Subjective confidence reveals the hierarchical nature of learning under uncertainty

Micha Heilbron and Florent Meyniel*

Affiliations
Cognitive Neuroimaging Unit, CEA DRF/I2BM, INSERM, Université Paris-Sud, Université Paris-Saclay, NeuroSpin center, 91191 Gif/Yvette, France

Correspondence
Micha Heilbron and Dr. Florent Meyniel
INSERM-CEA Cognitive Neuroimaging Unit
CEA/SAC/DRF/I2BM/NeuroSpin
Bat. 145 – Point Courrier 156
F-91191 Gif sur Yvette Cedex FRANCE
+33 1 6908 9501
micha.heilbron@gmail.com ; florent.meyniel@cea.fr

Short title
Confidence and hierarchical learning

Keywords
confidence, functional MRI, learning, probabilistic inference

ABSTRACT

Hierarchical processing is pervasive in the brain, but its computational significance for learning is disputed. We show that commonly used paradigms which rely on reports of learned quantities are insufficient to distinguish hierarchical from non-hierarchical models of learning under uncertainty, and present a novel paradigm which uses reports of confidence to arbitrate between the two. Our results provide clear evidence for hierarchical processing, and challenge influential neurocomputational accounts of learning.

MAIN TEXT

In real-world environments, learning is made difficult by two types of uncertainty[1]. First, there is inherent uncertainty in many real-world processes. For instance, the arrival of your daily commute may not be perfectly predictable but subject to occasional delays. Faced with such random fluctuations, learners should integrate as many observations as possible in order to obtain a stable, accurate estimate of the statistic of interest (e.g. the probability of delay)[2,3]. Second, there is the higher-order uncertainty related to sudden changes in those very statistics (*change points*). For instance, engineering works may increase the probability of delay for an extended period. Faced with such a change point, learners should discount older observations and rely on recent ones instead, in order to flexibly update their estimate[2,4,5].

Confronted with both forms of uncertainty, the optimal learning strategy is to track not only the statistic of interest but also the higher-order probability of change points. This allows learners to render their estimate stable when the environment is stable (i.e. between change points), and flexible when the environment changes[2,5–7]. Importantly, this approach assumes that learners use a *hierarchical model* of their environment comprising multiple levels: observations, the statistics of observations and the possibility of change points. The hierarchical approach is widely used to study learning in health[2,8] and disease[9]. However, efficient learning in dynamic environments is possible without tracking the likelihood of individual change points[3], and a large body of work indeed uses such a solution to model behavioral and brain responses[10,11]. Computationally, this approach is very different as it assumes that learners are oblivious to higher-order factors (change points) and hence use a non-hierarchical (*flat*) model of the world. The use of internal hierarchical models is an active area of research in cognitive science[12], and has important consequences for neurobiology, since hierarchical models[13,14] and non-hierarchical ones[3,10] require different neural architectures. In learning theory however, the issue is overlooked and the hierarchical approach remains largely untested, being assumed critical by some frameworks of learning[2,5,6] and unnecessary by others[3,10]. Here, we address this issue by developing a new way to test whether learners use a hierarchical model of the world.

Most earlier studies on learning under uncertainty investigated the statistics of observations estimated by subjects (e.g. reward probability), either directly via explicit reports[4,6] or indirectly via decisions guided by those statistics[2,9]. They reported that a metric derived from those reports and decisions, the *apparent learning rate,* is modulated by the occurrence of change points. This was argued a hallmark of hierarchical learning, as it supposedly revealed sensitivity to the higher-level likelihood of change points[2,4]. However, we show that a flat learning model can produce exactly such adjustments without tracking the likelihood of change points (see Fig. S1); therefore, modulations of the apparent learning rate are not diagnostic of hierarchical models. One reason for this may be that the apparent learning rate is derived from the observation statistics that subjects estimate (e.g. reward probability). Such *first-order statistics* only describe the level of observations and are seldom unique to a single model. By contrast, *second-order statistics*, such as the learner's confidence about her estimates, describe the learner's inference and can be more diagnostic about the underlying computations[15]. To illustrate this, we simulated an optimal hierarchical learning model and a near-optimal flat model in a task from a recent study on learning transition probabilities[6]. The (first-order) estimates of the flat model approximated those of the optimal hierarchical model extremely well: they were virtually indistinguishable (Fig. S2a). However, the (second-order) confidence values associated with these estimates were much less correlated, offering a more sensitive metric to discriminate the models (Fig. S2b). Altogether, our simulations show that some classic (first-order) metrics barely distinguish hierarchical from flat accounts of learning, and that second-order statistics like *confidence* may be better suited for this purpose.

We designed a task, inspired by a previous study[6] and Fig. S2, that uses confidence to reveal clearly dissociable signatures of hierarchical and flat models. Participants observed long sequences of two stimuli (A and B), the occurrence of which was governed by two transition probabilities which subjects had to learn: P(A|A) and P(B|B). The value of each probability was independent, but occasionally both simultaneously changed. Subjects passively observed the stimuli and occasionally reported their (first-order) estimate of P(A|A) or P(B|B) and their (second-order) confidence in that estimate (see Fig. 1a). First, we verified that probability estimates, and subjective confidence in those estimates, matched the predictions of an optimal hierarchical model (Fig. 1b) that 'inverts' the generative structure of the task using Bayes' rule (see methods). Linear regressions between subjects' and optimal values showed a tight agreement for probability estimates (β=0.66±0.06 s.e.m., t_{22}= 11.13, p=1.7 10^{-10}), and for confidence reports (β=0.10±0.03 s.e.m., t_{22}= 3.06, p=0.0058). Despite being somewhat noisier, confidence reports showed many key properties of optimal inference (see SI). Critically, if confidence reports are more diagnostic than probability estimates, they must convey information not already conveyed by probability estimates. This was confirmed by regressing out the (theoretically expected) covariance between confidence values and several first-order metrics (see SI), and finding that the residuals of this regression varied with the optimal confidence (β=0.028±0.012,

t_{22}=2.3, p=0.029). This result was replicated by performing the same analysis on another dataset[6] (0.023±0.010, t_{17}=2.2, p=0.0436).

