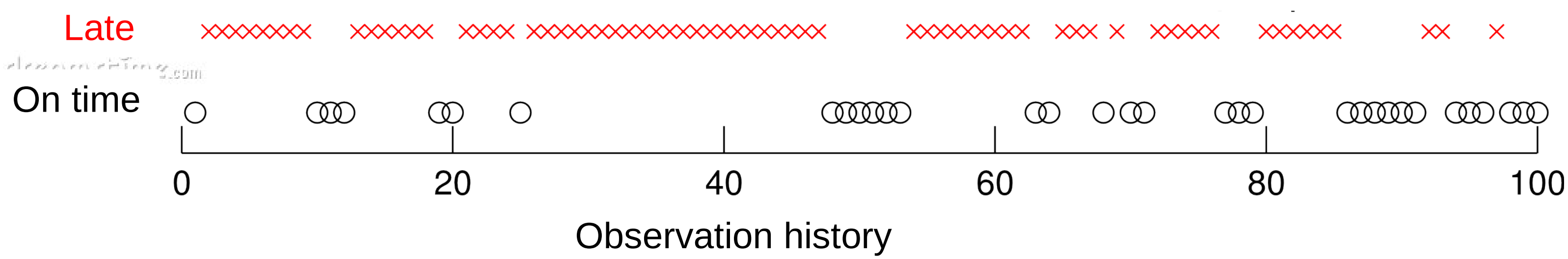
A rational sense of confidence during probabilistic inference in the human brain

Florent Meyniel
CEA Paris – France
Dehaene lab

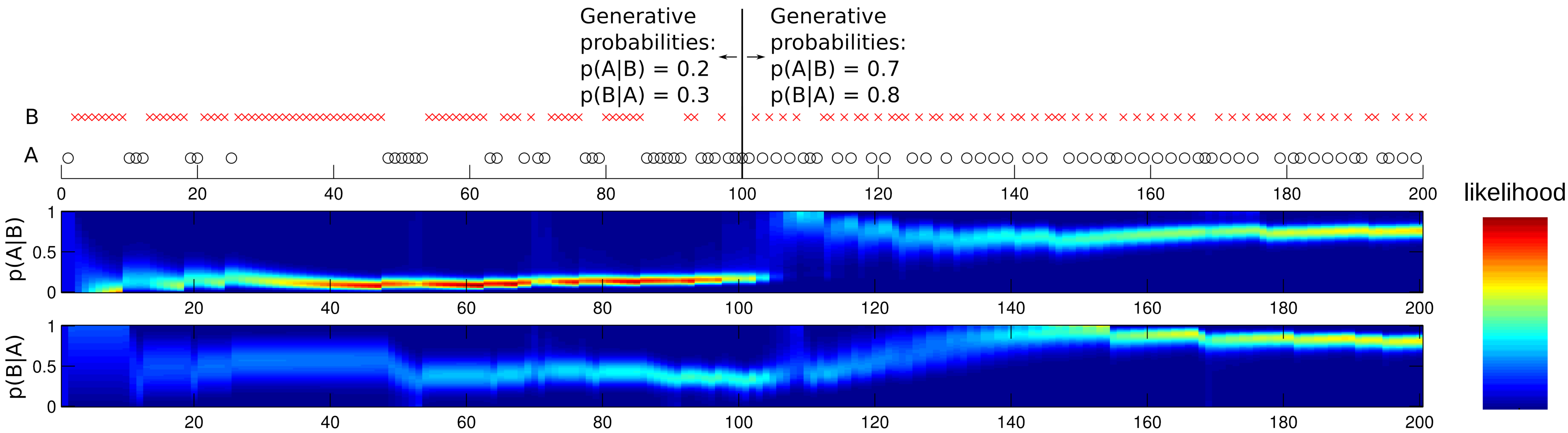
Probabilistic inference to extract the characteristics of our stochastic environments

Example: Will your colleague be late this morning?



- Humans and other animals learn in stochasticity world
  [Rescorla Wagner 1972, Dayan & Dolan Neuron 2013; Hyman 1953, Bornstein & Daw Plos Comp Biol 2013]
- Humans (at least) not only learn: they also have 'feelings of knowing', or confidence, about what they have learned
- Confidence has a clear definition and role in probabilistic inference

Bayesian probabilistic inference: principled account of learning and confidence about what has been learned



Characteristic of the ideal (Bayesian) observer model:
- performs Bayesian inference (it is optimal)
- infers the transition probabilities that generate the observed outcomes
- assumes that transition probabilities can change over time
- returns a posterior distribution of values

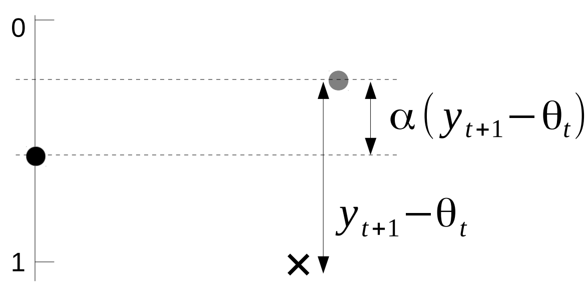
- Optimal Bayesian inference provides estimates of probabilities and a principled account of **confidence as the posterior precision** of the estimated distribution.
- Changes in probability induce **fluctuations of precision**, hence, of confidence in what has been learned

A computational role for confidence during learning: adjusting the weight of momentary evidence vs. prior knowledge

Delta rule (constant weighting)

The **update** is proportional to the deviation from prediction (**prediction error**)

$$\theta_{t+1} = \theta_t + \alpha \left(y_{t+1} - \theta_t \right)$$



Optimal Bayesian learning (confidence weighting)

