Human Inference in Changing Environments with Temporal Structure

Arthur Prat-Carrabin, Robert Wilson, Jonathan Cohen, Rava Azeredo da Silveira

Human Inference in Changing Environments With Temporal Structure

Arthur Prat-Carrabin[1,†], Robert C. Wilson[3,‡], Jonathan D. Cohen[3], and Rava Azeredo da Silveira[1,2,3,4*]

[1]Laboratoire de Physique de l'École Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris, 75005 Paris, France

[2]IOB, Faculty of Science, University of Basel, Basel, Switzerland

[3]Princeton Neuroscience Institute, Princeton University, Princeton, USA

[4]Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel

[†]*Present address:* Department of Economics, Columbia University, USA

[‡]*Present address:* Department of Psychology and Cognitive Science Program, University of Arizona, Tucson, USA

[*]*For correspondence:* rava@ens.fr

Abstract

To make informed decisions in natural environments that change over time, humans must update their beliefs as new observations are gathered. Studies exploring human inference as a dynamical process that unfolds in time have focused on situations in which the statistics of observations are history-independent. Yet temporal structure is everywhere in nature, and yields history-dependent observations. Do humans modify their inference processes depending on the latent temporal statistics of their observations? We investigate this question experimentally and theoretically using a change-point inference task. We show that humans adapt their inference process to fine aspects of the temporal structure in the statistics of stimuli. As such, humans behave qualitatively in a Bayesian fashion, but, quantitatively, deviate away from optimality. Perhaps more importantly, humans behave suboptimally in that their responses are not deterministic, but variable. We show that this variability itself is modulated by the temporal statistics of stimuli. To elucidate the cognitive algorithm that yields this behavior, we investigate a broad array of existing and new models that characterize different sources of suboptimal deviations away from Bayesian inference. While models with 'output noise' that corrupts the response-selection process are natural candidates, human behavior is best described by sampling-based inference models, in which the main ingredient is a compressed approximation of the posterior, represented through a modest set of random samples and updated over time. This result comes to complement a growing literature on sample-based representation and learning in humans.

Keywords: Bayesian inference, change-points, non-Poisson statistics, online inference, response variability, computational modeling

Human Inference in Changing Environments With Temporal Structure

In a variety of inference tasks, human subjects use sensory cues as well as prior information in a manner consistent with Bayesian models. In tasks requiring the combination of a visual cue (such as the shape, position, texture, or motion of an object) with a haptic (Battaglia, Kersten, & Schrater, 2011; Ernst & Banks, 2002), auditory (Battaglia, Jacobs, & Aslin, 2003), proprioceptive (van Beers, Sittig, & Denier, 1999), or a secondary visual cue (Hillis, Watt, Landy, & Banks, 2004; Jacobs, 1999; Knill, 2007), human subjects weigh information coming from each cue according to its uncertainty, in agreement with an optimal, probabilistic approach. Moreover, subjects appear also to integrate optimally prior knowledge on spatial (Körding & Wolpert, 2004, 2006) and temporal (Jazayeri & Shadlen, 2010; Miyazaki, Nozaki, & Nakajima, 2005) variables relevant to inference, in line with Bayes' rule.

The Bayesian paradigm hence offers an elegant and mathematically principled account of the way in which humans carry inference in the presence of uncertainty. In most experimental designs, however, successive trials are unrelated to each other. Yet, in many natural situations, the brain receives a stream of evidence from the environment: inference, then, unfolds in time. Moreover, natural mechanisms introduce sophisticated temporal statistics in the course of events (e.g., rhythmicity in locomotion, day-night cycles, and various structures found in speech). Are these temporal dynamics used by the brain to refine its *online* inference of the state of the environment?

Furthermore, most studies that support a Bayesian account of human inference discuss average behaviors of subjects, and, thereby, side-step the issue of the variability in human responses. While an optimal Bayesian model yields a unique, deterministic action in response to a given set of observations, human subjects exhibit noisy, and thus suboptimal, responses. Methods commonly used to model response variability, such as 'softmax' and probability-matching response-selection strategies, or, more recently, stochastic inference

processes, correspond to different forms of departure from Bayesian optimality. One would like to identify the nature of the deviations from Bayesian models that can account for the observed discrepancies from optimality in human behavior.

To explore these questions, we use an online inference task based on a 'change-point' paradigm, i.e., with random stimuli originating from a hidden state that is subject to abrupt, occasional variations, which are referred to as 'change points'. A growing theoretical and experimental literature examines inference problems for this class of signals (Adams & MacKay, 2007; Brown & Steyvers, 2009; Fearnhead & Liu, 2007; Gallistel, Krishan, Liu, Miller, & Latham, 2014; Glaze, Filipowicz, Kable, Balasubramanian, & Gold, 2018; Glaze, Kable, & Gold, 2015; Khaw, Stevens, & Woodford, 2017; Nassar et al., 2012; Nassar, Wilson, Heasly, & Gold, 2010; Piet, El Hady, & Brody, 2018; Radillo, Veliz-Cuba, Josić, & Kilpatrick, 2017, 2019; Veliz-Cuba, Kilpatrick, & Josić, 2016; Wilson, Nassar, & Gold, 2010, 2013). All these studies, with the exception of the work of Fearnhead and Liu (2007), focus on the history-independent case of random change points that obey Poisson temporal statistics. Such problems are characterized by the absence of temporal structure: the probability of occurrence of a change point does not depend on the realization of past change points. Piet et al. (2018), Wilson et al. (2010) and Glaze et al. (2018) extend their studies beyond this simple framework by considering 'hierarchical-Poisson' models in which the change probability is itself subject to random variations; but, here also, the occurence of a change point does not depend on the timing of earlier change points. The experimental studies among the ones cited above have investigated the way in which human subjects and rodents infer hidden states, and whether they learn history-independent change probabilities.