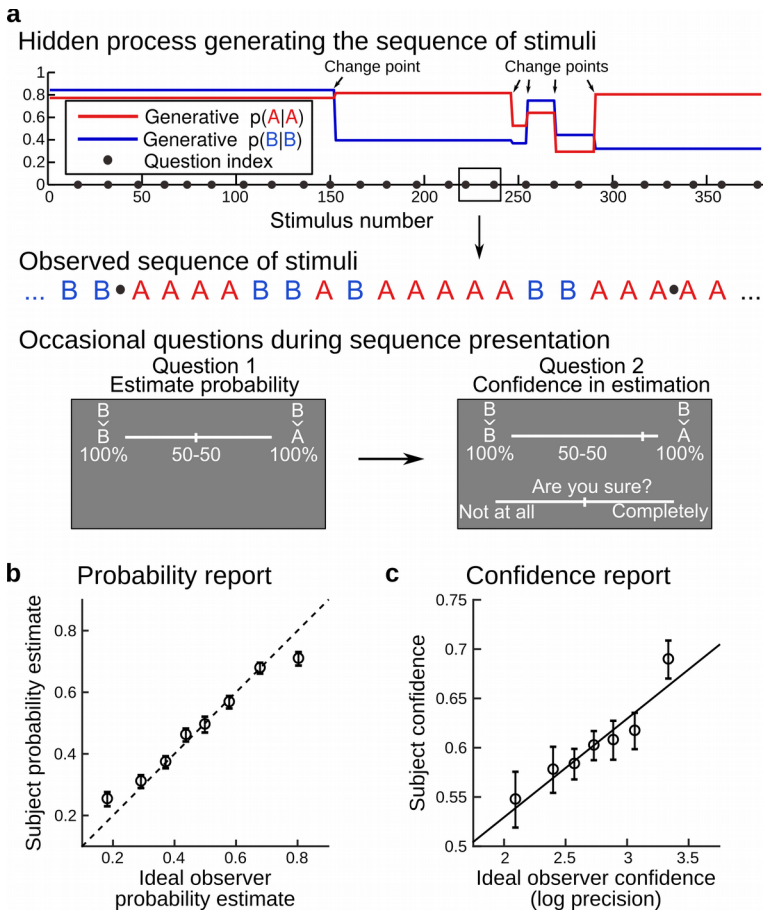

Figure 1. (a) Probability learning task. Human subjects (N=23) were presented with random sequences of two stimuli A and B. The stimuli were, in distinct sessions, either auditory or visual and they were perceived without ambiguity. At each trial, one of either stimulus was sampled according to a probability that depended on the identity of the previous stimulus: p(A|A) and P(B|B). These transition probabilities underwent occasional, abrupt changes (change points). A change point could occur at any trial with a probability that was fixed throughout the experiment. Subjects were instructed about this generative process and had to continuously estimate the (changing) transition probabilities given the observations received. Occasionally (see black dots in a), we probed their inferences by asking them, first, to report the probability of the next stimulus (i.e. report their estimate of the relevant transition probability) and second, to rate their confidence in this probability estimate. **(b, c)** Subjects' responses were compared to an ideal observer model that solved the same task using optimal Bayesian inference. Numeric values of confidence differ between subjects and models since they are on different scales (from 0 to 1 in the former, in log-precision unit in the latter). For illustration, the ideal observer values were binned, the dashed line (b) denotes the identity, the plain line (c) is a linear fit, and data points correspond to subjects' mean ± s.e.m.

We then analyzed confidence reports to test for hierarchy. This was possible because changes in the two transition probabilities (P(A|A) and p(B|B)) were *coupled*, changing simultaneously. A flat learner then only estimates the value of each transition probability, while a hierarchical model *also* estimates the probability of a global change point. Faced with a change point, the hierarchical model reacts optimally and makes its prior knowledge more malleable by becoming uncertain about *both* P(A|A) and P(B|B). Knowing that change points are coupled between transition probabilities allows for generalization: if a change point is suspected after observing just one type of transition (e.g.

AAAAAAA, when P(A|A) was estimated to be low) a hierarchical learner would *also* become uncertain about the other quantity, P(B|B), despite having acquired no direct evidence on this transition (Fig 2a). Critically, this form of indirect inference is unique to hierarchical models and hence offers a powerful test of hierarchical theories of learning.

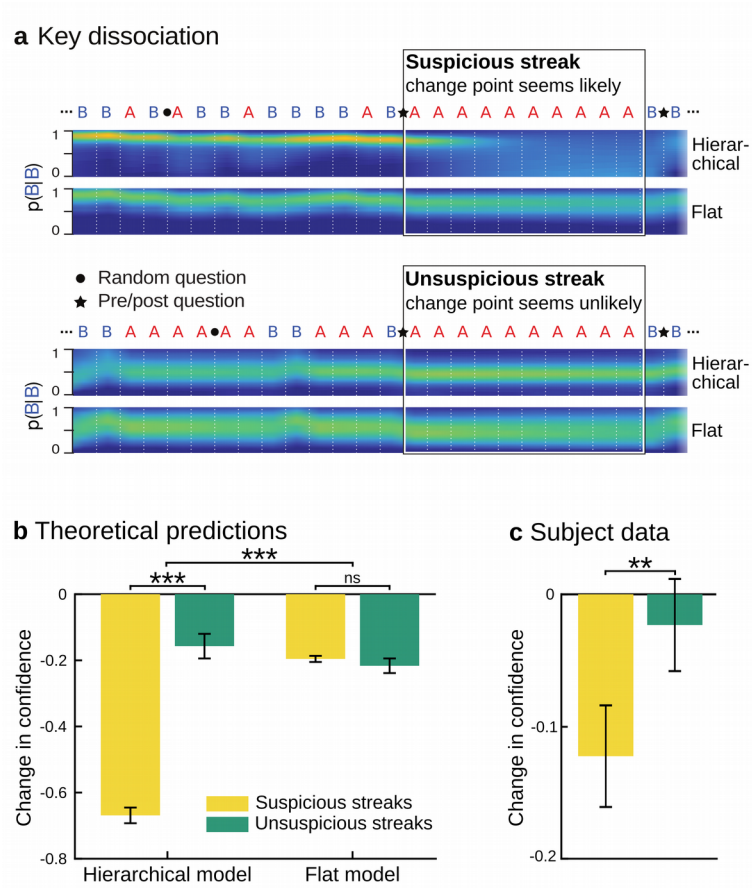

Figure 2. (a) Divergent predictions of hierarchical versus flat learning models. Two fragments of sequences are shown in which one stimulus ('A') is consecutively repeated 10 times. In the upper fragment, this streak of repetitions is highly unlikely (or 'suspicious') given the context, and may indicate that the underlying statistics changed. By contrast, in the lower fragment, the same streak is not unlikely, and does not suggest a change point. The heat maps show the posterior probability distribution of P(B|B) or the probability of a repetition of the other stimulus (B), estimated by a hierarchical (top heat map) and flat (bottom heat map) model. In a hierarchical model, unlikely streaks arouse the suspicion of a global change in statistics, causing the model to become uncertain about its estimates of both transition probabilities, despite having acquired no direct evidence on P(B|B). In a flat model, by contrast, a suspicious streak of A's will not similarly decrease the confidence in P(B|B), because a flat model does not track global change points. To test for this effect, pre/post questions (indicated by a star) were placed immediately before and after selected streaks, to obtain subjective estimates of the transition probability corresponding to the stimulus not observed during the streak. Streaks were categorized as suspicious if they aroused the suspicion of a change point from the hierarchical ideal observer's viewpoint. Note that the Flat model also shows a decrease in confidence, because it progressively forgets its estimates about P(B|B) during a streak of (A|A). Critically however, there is no difference between suspicious and unsuspicious streaks. **(b)** For the exact sequences presented to subjects, the change in confidence (post-streak minus pre-streak) was significantly modulated by streak type in the hierarchical model, but not in a flat model. **(c)** Subjects' confidence showed an effect of streak type predicted by the ideal hierarchical model. As in Fig. 1C, confidence values in subjects and models are on different scales. Data points are subjects' mean ± s.e.m; significance levels correspond to paired t-test with $p<0.005$ (**) and $p<10\text{-}12$ (***).

To test for this generalization effect, we focussed on streaks of repetitions, and distinguished between streaks that seem unlikely in context and may signal a change point (*suspicious streaks)* and streaks that do not (*unsuspicious streaks*). Stimulus sequences were carefully selected to contain enough suspicious and unsuspicious streaks and to control for confounds such as streak duration (see SI). Questions were inserted just before and after the streak, so that subjects reported their estimate of (and confidence in) the other, non-repeating transition (Fig. 2a). Exact theoretical predictions for both models are found in Figure 2b. In the hierarchical model, confidence decreases strongly after suspicious, but much less after unsuspicious streaks. But in the flat model, there is no such difference. Strikingly, subjective reports exactly followed the hierarchical account and falsified the flat one: confidence decreased strongly after suspicious (-0.12±0.04 s.e.m, t_{22}=-3.2, p=0.004) but not after unsuspicious streaks (-0.02±0.03 s.e.m., t_{22}=-0.7, p=0.51), and this interaction was significant (paired difference, 0.10±0.03 s.e.m., t_{22}=3.7, p=0.001).