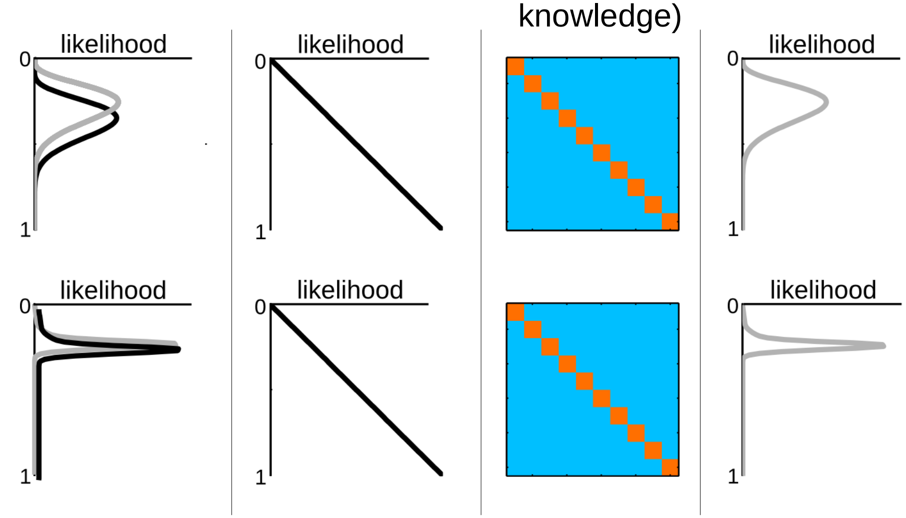
The **update** is determined by the unlikelihood of the observed event given what has been learned (**surprise**), the precision of the learned distribution (**confidence**) and the frequency of changes (**volatility**).

$$p\left(\boldsymbol{\theta}_{t+1} \mid y_{1:t+1}\right) \propto p\left(y_{t+1} \mid \boldsymbol{\theta}_{t+1}, y_t\right) \cdot \left(\boldsymbol{p}\left(\boldsymbol{\theta}_{t+1}, \boldsymbol{\theta}_t\right) p\left(\boldsymbol{\theta}_t \mid y_{1:t}\right) \right)$$

Updated knowledge | Momentary evidence | Volatility (overall stability of the world, hence, of knowledge) | Prior knowledge

Low confidence

High confidence



Volatility and *confidence* should not be conflated.

Previous studies mimicking unstable worlds: Behrens et al Nat Neuro 2007; Nassar & Gold J Neuro 2010; O'Reilly et al PNAS 2013, Gallistel et al 2014 Psych Rev, …

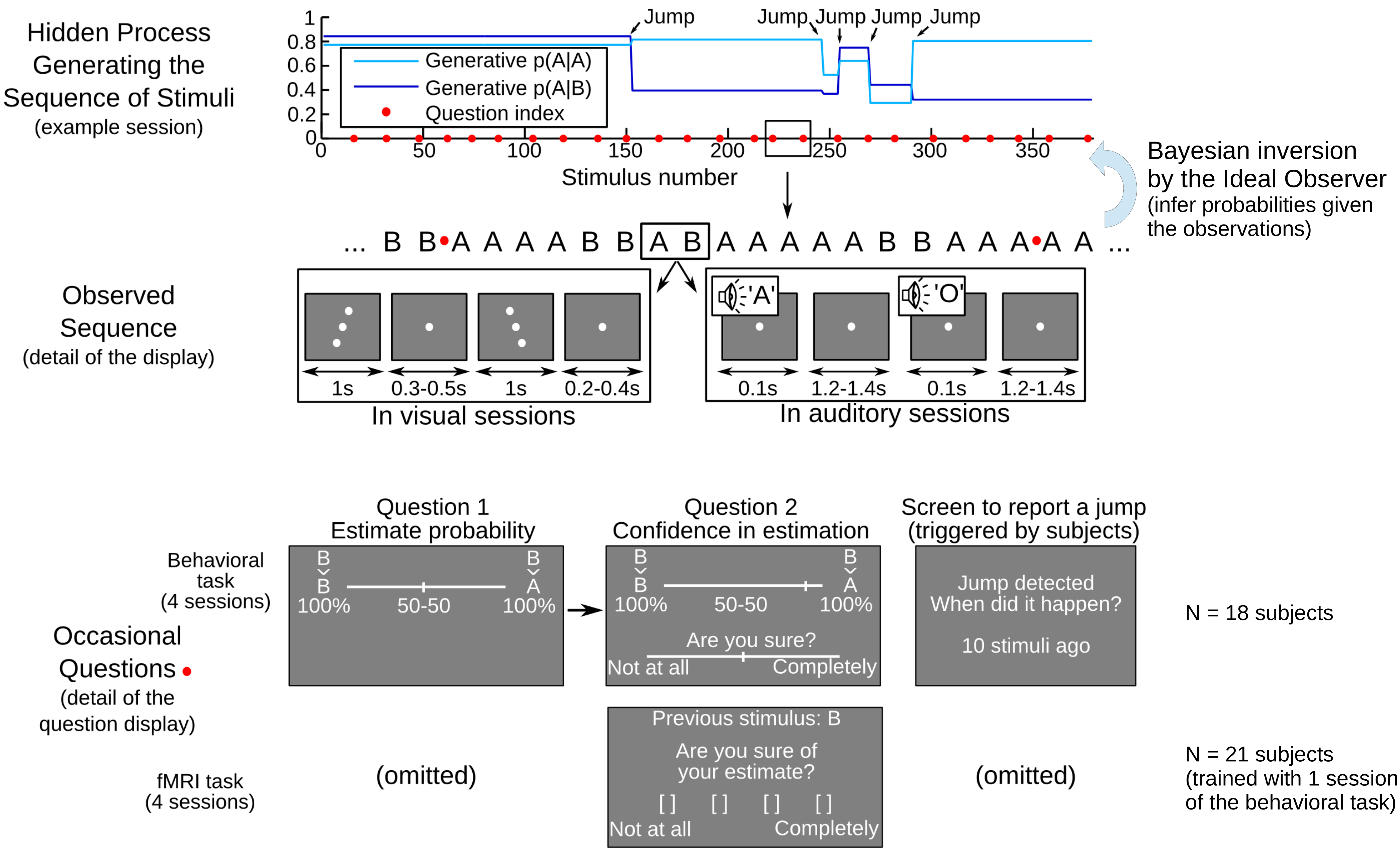
Topics addressed in this talk

In a stochastic and changing world (probabilistic learning task):