Because of the pervasiveness of temporal structure in natural environments, we decided to study human inference in the presence of 'history-dependent' statistics in which the occurrence of a change point depends on the timing of earlier change points. This introduces considerable complexity in the optimal inference model (as the hidden state is

no longer Markovian), and serves as a first step toward a more ecological approach to human inference. For the purpose of comparison, we consider two different statistics of change points: the first one is the Poisson statistics commonly used in earlier studies; the second is the simplest non-Markovian statistics, in which the probability of a change point is a function of the timing of the preceding change point. This setup allows us to examine the effect of the latent temporal structure on both human behavior and model responses.

In these two contrasting conditions, the behavior of the Bayesian model and that of human subjects exhibit both similarities and discrepancies. A salient departure from optimality exhibited by subjects is the variability in their responses. What is more, the shape of the distribution of responses is not constant, but, rather, subject to modulations during the course of inference. The standard deviation and skewness of the empirical response distribution are correlated with that of the optimal, Bayesian posterior; this suggests that the randomness in subjects' responses does not reflect some 'passive' source of noise but is in fact related to the uncertainty of the Bayesian observer.

To account for this non-trivial variability in human responses and other deviations from optimality, we investigate in what ways approximations of the Bayesian model alter behavior, in our task. The optimal estimation of a hidden state can be split into two steps: Bayesian posterior inference (computing optimally the belief distribution over the state space) and optimal response selection (using the belief distribution to choose the response that maximizes the expected reward). Suboptimal models introduce systematic errors or stochastic errors in the inference step or in the response-selection step, or in both, thus impacting behavior. Models discussed in the change-point literature, along with new models we introduce, provide a wide range of such deviations from optimality, which we compare to experimental data. This allows us to assess how different sources of suboptimality impact behavior, and to what extent they can capture the salient features in human behavior.

The paper is outlined as follows. We first present the main aspects of our task, in which

subjects observe a visual stimulus and infer an underlying, changing, hidden state. The susceptibility of subjects to a new stimulus is shown to differ appreciably between the two conditions (with and without latent temporal structure), and to adapt to the statistics of change points. We then analyze the variability in the subjects' responses, and show how it is modulated over the course of inference. After deriving the optimal, Bayesian solution of the inference problem in the context of our task, we examine its behavior in comparison with experimental data. We then turn to investigating a broad family of suboptimal models. In particular, motivated by the form of the variability present in our human data, we examine stochastic perturbations in both the inference step and in the response-selection step. These models reflect *different forms* of *sampling*: model subjects either perform inference using samples of probability distributions or select responses by sampling; the former option includes models with limited memory as well as sequential Monte Carlo (particle-filter) models. Finally, we discuss model fitting, from which we conclude that humans carry out stochastic approximations of the optimal Bayesian calculations through sampling-based inference (rather than sampling-based response selection).

Our observations confirm and extend the results reported in the change-point literature on human inference in the context of Poisson statistics, by exploring a more ecological (Anteneodo & Chialvo, 2009; Campione & Véronis, 2002; Griffin, West, & West, 2000; Hausdorff, Peng, Ladin, Wei, & Goldberger, 1995; Low, Grabe, & Nolan, 2000; Nakamura et al., 2007; Nakamura et al., 2008; Nunes Amaral et al., 2004; Ramus, Nespor, & Mehler, 1999), non-Poisson, temporally structured environment. Likewise, our results come to complement those of a number of studies on perception and decision-making that also investigate inference from stimuli with temporal statistics (Ghose & Maunsell, 2002; Janssen & Shadlen, 2005; Jazayeri & Shadlen, 2010; Li & Dudman, 2013; Miyazaki et al., 2005; ten Oever, Schroeder, Poeppel, van Atteveldt, & Zion-Golumbic, 2014). Our experimental results demonstrate that humans learn implicitly the temporal statistics of stimuli. Moreover, our work highlights the variability ubiquitous in behavioral data, and

shows that it itself exhibits structure: it depends on the temporal statistics of the signal, and it is modulated over the course of inference. We find that a model in which the Bayesian posterior is approximated with a set of samples captures the behavioral variability during inference. This proposal adds to the growing literature on cognitive 'sample-based representations' of probability distributions (Gershman, Vul, & Tenenbaum, 2012; Goodman, Tenenbaum, Feldman, & Griffiths, 2008; Moreno-Bote, Knill, & Pouget, 2011; Vul, Goodman, Griffiths, & Tenenbaum, 2014). Our results suggest that the brain carries out complex inference by manipulating a modest number of samples, selected as a low-dimensional approximation of the optimal, Bayesian posterior.