One alternative explanation for this effect is that suspicious streaks were more surprising and that subjects may become generally uncertain after surprising events. We therefore performed a control experiment in which both probabilities changed independently: here, suspicious streaks were equally surprising but no longer signalled a *global* change point – and indeed, generalization of a decrease in confidence was no longer observed (see Fig. S3). Another possibility is that the result may have been affected by subjects' understanding of the goal of the experiment. However, this was ruled out by interviews after the tasks which revealed that no subject had noticed that a sizable fraction (~30%) of questions purposefully surrounded target streaks (see SI). A final potential confound is that subjects may have used a flat model while tracking not transition probabilities, but some other (heuristic) quantity. However, detailed analysis revealed that subjects did track transition probabilities (see SI) and that no heuristic flat model could explain the selective decrease of confidence (Fig S4).

In sum, we have shown that commonly used paradigms are insufficient to distinguish hierarchical from non-hierarchical models of learning, and introduced a novel test to dissociate the two. Our results provide clear evidence for hierarchical learning, thereby challenging very common and influential non-hierarchical accounts[3,10,11], although they do not rule out the that flat algorithms may *also* exist within the brain. Critically, because our paradigm taps into a unique property of hierarchical models, both classes of theories make *qualitatively different predictions* at the level of individual trials. Such qualitative tests have previously proven useful[16], for instance to dissociate model-free and model-based reinforcement learning[17]. This is because qualitative tests are much simpler and more powerful than typical model comparison procedures, as they rely directly on subjects' reports rather than on model fitting. Our test was carried out in humans, but it can be applied to any learning model that computes uncertainty, and will therefore be of interest to experimentalists and theoreticians alike.

References

1. Yu, A.J. & Dayan, P. *Neuron* **46**, 681–692 (2005).
2. Behrens, T.E.J., Woolrich, M.W., Walton, M.E. & Rushworth, M.F.S. *Nat. Neurosci.* **10**, 1214–1221 (2007).
3. Yu, A.J. & Cohen, J.D. *Adv. Neural Inf. Process. Syst.* **21**, 1873–1880 (2008).
4. Nassar, M.R., Wilson, R.C., Heasly, B. & Gold, J.I. *J. Neurosci.* **30**, 12366–12378 (2010).
5. Mathys, C. *Front. Hum. Neurosci.* **5**, (2011).
6. Meyniel, F., Schlunegger, D. & Dehaene, S. *PLoS Comput. Biol.* **11**, e1004305 (2015).
7. Iigaya, K. *Elife* **5**, (2016).
8. Iglesias, S. et al. *Neuron* **80**, 519–530 (2013).
9. Lawson, R.P., Mathys, C. & Rees, G. *Nat. Neurosci.* **20**, 1293–1299 (2017).
10. Farashahi, S. et al. *Neuron* **94**, 401–414.e6 (2017).
11. Rescorla, Robert A. & Wagner, Allan R. *Classical conditioning II: Current research and theory* 64–99 (1972).
12. Tenenbaum, J.B., Kemp, C., Griffiths, T.L. & Goodman, N.D. *Science* **331**, 1279–1285 (2011).
13. Friston, K. *PLoS Comput. Biol.* **4**, e1000211 (2008).
14. Lee, T.S. & Mumford, D. *J. Opt. Soc. Am.* **20**, 1434 (2003).
15. Meyniel, F., Sigman, M. & Mainen, Z.F. *Neuron* **88**, 78–92 (2015).
16. Palminteri, S., Wyart, V. & Koechlin, E. *Trends Cogn. Sci.* **21**, 425–433 (2017).
17. Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P. & Dolan, R.J. *Neuron* **69**, 1204–1215 (2011).

Acknowledgements

We wish to thank Stefano Palminter, Maxime Maheu, Peter Latham and Alexandre Pouget for useful comments on this work.

Online Methods

Participants

Participants were recruited by public advertisement. They gave a written informed consent prior to participating and received 20 euros for volunteering in the experiment. The study was approved by the local Ethics Committee (CPP n°08–021 Ile de France VII). 26 participants (17 female, mean age 23.8, s.e.m.: 0.49) performed the main task and 21 other participants performed the control task (11 female, mean age 23.0, s.e.m.: 0.59). We excluded participants who showed poor learning performance, which we quantified as the Pearson ρ coefficient between their probability estimates and the ideal observer's estimates. We used a threshold corresponding to a 5% of the (lowest) values measured in this task ($\rho<0.18$, from a total of 105 participants in this study and others) This excluded 3 subjects from the main task, and none from the control task.

Main Task

The task was run using Octave (Version 3.4.2) and PsychToolBox (Version 3.0.11). Each participant completed a total of 5 blocks: 1 training block and 4 experimental blocks (2 auditory, 2 visual). Auditory and visual blocks alternated, with the modality of the first block randomised across participants. In each block, subjects were presented binary sequences of 380 stimuli (1520 total) denoted A and B, which were either visual symbols or sounds and were perceived without ambiguity (Fig. 1).

Sequences were generated according to the same principles as in previous studies[6, 18]. A and B were randomly drawn based on two hidden transition probabilities which subjects had to learn. These probabilities were stable only for a limited time. The length of stable periods was randomly sampled from a geometric distribution with average length of 75 stimuli, truncated at 300 stimuli to avoid overly long stable periods. Critically, and contrary to other studies[2] the volatility was thus fixed (at 1/75). Transition probabilities were sampled independently and uniformly between 0.1-0.9, with the constraint that, for at least one of the two probabilities, the change in odd ratio (p/1-p) between consecutive stable periods was at least fourfold, thus guaranteeing that the change was effective. Across sequences and subjects, the actually used generative values indeed covered the transition probability matrix 0.1-0.9 uniformly, without any correlation (Pearson $\rho = -0.0009$, $p = 0.98$). Occasionally, the sequence was interrupted and subjects had to estimate the probability that the next stimulus would be either an A or a B and report their confidence in that estimate. Questions were located quasi-randomly, semi-periodically once each 15 stimuli on average (100 in total). Of the 100 questions, 68 questions were randomly placed; the remaining 32 questions were intentionally located just before and after 16 selected streaks (8 suspicious, 8 non-suspicious) and functioned as pre/post-

questions to evaluate the effect of these streaks (see Fig 2). For details on the definition and selection of suspicious/non-suspicious streaks, see below.

To familiarize participants with the task they were carefully instructed and performed one training block of 380 stimuli (or ~12 minutes). To make sure they were fully aware of the volatile nature of the generative process, participants had to report when they detected changes in the hidden regularities. In the experimental sessions, reporting change points was omitted, but participants knew the underlying generative process was the same.

Control Task

The control task was very similar to the main one, with only two differences. (1) When a change occurred, it impacted only one of the two transition probabilities (randomly chosen). (2) During the training session, when subjects were required to report when they detected change points, they also reported which of the two transition probabilities had changed.