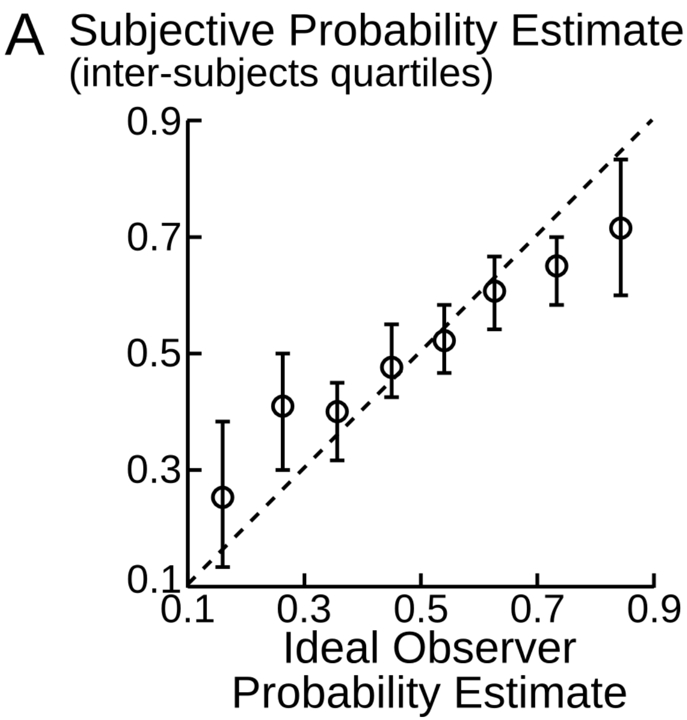
- Does the subjective confidence about what has been learned reflect the Bayesian notion of confidence?
- Do subjective estimates and confidence in those estimates, reveal properties expected from the (optimal) probabilistic inference?
- What are the functional correlates of confidence and learning in the brain?
- Do these correlates show evidence of a confidence-weighting of the momentary evidence as normatively prescribed?

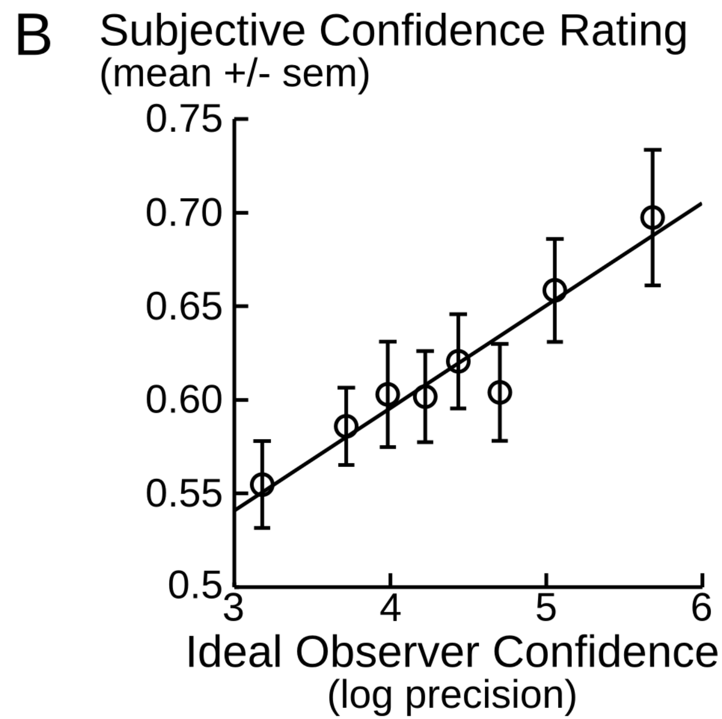
First behavioral part: Meyniel, Schlunneger & Dehaene, Plos Computational Biology 2015
Second fMRI part: work in progress

Task: estimation and confidence in a probabilistic learning task



Result #1: Accurate subjective prediction & confidence independent from the sensory modality

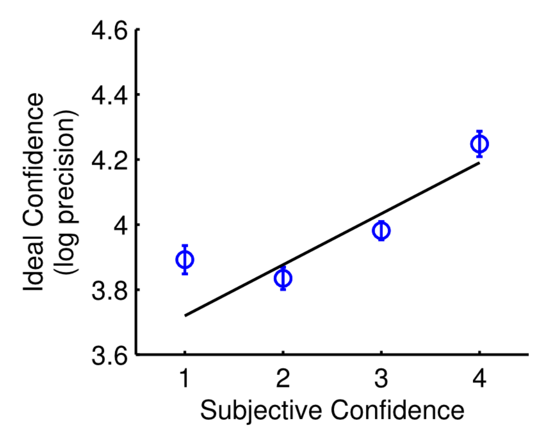




Are these two capabilities related?

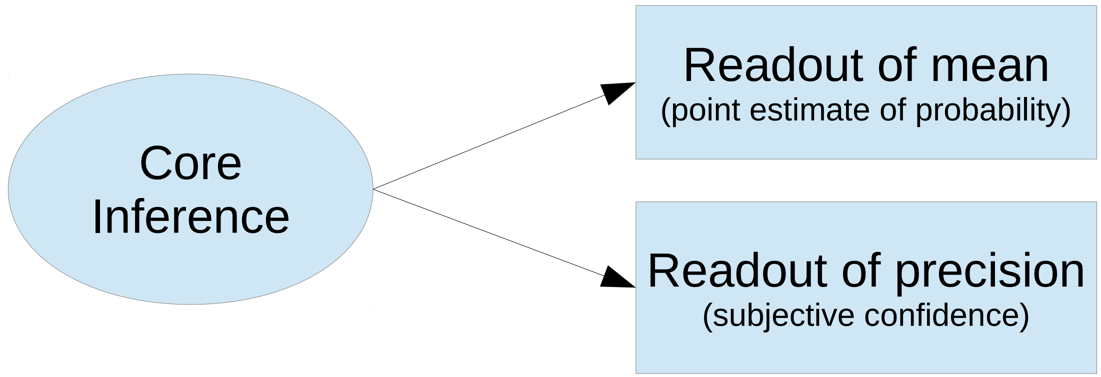
These strong correlations are found when each modality (visual, auditory) is tested separately. And results are highly correlated between modalities, arguing in favor of a high-level inference system.