Results

Behavioral task, and history-independent vs. history-dependent stimuli

In our computer-based task, subjects are asked to infer, at successive trials, t, the location, on a computer screen, of a hidden point, the state, s_t, based on an on-screen visual stimulus, x_t, presented as a white dot on a horizontal line (Fig. 1A,B). Subjects can only observe the white dots, whose positions are generated around the hidden state according to a likelihood probability, $g(x_t|s_t)$ (Fig. 1C,E, blue distribution). The state itself, s_t, follows a change-point process, i.e., it is constant except when it 'jumps' to a new location, which happens with probability q_t (the 'hazard rate' or 'change probability'). The dynamics of change points are, hence, determined by the change probability, q_t. To examine the behavior of models and human subjects in different 'environments', we choose two kinds of signals which differ in their temporal structure. History-independent (**HI**) signals are memoryless, Poisson signals: q_t is constant and equal to 0.1. Consequently, the intervals between two change points last, on average, for 10 trials, and the distribution of these intervals is geometric (Fig. 1D, blue bars). Conversely, history-dependent (**HD**) signals are characterized by temporal correlation. Change points also occur every 10 trials, on average, but the distribution of the duration of inter-change-point intervals is peaked

around 10. This corresponds to a change probability, q_t, that is an increasing function of the number of trials since the last change point — a quantity referred to as the 'run-length', τ_t. We thus denote it by $q(\tau_t)$. In HD signals, change points occur in a manner similar to a 'jittered periodic' process, though the regularity is not readily detected by subjects.

When a change point occurs, the state randomly jumps to a new state, s_{t+1}, according to a state transition probability, $a(s_{t+1}|s_t)$ (Fig. 1C,E, green distribution). The likelihood, g, and the state transition probability, a, overlap, thus allowing for ambiguity when a new stimulus is viewed: is it a random excursion about the current state, or has the state changed? At each trial, subjects click with a mouse to give their estimate, $\hat{s}_t$, of the state. The reward they receive for each response is a decreasing function, R, of the distance between the state and the estimate, $|\hat{s}_t - s_t|$: one reward point if the estimate falls within a given, short distance from the state, 0.25 point if it falls within twice that distance, and 0 point otherwise (Fig. 1E). The task is presented as a game to subjects: they are told that someone is throwing snowballs at them. They cannot see this hidden person (whose location is the state, s_t), but they observe the snowballs as white dots on the screen (the stimulus, x_t). After several tutorial runs (in some of which the state is shown), they are instructed to use the snowballs to guess the location of the person (i.e., produce an estimate, $\hat{s}_t$). Additional details on the task are provided in Methods.

Learning rates adapt to the temporal statistics of the stimulus

A typical example of a subject's responses is displayed in Fig. 2A. To describe the data, we focus, throughout this paper, on three quantities: the learning rate, defined as the ratio of the 'correction', $\hat{s}_{t+1} - \hat{s}_t$, to the 'surprise', $x_{t+1} - \hat{s}_t$; the repetition propensity, defined as the proportion of trials in which the learning rates vanishes ($\hat{s}_{t+1} = \hat{s}_t$); and the standard deviation of the responses of the subjects. The learning rate represents a normalized measure of the susceptibility of a subject to a new stimulus. If the new