Selection of sequences

Each randomly generated sequence was evaluated computationally and carefully selected to ensure that each subject encountered enough target moments during which the models make qualitatively different predictions, and that all sequences were balanced in terms of potential confounds such as streak duration and location. To this end, 4 random sequences of 380 stimuli long (each corresponding to one block) were analyzed computationally with the hierarchical and flat learning models, yielding 4 simulated 'blocks'. The sequences, and associated trial-by-trial transition probability estimates from both models, were concatenated to form a single experimental sequence (of 1520 stimuli). This experimental sequence was then submitted to several selection criteria. First, we assessed whether the sequence contained at least 8 suspicious and 8 non-suspicious 'streaks'. Consecutive repetitions were defined as 'streaks' if they consisted of at least 7 or more stimuli, and started after the 15th stimulus of a block. Streaks were classified as 'suspicious' if they aroused the suspicion of a change in the hierarchical ideal observer. Computationally, this was defined via the confidence in the probability of the observed repetition decreasing at least once during the streak. Following this criterion, even streaks of repetitions are that just *slightly* surprising are considered 'suspicious'. To ensure the effect would be strong enough to observe, we only selected the four blocks if they produced a significant ($p<0.05$) interaction in the hierarchical model, taking into account typical deviations between subjects and the model. We used as an estimate of those deviations the variance of the residuals from a linear regression between subjects' confidence report and the model confidence, in a previous data set[Citation error]. Finally, only selected sequences in which the average duration of suspicious and non-suspicious streaks were approximately identical, and in which there was at least one streak of each type in each block, were selected.

Ideal observer models

The models used in this study are implemented in a Matlab toolbox available on GitHub and described in a previous publication[19]. The model termed "hierarchical" and "flat" here correspond respectively to the hidden Markov model (HMM) and the leaky integrator model in the toolbox. Here, we summarize the essential aspects of those models.

The hierarchical and flat models are both ideal observer models: they use Bayes rule to estimate the posterior distribution of the statistic they estimate, θ_t, based on a prior on this statistic and the likelihood provided by previous observations, $y_{1:t}$ (here, a sequence of As and Bs). Subscripts denote the observation number within a sequence.

$$p(\theta_t|y_{1:t},M) \propto p(y_{1:t}|\theta_t,M)\,p(\theta_t,M) \qquad \text{(Eq 1)}$$

In the main text, the models estimate the transition probabilities between successive stimuli, so that θ is a vector with two elements: θ = [p(A|A), p(B|B)]. Note that those two probabilities suffice to describe all transitions, since the others can be derived as p(B|A) = 1-p(A|A) and p(A|B)=1-p(B|B). In Figure S2 and S4, we also consider variants in which the model estimate another statistic, the frequency of stimuli: θ = p(A). Note that p(B) is simply 1-p(A).

The estimation of θ depends on the assumption of the ideal observer model. The flat model considers that θ is fixed, and evaluates its value based on a leaky count of observations. The internal representation of this model therefore has only one level: θ, the statistic of observations. When the true generative statistic is in fact changing over time, the leakiness of the model allows it to constantly adapt its estimate of the statistic and therefore to cope with changes. If the leakiness is tuned to the rate of change, the estimate can approach optimality (see Fig. S2).

By contrast, the hierarchical model entertains the assumption that θ can abruptly change at any moment. The internal representation of the model therefore has several levels beyond observations: a level characterizing the statistic of observations at a given moment (θ_t) and a level describing the probability that of a change in θ occurs (p_c). Conceivably, there could be higher-order levels describing changes in p_c itself[Citation error]; however this sophistication is unnecessary here and we consider that p_c is fixed.

Flat model

The flat model assumes that the true value of θ is fixed, and it constantly infers its value given the evidence received. Therefore, the likelihood function can be decomposed as follows:

$$\begin{aligned} p(y_{1:t}|\boldsymbol{\theta}) &= p(y_1|\boldsymbol{\theta})\prod_{i=2}^{t} p(y_i|\boldsymbol{\theta}, y_{i-1}) \\ &= \frac{1}{2}\left[\theta_{A|A}^{N_{A|A}}(1-\theta_{A|A})^{N_{B|A}}\right]\left[\theta_{B|B}^{N_{B|B}}(1-\theta_{B|B})^{N_{A|B}}\right] \end{aligned} \qquad \text{(Eq 2)}$$

Where $N_{A|A}(t)$ denotes the number of AA pairs in the sequence $y_{1:t}$. A convenient parametrization for the prior distribution is the beta distribution: $p(\theta) = \text{beta}(\theta_{A|A} \mid N^{prior}_{A|A}, N^{prior}_{B|B})$. This parametrization allows for an intuitive interpretation of $N^{prior}_{A|A}$ and $N^{prior}_{B|B}$ as prior observation counts, and due to its conjugacy with the likelihood function (Eq2), inserting Eq2 into Eq1 yields that the posterior probability of θ is the product of two beta functions:

$$p(\theta|y_{1:t})=\frac{1}{2}Beta(\theta_{A|A}|N_{A|A}+N^{prior}_{A|A},N_{B|A}+N^{prior}_{B|A})Beta(\theta_{B|B}|N_{B|B}+N^{prior}_{B|B},N_{A|B}+N^{prior}_{A|B})$$

(Eq 3)

We consider here that the count of observations (the number of AA and BB pairs) is leaky, so that observations that are further in the past have a lower weight than recent ones. We modeled this leakiness as an exponential decay ω, such that the k-th past stimulus has a weight $e^{-k/\omega}$. Note a perfect integration, in which all observations are given the same weight, corresponds to the special case with ω being infinitely large. Also note that ω*ln(2) corresponds to the "half-life", i.e. the number of observations after which the weight of a past observation is reduced by half.

In the main text, we choose $[N^{prior}_{A|A}, N^{prior}_{B|B}] = [0\ 0]$, for in this case, the mean estimate of the flat model becomes strictly equivalent to the estimate of a "delta rule" as the number of observations increases (see Supplementary text). An alternative choice for the prior is the so-called Laplace-Bayes prior [1 1], which is uninformative in that it gives the same prior probability to any value of θ[20]. This choice is important for Fig. S2, but not for the results in the main text (see Fig S4 and Supplementary text).

Hierarchical model

The hierarchical model evaluates the current value of the generative statistic θ under the assumption that it may change at any new observation with a fixed probability p_c. Note that, would the location of the change point be known, the inference of θ would be simple: one would simply need to count the number of pairs ($N_{A|A}$, $N_{B|B}$, $N_{A|B}$, $N_{B|A}$) since the last change point and apply Eq. 2. However, without knowing the location of change points, one should in principle average the estimates given all possible locations of change points, which is in practice far too large a number. The computation is rendered tractable by the so-called Markov property of the generative process. If one knows θ at time t, then the next observation y_{t+1} is generated with $\theta_{t+1} = \theta_t$ if no change occurred and with another value drawn from the prior distribution otherwise. Therefore, if one knows θ_t, previous observations are not needed to estimate θ_{t+1}. Casting the generative process as a Hidden Markov Model (HMM) allows to compute the joint distribution of θ and observations iteratively, starting from the prior, and updating this distribution by moving forward in the sequence of observations:

$$p(\theta_{t+1},y_{1:t+1})=p(y_{t+1}|\theta_{t+1},y_t)\int p(\theta_t,y_{1:t})p(\theta_{t+1}|\theta_t)d\theta_t \qquad \text{(Eq 4)}$$

This integral can be computed numerically by discretization on a grid. The posterior probability can be obtained by normalizing this joint distribution.