Replication in the fMRI:



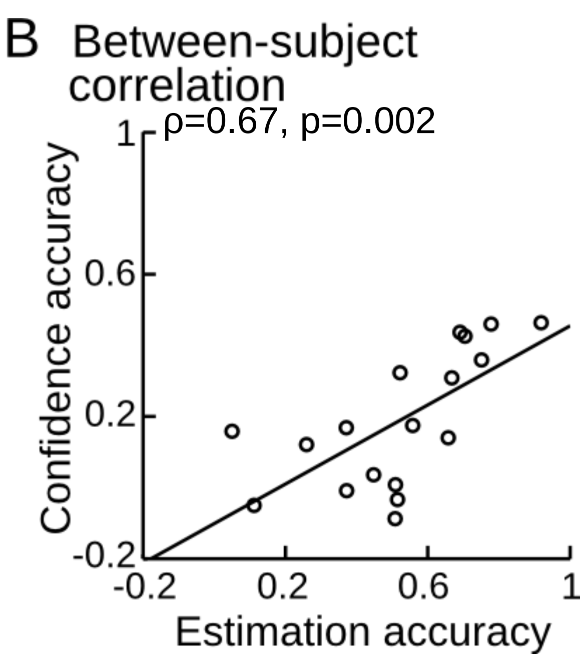
Results #2: Links between estimation of probability and confidence

The hypothesis



Accuracies in each estimate should be correlated across subjects

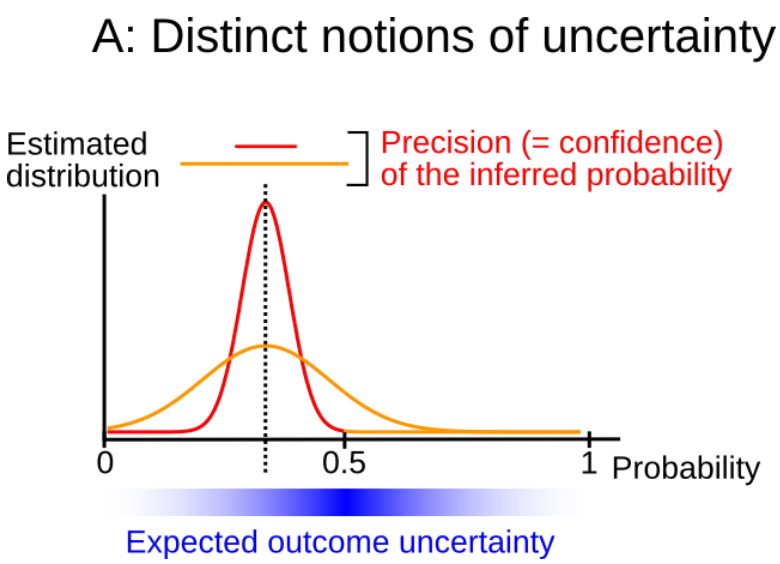
NB: the accuracy is characterized with respect to the optimum (correlation with the Ideal Observer)

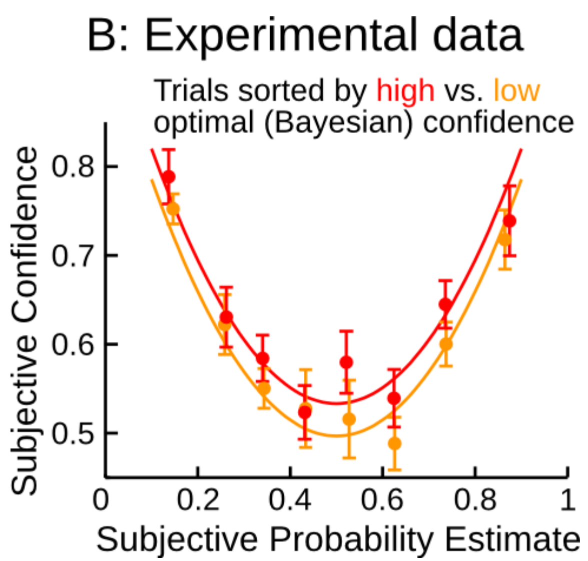


Accuracies in each estimate should be correlated across trials

- Weak correlation ($r=0.11 \pm 0.04$) but consistent across subject ($p<0.002$)
- Control analysis: the correlation survives when any systematic mapping between probability estimates and confidence level is explained away.

Results #2: Links between estimation of probability and confidence





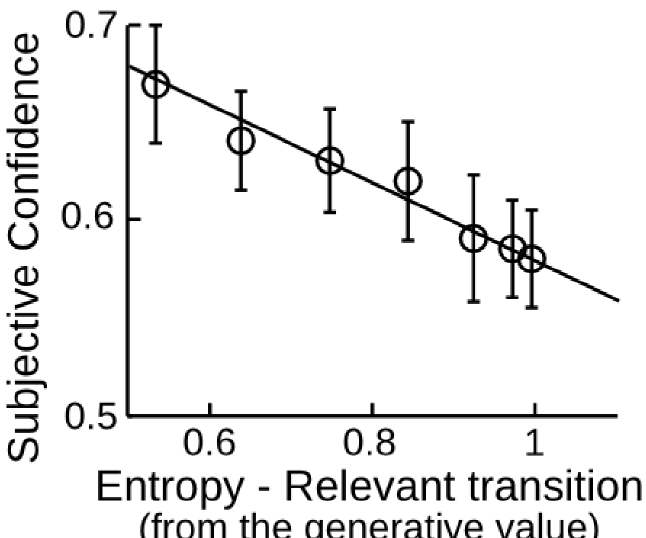
Dissociable components of subjective confidence:
→ the uncertainty arising from the unpredictability of the environment (inescapable, normative property of probabilistic inference)
→ the uncertainty about knowing this unpredictability

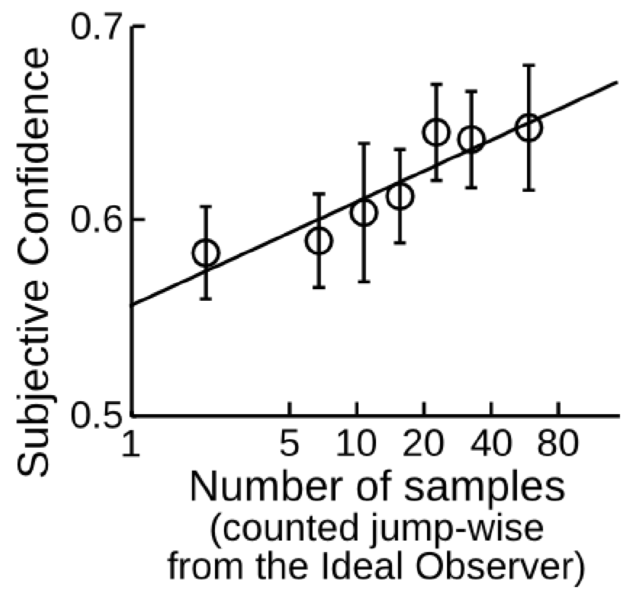
Result #3: Subjective confidence is impacted by several factors, similarly to the optimal inference