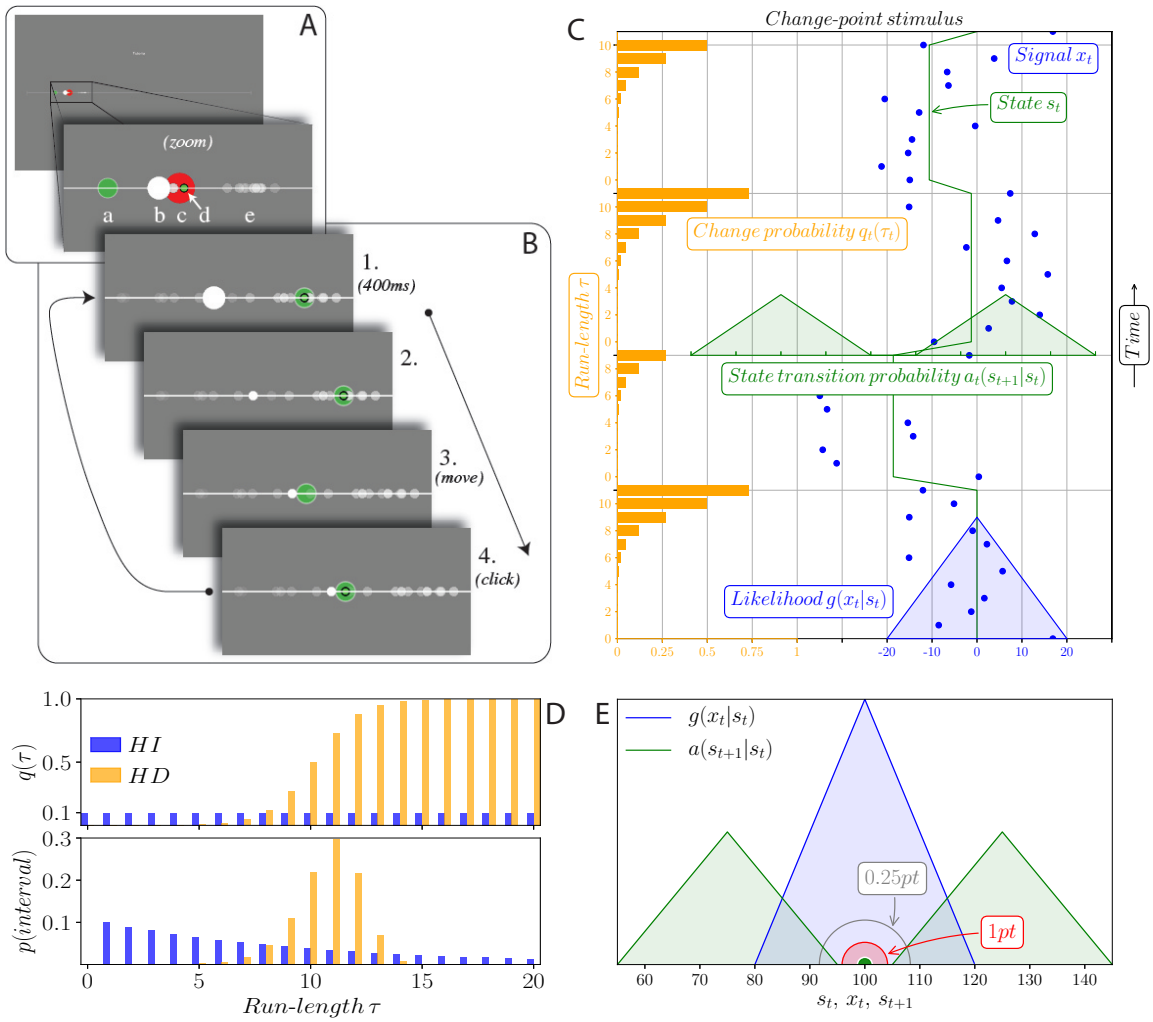

Figure 1. **Inference task and change probability, q, in the HI and HD conditions.** **A.** The various elements in the task appear on a horizontal white line in the middle of a gray screen. **a:** subject's pointer (green disk). **b:** new stimulus (white disk). **c:** state (red disk, only shown during tutorial runs). **d:** position of subject's previous click (green dot). **e:** for half of subjects, previous stimuli appear as dots decaying with time. **B.** Successive steps of the task: **1, 2:** a new stimulus is displayed; to attract the subject's attention, it appears as a large, white dot for 400ms, after which it becomes smaller. **3:** the subject moves the pointer. **4:** The subject clicks to provide an estimate of the position of the state. After 100ms, a new stimulus appears, initiating the next trial. **C.** The position of the stimulus on the horizontal axis, x_t, is generated randomly around the current state, s_t, according to the triangle-shaped likelihood function, $g(x_t|s_t)$. The state itself is constant except at change points, at which a new state, s_{t+1}, is generated around s_t from the bimodal-triangle-shaped transition probability, $a(s_{t+1}|s_t)$. The run-length, τ_t, is defined as the number of trials since the last change point. Change points occur with the change probability $q(\tau_t)$ (orange bars), which depends on the run-length in the HD condition (depicted here). *(Continued on the following page.)*

Figure 1. **D.** Top panel: change probability, $q(\tau)$, as a function of the run-length, τ. It is constant and equal to 0.1 in the HI condition, while it increases with the run-length in the HD condition. Consequently, the distribution of intervals between two consecutive change points (bottom panel) is geometric in the HI condition whereas it is peaked in the HD condition; in both conditions, the average duration of inter-change-point intervals is 10. **E.** Compared extents of the likelihood, $g(x_t|s_t)$ (green), the state transition probability, $a(s_{t+1}|s_t)$ (blue), the 'shot' resulting from a click (green dot), and the radii of the 1-point (red disk) and 0.25-point (grey circle) reward areas. A shot overlapping the red (gray) area yields 1 (0.25) point.

estimate, $\hat{s}_{t+1}$, is viewed as a weighted average of the previous estimate, $\hat{s}_t$, and the new stimulus, x_{t+1}, the learning rate is the weight given to x_{t+1}. A learning rate of 0 means that the subject has not changed its estimate upon observing the new stimulus; a learning rate of 0.5 means that the new estimate is equidistant from the previous estimate and the new stimulus; and a learning rate of 1 means that the new estimate coincides with the new stimulus, and the past is ignored (Fig. 2A).

Our data show that for human subjects the learning rate is not constant, and can vary from no correction at all (learning rate ≈ 0) to full correction (learning rate ≈ 1). We investigated how the average learning rate behaved in relation to the run-length, in the HI and HD conditions. As the run-length is not directly accessible to subjects, in our analyses we used the empirical run-length, $\tilde{\tau}$, a similar quantity derived from the subjects' data (see Methods). Unless otherwise stated, we focus our analyses on cases in which the surprise, $x_{t+1} - \hat{s}_t$, is in the [8,18] window, in which there is appreciable ambiguity in the signal.