Probability reports and confidence ratings with the models

Both the flat and hierarchical models estimate a full posterior distribution for θ. The probability of the next stimulus (question #1 asked to subjects) was computed from the posterior using Bayes rule:

$$p(y_{t+1}|y_{1:t}) = \int p(y_{t+1}|\boldsymbol{\theta}_{t+1}, y_t)\, p(\boldsymbol{\theta}_{t+1}|y_{1:t})\, d\boldsymbol{\theta}_{t+1} \qquad \text{(Eq 5)}$$

Note that the first term in the integral, the likelihood, is nothing but the relevant transition probability itself (conditioned on the actual previous observation). This integral is therefore simply the mean of the posterior distribution of the relevant transition probability. The confidence in the reported probability estimate was computed as the log-precision of this posterior distribution[6, 15, 18].

Model fit

The flat and the hierarchical models have one free parameter each, respectively ω (the leakiness) and p_c (the prior probability of change point).

The analysis reported in the main text used the parameters that best fit the true probabilities used to generate the sequences of observations presented to subjects. More precisely, for each sequence of observations, we computed the probability of each new observation given the previous ones, as estimated by the models using Eq. 5 and we compared it to the true generative probability. We adjusted the free parameters ω and p_c with grid search to minimize the sum of squared errors (SSE) over all the sequences used for all subjects. The resulting values, ω=20.3 and p_c=0.014 (indeed close to the generative value 1/75) were used to run the simulations presented in the main text.

We also fitted the parameters to the responses of each subject (Fig. S4). For probability estimates, the above grid-search procedure was repeated after replacing generative values with the subject's estimates of probabilities at the moment of questions. For confidence reports, we used a similar procedure; note however that subjects used a bounded qualitative slider to report confidence whereas the model confidence is numeric and unbounded, so that there is not a direct mapping between the two. Therefore, the SSE was computed with the residuals of a linear regression between subject's confidence and the model's confidence.

Statistical analyses

All linear regressions between dependent variables (e.g. probability estimates, confidence ratings) and explanatory variables (optimal estimates of probabilities and confidence, surprise, prediction error, entropy) included a constant and were estimated at the subject level. The significance of regression coefficients was estimated at the group level with t-tests. For multiple regressions,

explanatory variables were z-scored so that regression coefficients can be compared between the variable of a given regression.

Unless stated otherwise, all t-tests are two-tailed.

SUPPLEMENTARY INFORMATION

Supplementary figures and legends

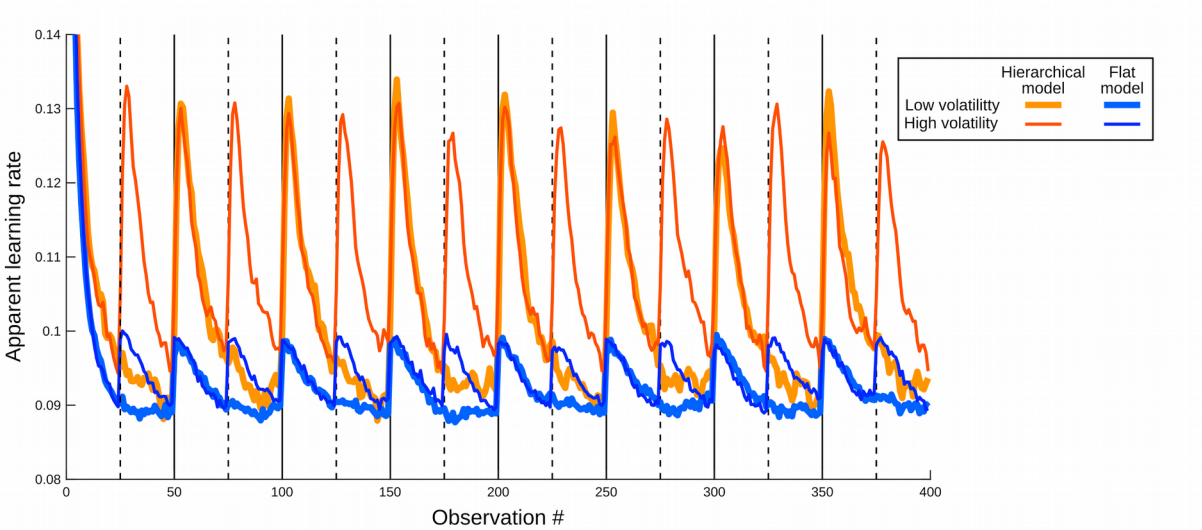

Figure S1

This simulation is inspired by a previous study[2], in which subjects carried out a one-arm bandit task. The reward probability was not fixed but changed abruptly. Subjects had to learn this reward probability through experience in order to optimize their payoff. The authors used different volatility levels (i.e. different numbers of change points) and found that subjects increased their apparent learning rate at higher volatility.

Similarly, we generated two types of sequences: one with low volatility (7 change points, see vertical black plain lines), and one with high volatility (see additional change points, vertical black dashed lines). The sequences were binary (absence or presence of reward) and the reward probability was resampled randomly after each change point. We consider two learning models: a hierarchical model, which estimates the reward rate, taking into account the possibility of change points; and a flat model that computes the reward rate near-optimally based on a fixed leaky count of observations, and a prior count of 1 for either outcome (see Methods). Each model has a single free parameter which we fit to return the best estimate of the actual generative reward probabilities in both the low and high volatility conditions together. Keeping those best fitting parameters equal across both conditions, we simulated sequences and measured the dynamic of the apparent learning rates of the models. The apparent learning rate is defined as the ratio between the current update of the reward estimate (θ_{t+1}-θ_t) and the prediction error leading to this update (y_{t+1}-θ_t). The hierarchical model shows a transient increase in its apparent learning rate whenever a change point occurs, reflecting that it gives more weights to the observations that follow a change point. Such a dynamic adjustment of the apparent learning rate was reported in humans [Nassar 2010]. The flat model showed a similar effect, despite the leakiness of its count being fixed. Note that since there are more change points in the higher

volatility condition (dashed lines), the average learning rates of both models also increase overall with volatility, as previously reported in humans[2].

The lines show mean values across simulations (n=1000); s.e.m. was about the line thickness and therefore omitted.

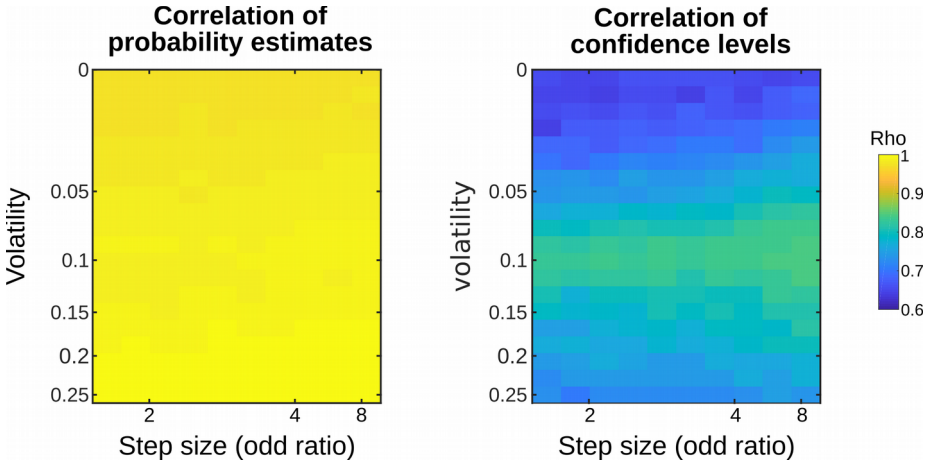

Figure S2

We ran simulations of our experiment to assess which metric (probability estimates or confidence in those estimates) better distinguishes the hierarchical and flat learning models. We simulated binary sequences of observations, using changing transition probabilities between successive items in the sequence. For the sake of generality, we varied the volatility (probability of a change point in a sequence of 380 stimuli) and the step size of those changes (minimum difference, in odd ratio, affecting the transition probabilities). For each combination of volatility and step size, we simulated 100 sequences to achieve stable results. For each volatility-step size combination, we first fit the single free parameter of each model (prior estimate of change point probability p_c in the flat model; and leak factor ω in the flat model) onto the actual generative probabilities of the observed stimuli in the sequences. The resulting parameterized models therefore return their best possible estimate of the hidden regularities, in each volatility-step size condition. We then simulated simulated new sequences (again, 100 per condition) to measure the correlation between the estimated probabilities of stimuli in the sequence between the two models (left), and the correlation between the confidence (log-precision) that those models entertained in those estimates (right). The correlations (Pearson's rho) indicate that probability estimates are nearly indistinguishable between the two models, whereas their confidence levels are much more different.