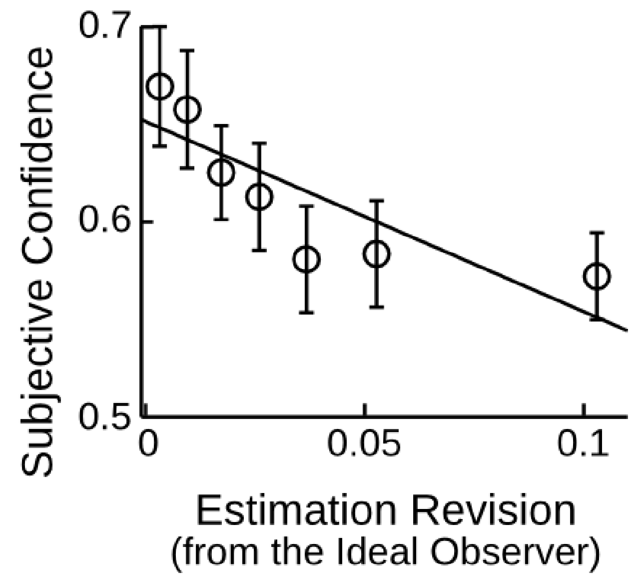
Normative Property #1
When outcomes are more difficult to predict (low predictability) confidence should be lower.

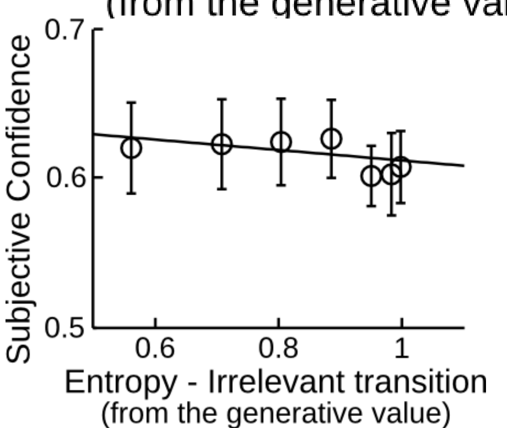
Normative Property #2
When more data support the inference, confidence should be higher.

Normative Property #3
When the current estimates need to be profoundly revised, confidence should be low.









Intermediate summary

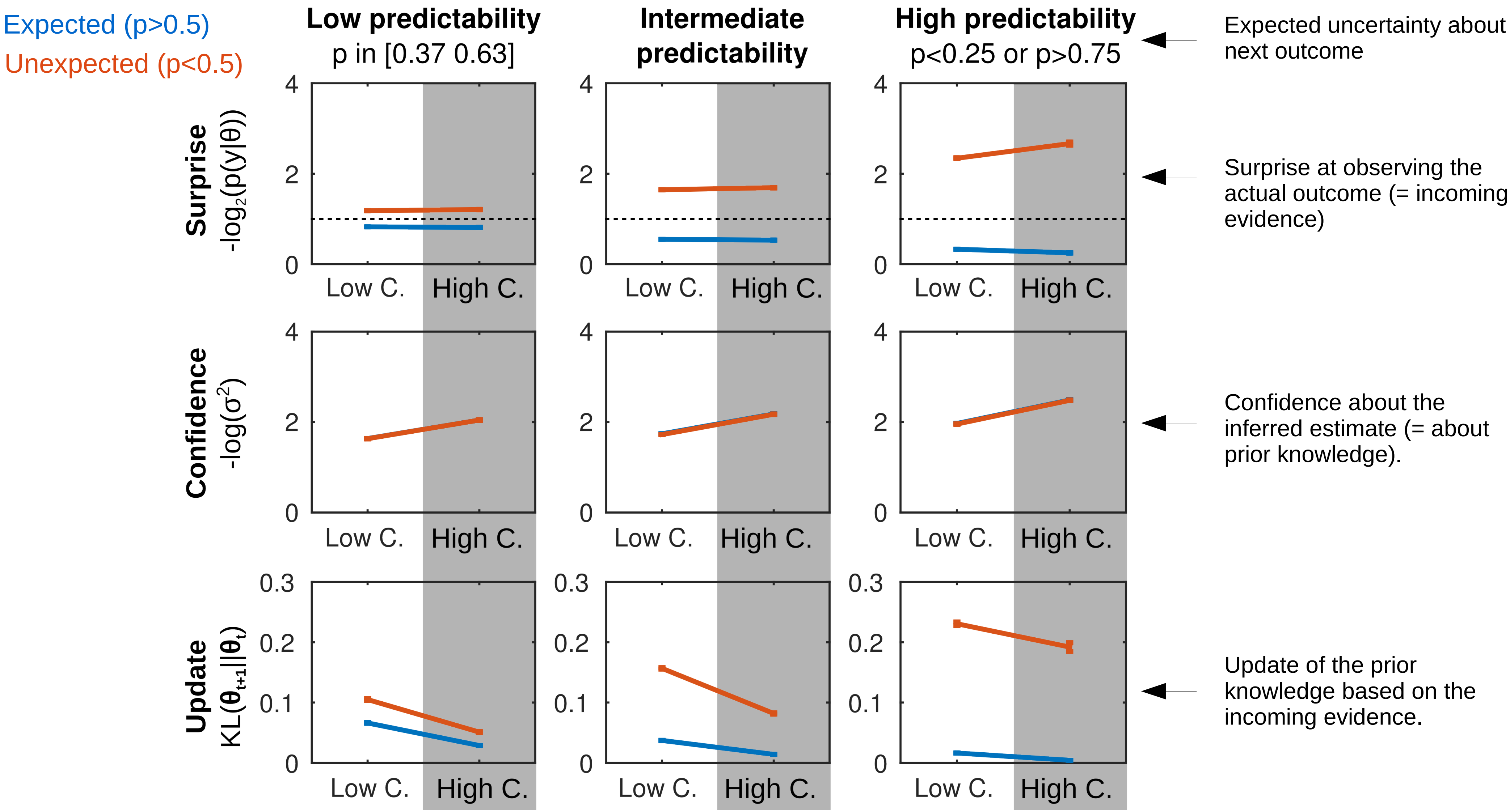
- Subjects are not only able to infer accurately probabilities, they are also able to estimate confidence levels in their inferences, in tight parallel with the optimal Bayesian inference.
- Their reports conform to several properties of a probabilistic inference.
- Additional results can be found in the publication, e.g.
  - subjects accurately detect changes in the generative probabilities
  - Several heuristics for confidence (non-probabilistic strategies) were ruled out

Meyniel, Schlunneger & Dehaene (2015) *Plos Computational Biology*
“The sense of confidence during probabilistic learning: a normative account”

Part 2: Does confidence-weighting contribute to balance prior knowledge and current evidence in the brain?