A first observation emerging from our data is that the learning rate changes with the run-length, in a quantitatively different fashion depending on the condition (HI or HD). In the HI condition, learning rates at short run-length ($\tilde{\tau} \in [5, 6]$) are significantly higher than at long run-length ($\tilde{\tau} \in [9, 10]$), i.e., the learning rate decreases with run-length (Fig. 2B, blue bars). In the HD condition, the opposite occurs: learning rates are significantly higher at long run-lengths (Fig. 2B, orange bars), indicating that subjects modify their inference depending on the temporal structure of the signal. In addition, at short run-lengths,

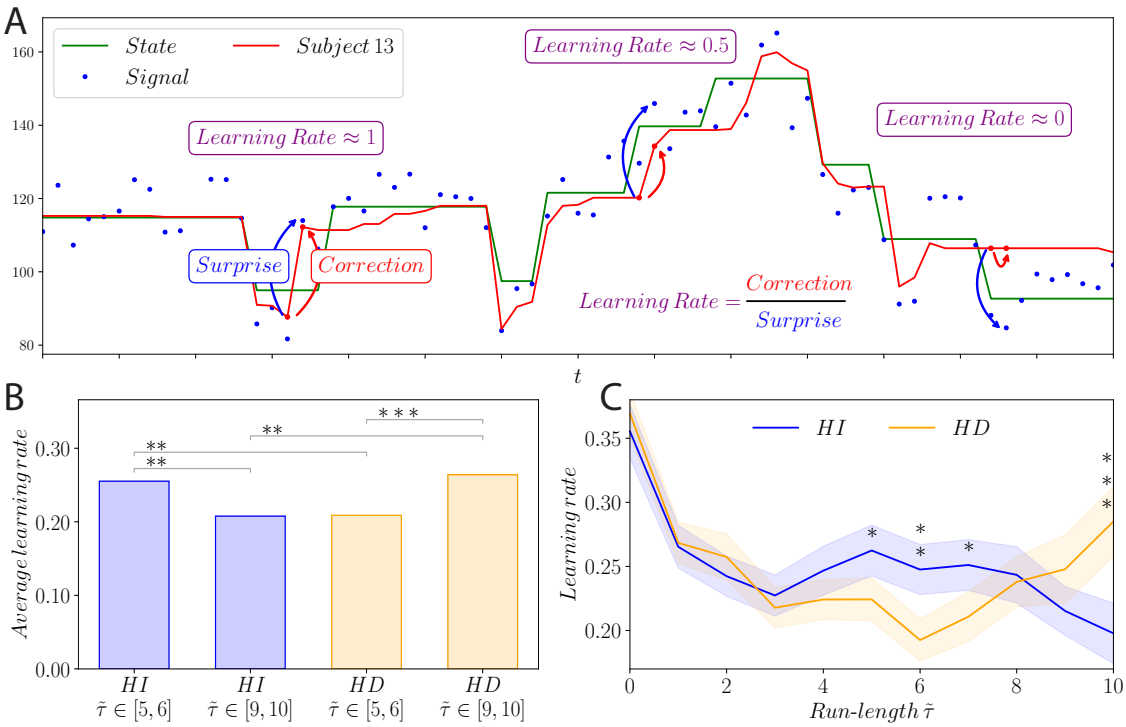

Figure 2. **Human learning rates depends on the temporal statistics (HI or HD) of the stimulus. A.** Illustration of the learning rate using a sample of a subject's responses (red line). The 'surprise' (blue arrow) is the difference, $x_{t+1} - \hat{s}_t$, between the estimate at trial t, $\hat{s}_t$ (red), and the new stimulus at trial $t+1$, x_{t+1} (blue). The 'correction' (red arrow) is the difference between the estimate at trial t and the estimate at trial $t+1$, $\hat{s}_{t+1} - \hat{s}_t$. The 'learning rate' is the ratio of correction to surprise. **B.** Average learning rates in HI (blue) and HD (orange) conditions, at short run-lengths ($\tilde{\tau} \in [5, 6]$) and long run-lengths ($\tilde{\tau} \in [9, 10]$). In the HD condition the change probability increases with the run-length, which advocates for higher learning rates at long run-lengths. **C.** Average learning rates in HI (blue) and HD (orange) conditions, vs. run-length $\tilde{\tau}$. Shaded bands indicate the standard error of the mean. **B**,**C.** Stars indicate p-values of one-sided Welch's t-tests, which do not assume equal population variance. Three stars: $p < 0.01$; two stars: $p < 0.05$; one star: $p < 0.1$. Bonferroni-Holm correction (Holm, 1978) is applied in panel **B**.

learning rates are significantly lower in the HD condition than in the HI condition; this suggests that subjects take into account the fact that a change is less likely at short run-lengths in the HD condition. The opposite holds at long run-lengths: HD learning rates are markedly larger than HI ones (Fig. 2B).