Note that the volatility level used in the experimental study (0.013) and step size (4) ensure that confidence levels greatly differ between models. Those simulations used prior [1 1] for the flat model, but the results are qualitatively similar with prior [0 0] (see Methods).

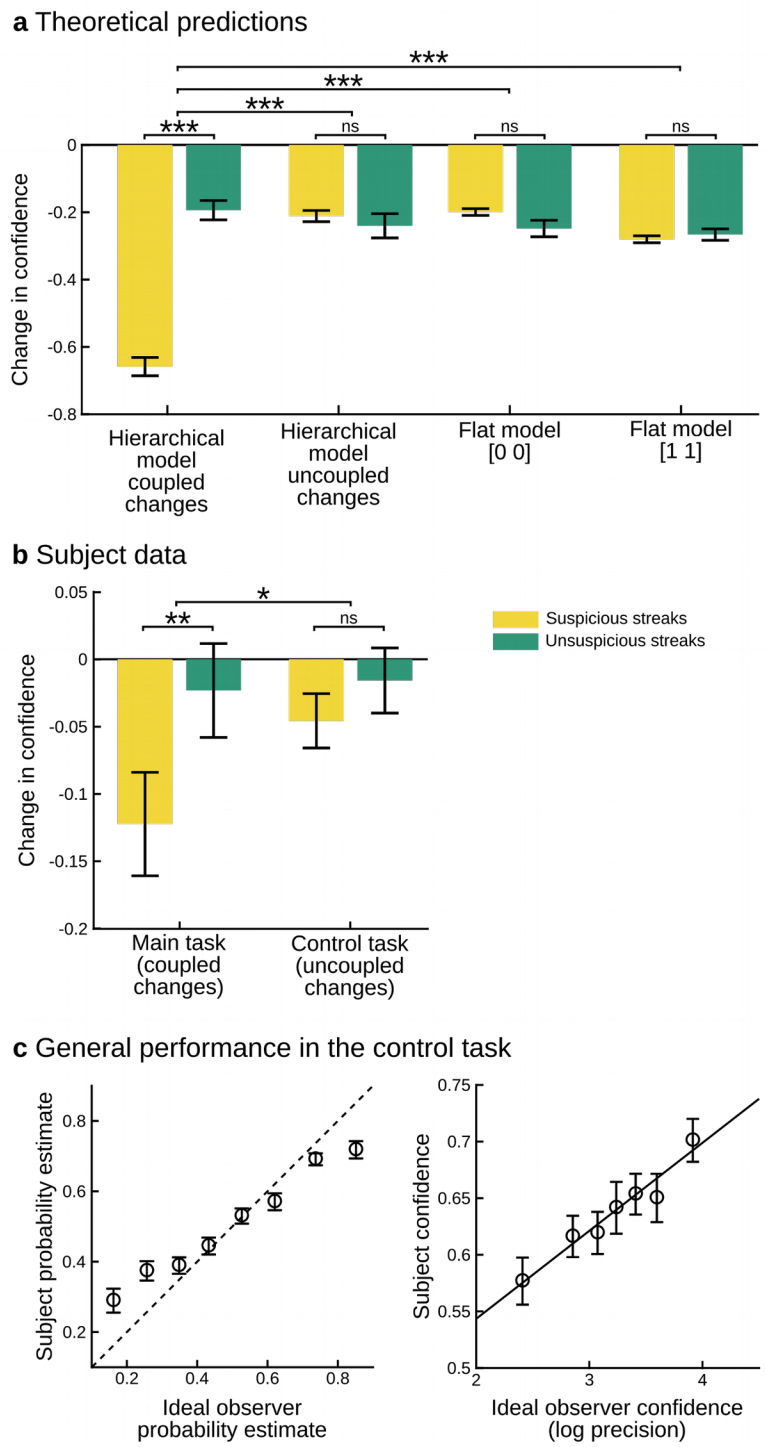

Figure S3

Results of the control experiment, in which change points were uncoupled between the two transition probabilities, thereby abolishing the possibility to infer a change in one transition probability by only observing the other transition type. (**a**) Theoretical predictions for changes in confidence around the target streaks. The optimal hierarchical model for the main task assumes that change points are coupled ("hierarchical model, coupled changes"), which is no longer optimal in the case of uncoupled change points. This model was nevertheless used to identify suspicious and unsuspicious streaks and indeed it showed an effect of streak type on the change in confidence as in

the main task (Fig. 2c) and here in the control task. The optimal hierarchical Bayesian model for this control experiment is similar to this first model, the only difference is that it assumes that change points are uncoupled ("hierarchical model, uncoupled changes"). As expected, this model correctly showed no effect of streak type on the change in confidence. The flat model, by definition, ignores change points and therefore whether they are coupled or uncoupled, as a result they show no effect of streak type (as in the main experiment). (**b**) Subjects showed no difference between streak types (paired difference between suspicious and unsuspicious streaks: 0.03±0.02 s.e.m., t_{20}=1.5, p=0.15), like the hierarchical model for uncoupled changes. This absence of effect is significantly different from the effect found in the main task (difference of paired differences, two-sample t-test, t_{42}=-2.03, p=0.048). The results of the main task are reproduced from Fig. 2c here to facilitate visual comparison. (**c**) One concern may be that subjects showed no effect of streak type here (panel b) because they overall perform badly in the control experiment. This was not the case: linear regression between the optimal hierarchical model for uncoupled change (the ideal observer in this task) and subjects showed tight agreement for both predictions (left, β=0.61±0.06 s.e.m., t_{20}= 10.27, p=2 10^{-9}) and confidence (right, β=0.08±0.01 s.e.m., t_{20}= 6.83, p=1.2 10^{-6}), as in the main task (see Fig. 1b, 1c).

In all panels, data points are mean ± s.e.m across subjects; significance levels correspond to p<0.05 (*), p<0.01 (**), p<0.001(***) in a t-test.