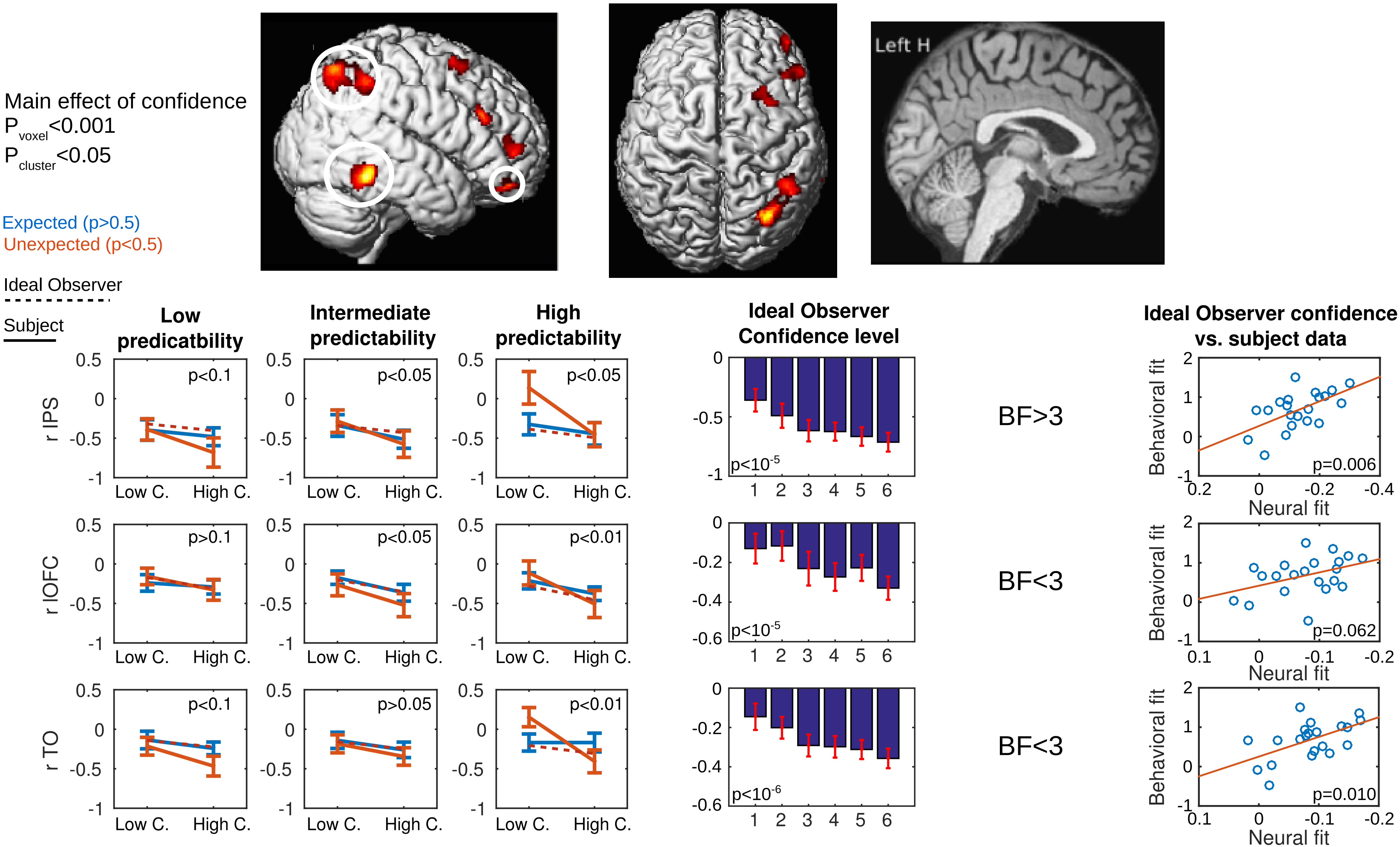
Dissecting confidence-weighting in Bayesian inference: update, confidence, surprise, predictability

The Ideal Observer estimates are sorted into bins to illustrate the **expected patterns** for confidence, surprise and update.



→ **A theory-driven approach to look for functional correlates of confidence-weighting**

Specific computational signatures of confidence in the brain



Confidence is distinct from predictability and surprise

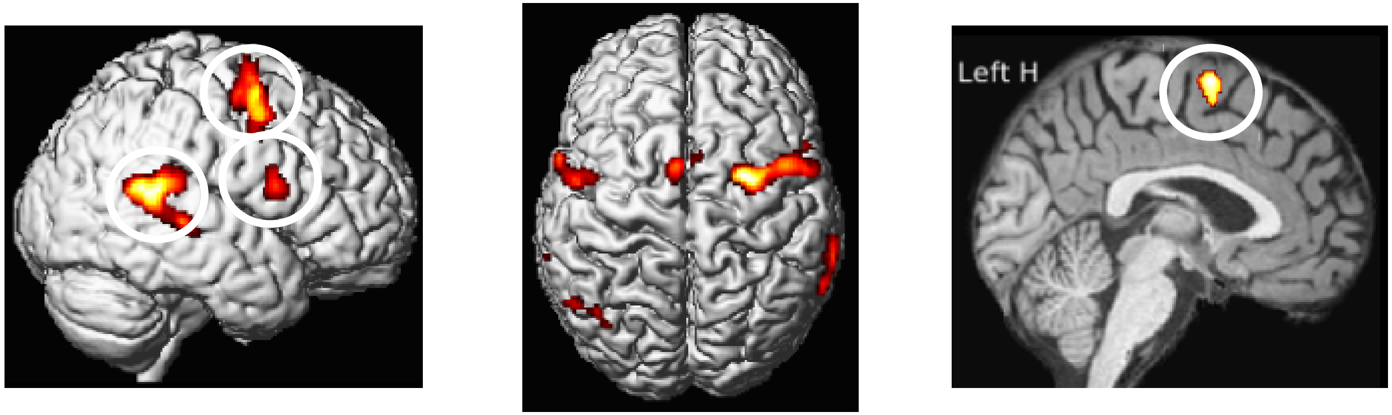
Parametric relation to optimal confidence