Inspecting the dependence of the learning rate on the run-length (Fig. 2C), we note that the HD learning-rate curve adopts a 'smile shape', unlike the monotonic curve in the HI condition. (A statistical analysis confirms that these curves have significantly different

shapes; see Methods.) The HI curve is consistent with a learning rate that simply decreases as additional information is accumulated on the state. In the HD condition, initially the learning rate is suppressed, then boosted at longer run-lengths, reflecting the modulation in the change probability.

These observations demonstrate that subjects adapt their learning rate to the run-length, and that in the HD condition subjects make use of the temporal structure in the signal. These results are readily intuited: shortly after a change point, the learning rate should be high, as little is known about the new state, while at longer run-lengths the learning rate should tend to zero as the state is more and more precisely inferred. This decreasing behavior is observed, but only in the HI condition. The HD condition introduces an opposing effect: as the run-length grows, new stimuli are increasingly likely to divulge the occurrence of a new state, which advocates for adopting a higher learning rate. This tradeoff is reflected in our data in the 'smile shape' of the HD learning-rate curve (Fig. 2C; these trends subsist at longer run-lengths, see Supplementary Fig. B1.). The increase in learning rate at long run-lengths is reminiscent of the behavior of a driver waiting at a red light: as time passes, the light is increasingly likely to turn green; as a result, the driver is increasingly susceptible to react and start the car.

Human repetition propensity

A closer look at the data presented in the previous section reveals that in a number of trials the learning rate vanishes, i.e., $\hat{s}_{t+1} = \hat{s}_t$. The distribution of the subjects' corrections, $\hat{s}_{t+1} - \hat{s}_t$, exhibits a distinct peak at zero (Fig. 3A). In other words, in a fraction of trials, subjects click twice consecutively on the same pixel. We call such a response a 'repetition', and the fraction of repetition trials the 'repetition propensity'. The latter varies with the run-length: it increases with τ in both HI and HD conditions, before decreasing in the HD condition for long run-lengths (Fig. 3B).

What may cause the subjects' repetition behavior? The simplest explanation is that,

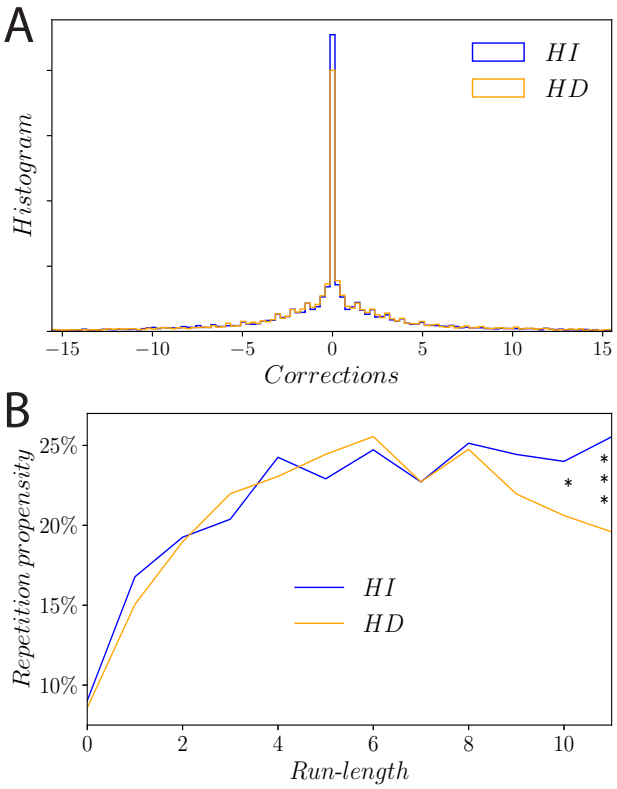

Figure 3. **Human repetition propensity depends on the temporal statistics, and dynamically on the run-length. A.** Histogram of subject corrections (difference between two successive estimates, $\hat{s}_{t+1} - \hat{s}_t$), in the HI (blue) and HD (orange) conditions. The width of bins corresponds to one pixel on the screen, thus the peak at zero represents the repetition events ($\hat{s}_{t+1} = \hat{s}_t$). **B.** Repetition propensity, i.e., proportion of occurrences of repetitions in the responses of subjects, as a function of run-length, in the HI (blue) and HD (orange) conditions. Stars indicate p-values of Fisher's exact test of equality of the repetition propensities between the two conditions, at each run-length.

after observing a new stimulus, a subject may consider that the updated best estimate of the state lands on the same pixel as in the previous trial. The width of one pixel in arbitrary units of our state space is 0.28. As a comparison, the triangular likelihood, g, has a standard deviation, σ_g, of 8.165. An optimal observer estimating the center of a Gaussian density of standard deviation σ_g, using 10 samples from this density, comes up with a posterior density with standard deviation $\sigma_g/\sqrt{10} \approx 2.6$. Therefore, after observing even 10 successive stimuli, the subjects' resolution is not as fine as a pixel (it is, in fact, 10 times coarser). This indicates that the subjects' repetition propensity is higher than the optimal one (the behavior of the optimal model, presented below, indeed exhibits a lower average

repetition propensity than that of the subjects). Another possible explanation is that even though the new estimate falls on a nearby location, a motor cost prohibits a move if it is not sufficiently extended to be 'worth it' (Morasso, 1981; Rigoux & Guigon, 2012; Shadmehr, 2009; Wolpert, 1997). A third, heuristic explanation is that humans are subject to a 'repetition bias' according to which they repeat their response irrespective of their estimate of the state.