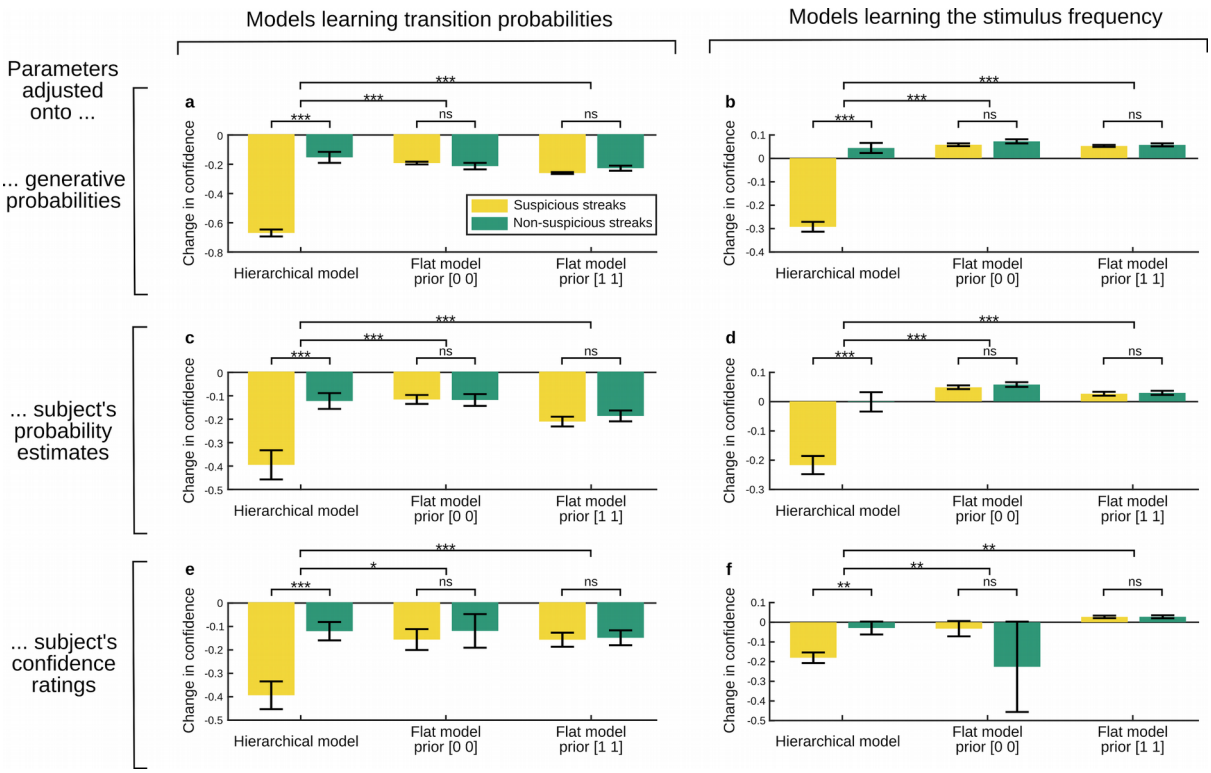

Figure S4

Simulated changes in confidence around the target streaks in the main task. We consider three ideal observers models: the hierarchical model, the flat model with prior [0 0] and with prior [1 1]. Those models tracked either the transition probabilities between successive stimuli (**a**, **c**, **e**) or the frequency of stimuli (**b**, **d**, **f**). The free parameters of the models (the prior probability of change point p_c in the hierarchical model; the leak factor ω in the flat model) were fitted following three procedures: so as to provide the best estimation of the actual generative probabilities of the sequences at all trials (**a**, **b**) or the answers of subjects at the moment of all questions regarding probability estimates (**c**, **d**) or their confidence ratings (**e**, **f**). Panels **a** and **b** therefore show the results of models that were optimized to solve the probability estimation task. By contrast, panels **c**, **d**, **e**, **f**, show the results of models that were optimized to be as close as possible to subjects, which can in principle deviate from **a** and **b**. The rationale for those simulations is to verify that the effect of streak type on the subject's change in confidence (Fig. 2c) is a landmark signature of a hierarchical estimation and cannot be reproduced by other models, even when they are optimized to account for the subject's reports.

Note that panel **a** corresponds to Fig. 2b, expanded with a new case (simulation of a flat model with priors [1 1]). None of the flat models in all plots shows an effect of streak type; some even predict increase (not decrease) in confidence (**b**, **d**, **f**). By contrast, a hierarchical model learning the stimulus frequency (**b**, **d**, **f**) seems more compatible with the subject's data: they indeed predict an effect of streak type. One could therefore wonder whether subjects actually monitor the item

frequency, instead of transition probabilities in the task. Several pieces of evidence argue against this possibility (see Supplementary text). In addition, this possibility would also be incompatible with the results of the control experiment (Fig. S3): a model that estimates a single statistic, namely the item frequency (and the associated confidence), shows the same effect of streak type no matter whether change points are coupled between transition probabilities (main experiment) or uncoupled (control experiment). By contrast, subjects showed different responses in the two tasks (see Fig.S4c).

Data points are mean ± s.e.m of model predictions for each subject; significance levels correspond to $p<0.05$ (*), $p<0.01$ (**), $p<0.001$(***) in paired t-test, excepted for comparisons involving the flat model with prior [0 0] in **f** for which we used a Wilcoxon sign rank test because of an outlier point (see large error bar).

Supplementary results

Relation between the flat model and the delta rule

The equations of the flat model can be re-arranged so as to show the link with a leaky integrator, and hence, to a delta rule. For simplicity we derive those equations for the Bernoulli case (when one seeks to infer the frequency of items in a sequence), noting that the case of transition probabilities between successive items (p(A|A), p(B|B)) is nothing but the Bernoulli case when looking at each transition type separately (from A, from B).

Let's recode the binary sequences as 1s and 0s, and estimate the probability $P_1(n)$ of observing a 1 after a sequence of observations $y_1, \ldots, y_n$. $P_1(n)$ is the mean of a beta distribution, whose parameters are the (leaky) counts of observations and the prior count (cf. Eq. 3 and 5). Using the analytical solution for beta distributions, and recalling that the observation count is leaky, with exponential decay ω:

$$P_1(n)=\frac{N_1+N_1^{prior}}{N_1+N_0+N_1^{prior}+N_0^{prior}}$$

$$P_1(n)=\frac{\sum_{t=1}^{n} y_t\left(e^{-1/\omega}\right)^{n-t}+N_1^{prior}}{\sum_{t=1}^{n}\left(e^{-1/\omega}\right)^{n-t}+N_1^{prior}+N_0^{prior}} \qquad \text{(Eq 6)}$$

$$P_1(n)=c\sum_{t=1}^{n} y_t\left(e^{-1/\omega}\right)^{n-t}+c\,N_1^{prior}$$

Where *c* can be approximated by a constant since $e^{-1/\omega}<1$ and *n* is typically large:

$$c=\frac{1}{\frac{1-e^{-n/\omega}}{1-e^{-1/\omega}}+N_1^{prior}+N_0^{prior}}$$

$$c\approx\frac{1}{\frac{1}{1-e^{-1/\omega}}+N_1^{prior}+N_0^{prior}}$$

A delta-rule with learning rate α reads as follow:

$$P_1(n)=(1-\alpha)^n P_1(0)+\alpha\sum_{t=1}^{n} y_t(1-\alpha)^{n-t}$$

$$P_1(n)\approx\alpha\sum_{t=1}^{n} y_t(1-\alpha)^{n-t} \qquad \text{(Eq 7)}$$

Note that for large *n*, the term before the sum vanishes since (1-α)<0, hence the approximation on the second line.

Comparison of Eq. 6 and 7 shows that the flat model and the delta rule are very similar, the only difference is that the leaky integration constantly adds a prior count in the flat model, whereas in the delta-rule, the impact of the starting point $P_1(0)$, which can be thought of as a prior, vanishes as more observations are accumulated. Stated differently, in the flat model, the inference progressively forgets about previous observations (like in the delta rule) and constantly factors in a prior about the estimated quantity (unlike the delta rule). Note that with a prior $N^{prior}_1 = N^{prior}_0 = 0$, both models become asymptotically identical for large *n*.