Modality independent

The neural data predict inter-subject variability

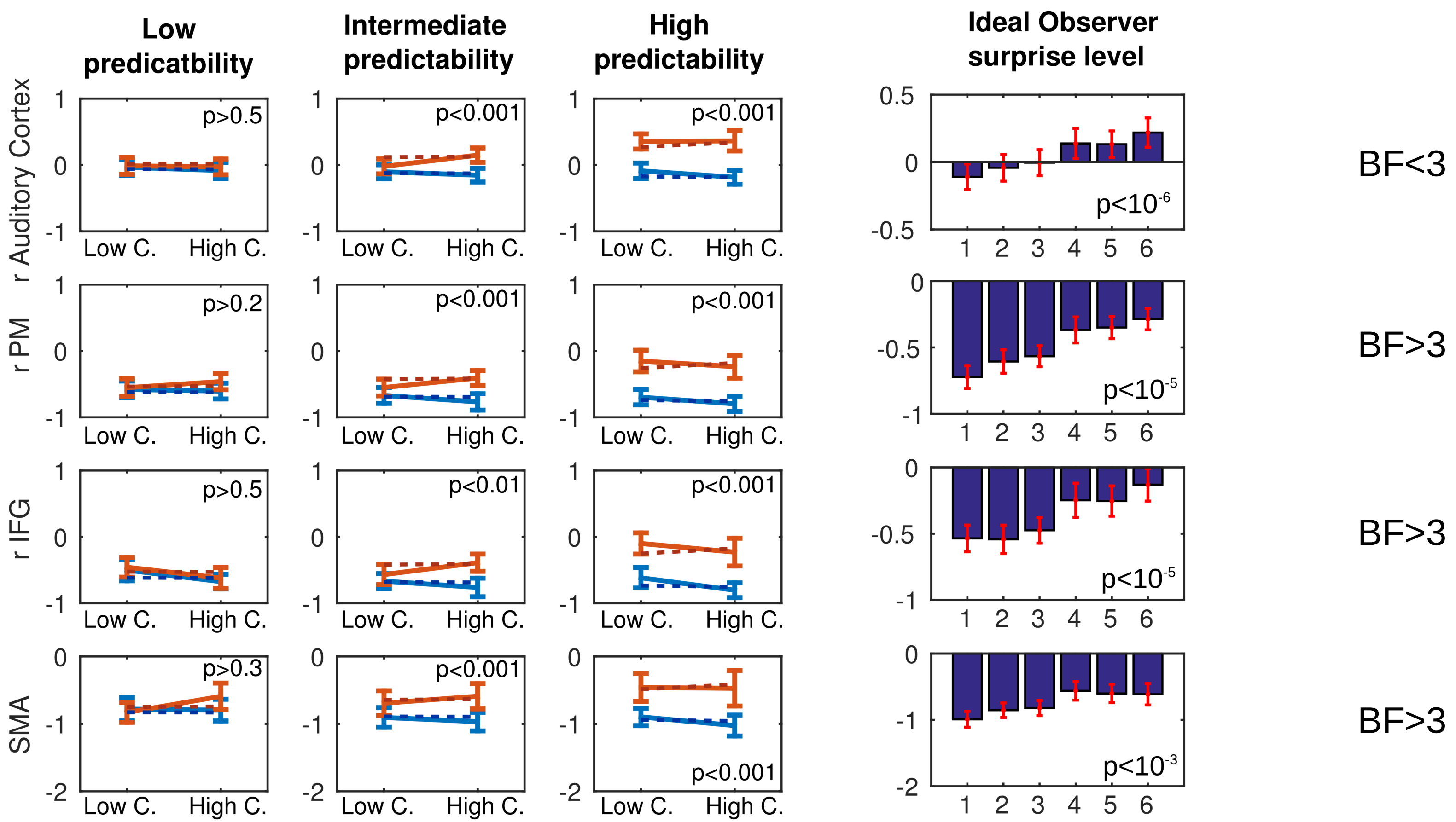
Specific computational signatures of surprise in the brain

Main effect of un/expected
$P_{voxel}<0.001$
$P_{cluster}<0.05$



Expected (p>0.5)
Unexpected (p<0.5)

Ideal Observer
Subject



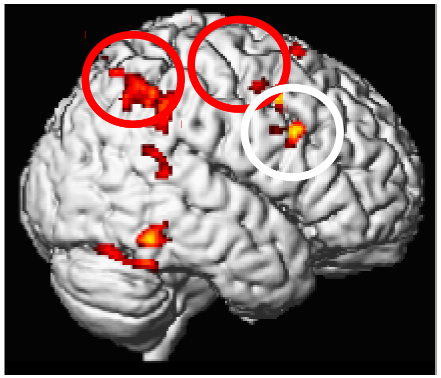
Surprise is modulated by predictability, not confidence

