence:
  - We entertain degree of belief, even about probabilities
  - The inference follows Bayes' rule
  - We constantly switch back and forth between estimates and predictions
- This statistical inference is **accessible to introspection**.
- **Confidence judgements** offer a window on our Bayesian brain *(Meyniel, Sigman and Mainen, Neuron 2015)*.