Computing the apparent learning rate (the ratio between update $P_1(n)-P_1(n-1)$ and prediction error $y_n-P_1(n-1)$ leading to this update) shows, with a bit of math, that for typical choices of ω, $N^{prior}_1 > 0$ and $N^{prior}_0 > 0$, the apparent learning rate increases whenever two observations in a row are identical, and decreases whenever they differ. Considering that consecutive observations are more likely to differ after a change point in the underlying generative probability, the learning rate of the flat model typically increases, on average, immediately after change points, as can be seen in Fig. S2.

Robustness of the results

Each subject performed four sessions, two with auditory stimuli and two with visual stimuli. We tested the robustness of the linear relations between subject's and optimal values (Fig. 1b) by testing them separately in each session type. We found that the results were replicated in each sensory modality. For probability estimates, in the auditory modality $\beta=0.69\pm0.07$ s.e.m., $t_{22}=9.27$, $p=4.7\ 10^{-9}$; in the visual modality $\beta=0.64\pm0.06$ s.e.m., $t_{22}=10.32$, $p=6.8\ 10^{-10}$. For confidence, in the auditory modality $\beta=0.10\pm0.04$ s.e.m., $t_{22}=2.85$, $p=0.009$; in the visual modality $\beta=0.09\pm0.03$ sem, $t_{22}=2.81$, $p=0.010$. Interestingly, the regression coefficients were correlated across subjects between modalities (probability estimates: $\rho_{23}=0.45$, $p=0.031$; confidence ratings: $\rho_{23}=0.81$, $p=2.6\ 10^{-6}$), suggesting that inference in this task operates at an abstract, amodal level.

We further tested the robustness of the correlation between subject's and optimal values by restricting the regression analysis to a subset of data points, namely, the questions that surround the target streaks (Fig 2C). The significant correlations were replicated on this subset of data for both probability estimates ($\beta=0.57\pm0.07$ s.e.m., $t_{22}=7.6$, $p=1.4\ 10^{-7}$) and confidence ratings ($\beta=0.12\pm0.05$ s.e.m., $t_{22}=2.6$, $p=0.017$).

We also tested the robustness of our central analysis of the effect of streak type on the change in confidence. In the main text, we report a dichotomy between suspicious and non-suspicious streaks, but in reality the 'suspiciousness' of a streak is a matter of degree: the more a streak arouses the suspicion of a change, the larger the decrease in confidence. We therefore regressed the subject's changes in confidence onto the hierarchical model's changes in confidence across all streaks. In order to test whether the hierarchical model or the flat model provides a better account of the subjects' data, we also include the change in confidence of the flat model as a competing explanatory variable in the

multiple regression. Regression coefficients were significant for the hierarchical model (β=0.07±0.02 s.e.m., t_{22}=3.7, p=0.001), but not for the flat model (β=0.01± 0.01 s.e.m., t_{22}=0.8, p=0.44), and the regression coefficients of the hierarchical model were significantly larger than those of the flat model (paired difference of βs=0.06±0.03 s.e.m., t_{22}=2.3, p=0.031), indicating that the hierarchical model provides a significantly better account of the subjects' data.

Normative properties of confidence reports

In the ideal observer model, the answers to questions #1 and #2 (probability estimate and confidence rating) are different readouts of the same posterior distribution, namely its mean and log-precision. If the subjects' answers to those questions also derive from the same inference process, then we expect that subjects who are closer to the optimal probability estimates are also those who are closer to the optimal confidence ratings. We therefore tested whether the linear regression coefficients (βs) linking subjects and the optimal hierarchical model were correlated between probability estimates and confidence ratings. The between-subject correlations was indeed significant: ρ_{23}=0.53, p=0.009.

We tested for further normative properties of the subject's confidence ratings. Several factors, notably factors pertaining to first-order estimates, are expected to impact confidence ratings in this task from a normative viewpoint. We first show that those factors indeed impact the optimal confidence levels at the moments of questions during the task, and then report a similar analysis for the subject's confidence ratings. Optimal confidence levels were entered into a multiple regression model which included the estimated probability itself, the entropy of this probability (which quantifies the estimated unpredictability of the next stimulus, it culminates when the estimated probability is 0.5), and the extent to which the current observation deviates from the previous estimate, as quantified by the surprise (negative log likelihood of the observations, [Shannon]) and the prediction error (one minus the likelihood of the current observation). Optimal confidence was lower when the estimated entropy was higher (β=-0.051±0.008, t_{22}=-6.1, p=4.2 10^{-6}), lower when surprise was larger (β=-0.142±0.026, t_{22}=-5.4, p=1.8 10^{-5}) and when prediction error was larger (β=-0.128±0.030, t_{22}=-4.2, p=3.5 10^{-4}).

To analyze subjects' confidence ratings, we added other explanatory variables to this multiple linear regression model, which correspond to subject's estimates: the subject's probability estimate, and the entropy corresponding to this estimate. Note that questions are asked only occasionally, so that we don't know the probability estimate of the subject at the *previous trial*, and therefore, we cannot compute the subject's surprise and prediction error elicited by the last observation. Subject's confidence was lower when the entropy of his estimate was higher (β=-0.125±0.009, t_{22}=-14.5, p=1.0 10^{-12}) and when the optimal surprise level was higher (β=-0.090±0.022, t_{22}=-4.1, p=5.0 10^{-4}).

Another aspect of subjects' accuracy is that their report of confidence is specific to the relevant statistics. In the experiment, subjects monitor two transition probabilities, there are therefore two

confidence levels, each being attached to one transition type. Questions asked subjects to estimate the likelihood of the next stimulus, which depends on only one of the two transition probabilities: the one that is relevant given the identity of the previous stimulus at the moment of the question. We estimated a multiple linear regression in which the subject's confidence ratings was regressed onto both the optimal relevant confidence levels, and the optimal irrelevant confidence level (those attached to the irrelevant transition). The regression coefficients corresponding to the relevant confidence levels were significant (β=0.039±0.014 s.e.m., t_{22}= 2.8, p=0.011), those for the irrelevant confidence were not (β=0.001±0.008 s.e.m., t_{22}=0.1, p=0.89) and the difference between the two was significant (paired difference of βs=0.038±0.018 s.e.m., t_{22}=2.0, p=0.027, one-tailed test). A different multiple linear regression indicates, in addition, that confidence ratings are selectively modulated by the (optimal) entropy of the relevant transition probability, as opposed to the irrelevant one (paired difference of βs=-0.020±0.007 s.e.m., t_{22}=-2.9, p=0.008). Optimal confidence levels show the same effect: replacing the subjects' confidence ratings in this latter regression with the optimal confidence levels also reveals a significant difference (paired difference of βs=-0.050±0.016, t_{22}=-3.2, p=0.0045). By contrast, such a difference is not observed when replacing the subjects' confidence ratings with the optimal confidence of a model that monitors solely the frequency of items (paired difference of βs=0.005±0.008, t_{22}=0.6, p=0.54). Together, those results indicate that subjects reported specifically the confidence attached to the transition probability relevant at the moment of the question.

Supplementary references

18. Meyniel, F., Dehaene, S. *Proc Natl Acad Sci USA* **114**,3859-3868 (2017)
19. Meyniel, F., Maheu, M., Dehaene S., PloS Comput. Biol. **12**: e1005260 (2016)
20. Gelman, Bayesian Data Analysis, Third Edition. CRC Press (2013)