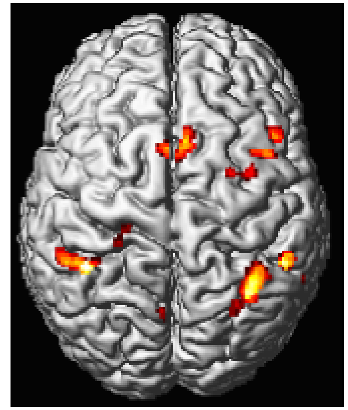
Parametric relation to optimal surprise

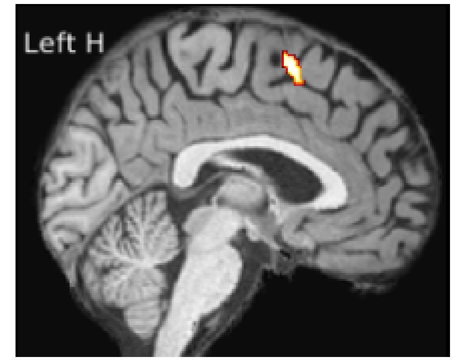
Modality independent

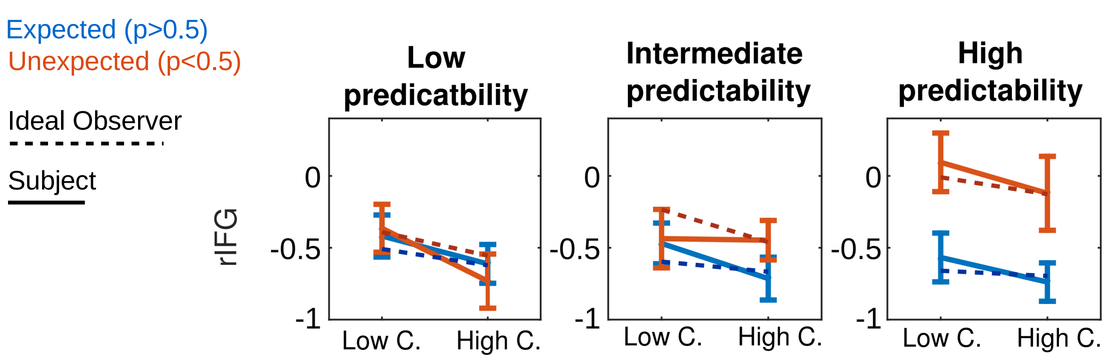
A combination of surprise and confidence signals: Specific computational signatures of confidence weighting

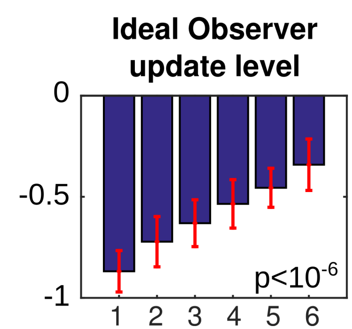
Physio-physiological Interaction between rIPS and rPM
$FWEP_{voxel}<0.001$
$P_{cluster}<0.05$











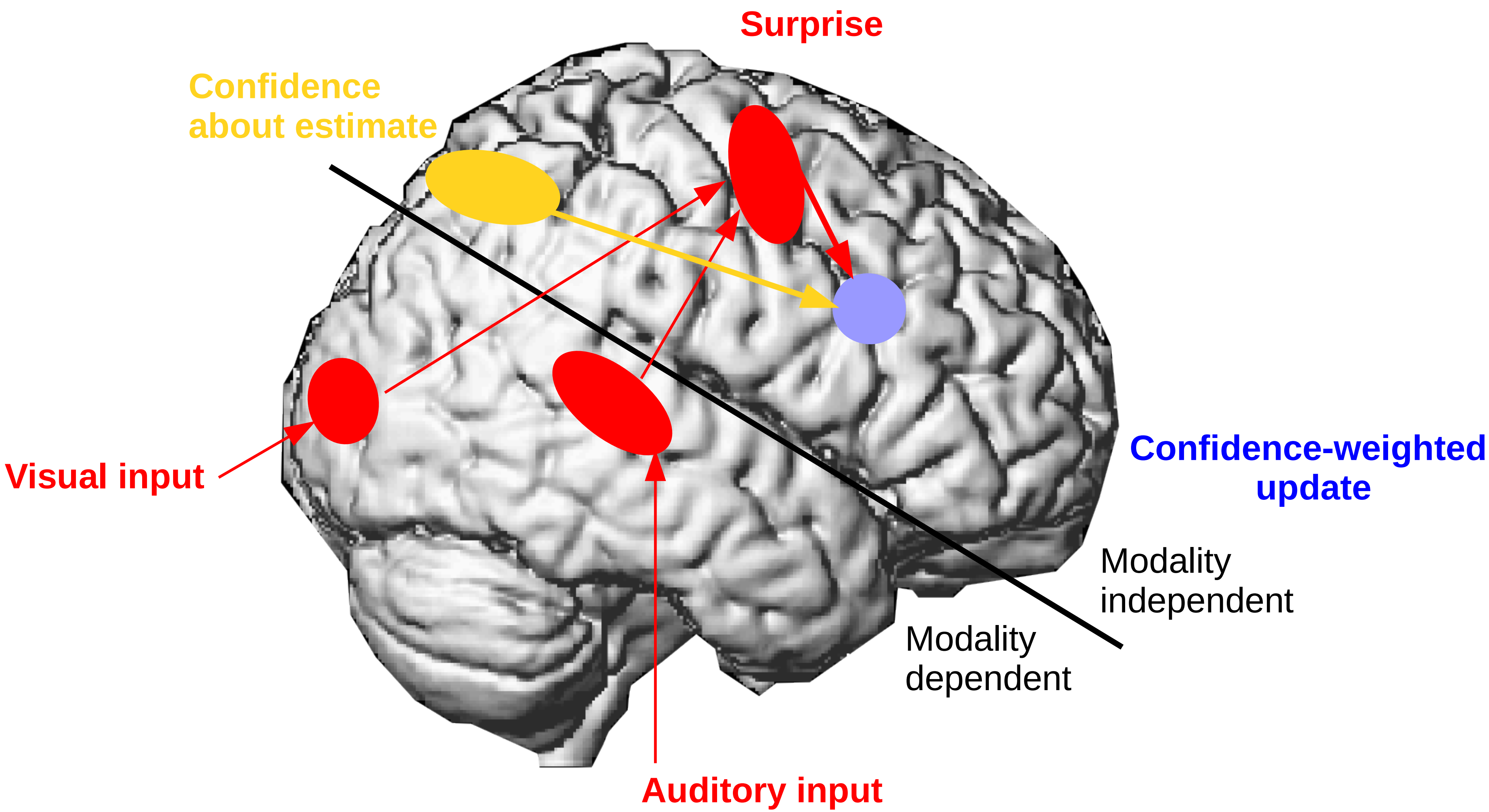
BF>3

Update as a combination of surprise and confidence

Parametric relation to optimal update

Modality independent

Graphical summary



Summary of the main findings

- Humans can accurately learn probabilities and assign rational confidence levels to their estimates, independent from a specific sensory modality.
- The properties of these estimates and confidence levels suggest that they both derive from the same probabilistic inference.
- Fluctuations of subjective and ideal confidence are driven by:
  - A first-level environmental uncertainty (predicting a stimulus given its probability of occurrence)
  - A second-level environmental uncertainty (changes in the probability of occurrence). Note that this second-level uncertainty may itself change (changes in the volatility).
- The confidence about what has been learned is tracked continuously in the brain.
- Confidence serves to weight the incoming evidence (surprising outcome) and update the internal knowledge. This process seems to rely on a fronto-parietal network.

Thanks

Funding: The Human Brain Project
Institution: CEA

Collaborators: Stanislas Dehaene
Daniel Schlunegger