Regardless of its origin, the high repetition propensity in data raises the question of whether it dominates the behavior of the average learning rate. As a control, we excluded all occurrences of repetitions in subjects' data and carried out the same analyses on the truncated dataset. We reached identical conclusions, namely, significant discrepancies between the HI and HD learning rates at short and long run-lengths, albeit with, naturally, higher average rates overall (see Supplementary Fig. B2).

The variability in subjects' responses evolves over the course of inference

In the previous two sections, we have examined two aspects of the distribution of responses: the average learning rate and the probability of response repetition. We now turn to the variability in subjects' responses. Although all subjects were presented with identical series of stimuli, x_t, their responses at each trial were not the same (Fig. 4A). This variability appears in both HI and HD conditions. The distribution of responses around their averages at each trial has a width comparable to that of the likelihood distribution, $g(x_t|s_t)$ (Fig. 4B). More importantly, the variability in the responses (as measured by the standard deviation) is not constant, but decreases with successive trials following a change point, at short run-lengths (Fig. 4C). Comparing the HI and HD conditions, we observe that for run-lengths shorter than 7, the standard deviation in the HD condition is significantly lower than that in the HI condition. At longer run-lengths, the two curves cross and the variability in the HD condition becomes significantly higher than in the HI condition. The HD curve adopts, again, a 'smile shape' (Fig. 4C). What is

the origin of the response variability? Because it changes with the run-length and the HI vs. HD condition, it cannot be explained merely be the presence of noise independent from the inference process, such as pure motor noise. In order to encompass human behavior in a theoretical framework and to investigate potential sources of this inference-dependent variability, we start by comparing the recorded behavior with that of an optimal observer.

Optimal estimation: Bayesian update and maximization of expected reward

We derive the optimal solution for the task of estimating the hidden state, s_t, given random stimuli, x_t. The first step (the 'inference step') is to derive the optimal posterior distribution over the state, s_t, using Bayes' rule. Because the state is a random variable coupled with the run-length, τ_t, another random variable, it is convenient to derive the Bayesian update equation for the (s_t, τ_t) pair (more precisely, the (s_t, τ_t) pair verifies the Markov property, whereas s_t alone does not, in the HD condition). We denote by $x_{1:t}$ the string of stimuli received between trial 1 and trial t, and by $p_t(s, \tau|x_{1:t})$ the probability distribution over (s, τ), at trial t, after having observed the stimuli $x_{1:t}$. At trial $t+1$, Bayes' rule yields $p_{t+1}(s, \tau|x_{1:t+1}) \propto g(x_{t+1}|s)p_{t+1}(s, \tau|x_{1:t})$. Furthermore, we have the general transition equation,

$$p_{t+1}(s, \tau|x_{1:t}) = \sum_{\tau_t} \int_{s_t} p_{t+1}(s, \tau|s_t, \tau_t) p_t(s_t, \tau_t|x_{1:t}) \mathrm{d}s_t, \tag{1}$$

given by the change-point statistics. As the transition probability, $p_{t+1}(s, \tau|s_t, \tau_t)$, can be expressed using $q(\tau_t)$ and $a(s|s_t)$ (see Methods for details), we can reformulate the update equation as

$$\begin{aligned} p_{t+1}(s, \tau|x_{1:t+1}) = \frac{1}{Z_{t+1}} g(x_{t+1}|s) \Bigg[& \mathbb{1}_{\tau=0} \sum_{\tau_t} q(\tau_t) \int_{s_t} a(s|s_t) p_t(s_t, \tau_t|x_{1:t}) \mathrm{d}s_t \\ & + \mathbb{1}_{\tau>0} \left(1 - q(\tau - 1)\right) p_t(s, \tau - 1|x_{1:t}) \Bigg], \end{aligned} \tag{2}$$

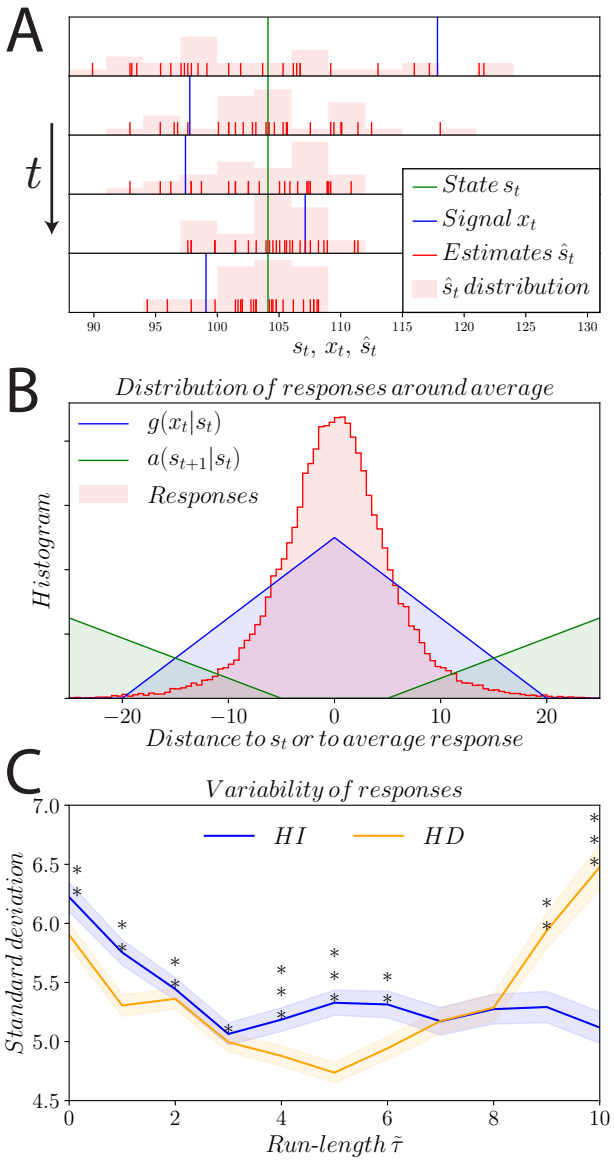

Figure 4. **The variability in subjects' responses is modulated during inference, and these modulations depend on the temporal statistics of the stimulus. A.** Responses of subjects in an example of 5 consecutive stimuli. In this example, there is no change point and the state (green) is constant. At each trial (from top to bottom), subjects observe the stimuli (blue) and provide their responses (red bars). A histogram of the locations of the responses is obtained by counting the number of responses in bins of width 3 (light red). **B.** Distribution of the responses of subjects around their average (red), compared to the likelihood, g (blue), and the state transition probability, a (green). **C.** Standard deviation of the responses of subjects vs. run-length, $\tilde{\tau}$, in the HI (blue) and HD (orange) conditions. Stars indicate p-value of Levene's test of equality of variance between the two conditions, at each $\tilde{\tau}$. Shaded bands indicate the standard error of the standard deviation (Ahn & Fessler, 2003).

where $\mathbb{1}_C = 1$ if condition C is true, 0 otherwise; and Z_{t+1} is a normalization constant. This equation includes two components: a 'change-point' one ($\tau = 0$) and a 'no change-point' one ($\tau > 0$). We call the model that performs this Bayesian update of the posterior the ***OptimalInference*** model.

Finally, following the inference step just presented (i.e., the computation of the posterior), a 'response-selection step' determines the behavioral response. At trial t and for a response $\hat{s}_t$, the expected reward is $\mathbb{E}_s R = \int R(|\hat{s}_t - s|) p_t(s|x_{1:t}) \mathrm{d}s$. The optimal strategy selects the response, $\hat{s}_t$, that maximizes this quantity. Before exploring the impact of relaxing the optimality in the inference step, in the response-selection step, or both, we examine, first, the behavior of the optimal model.

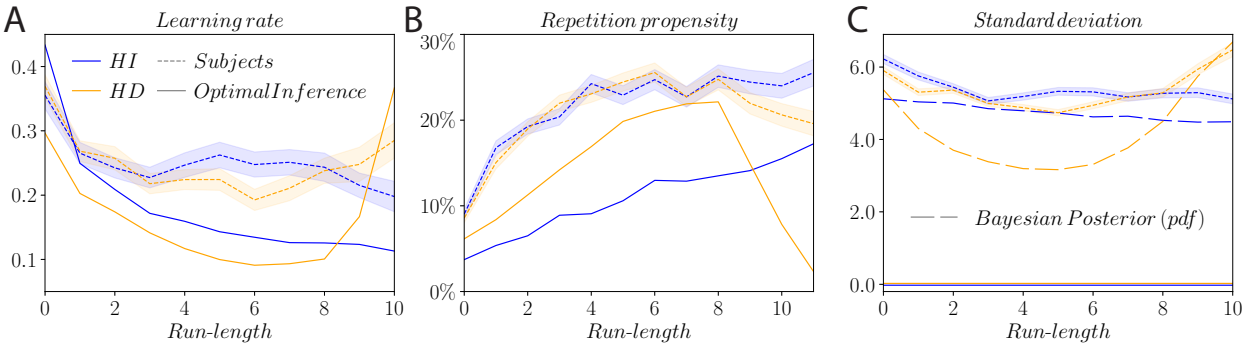

Figure 5. **The optimal model captures qualitatively the behavior of the learning rate and of the repetition propensity in subjects, but does not account for their variability.** **A.** Average learning rate as a function of the run-length. In the HI condition, the learning rate decreases with the run-length, for both the optimal model and the subjects. In the HD condition, learning rates in the optimal model are lower than in the HI condition, for short run-lengths, and higher for long run-lengths. The learning rate of subjects exhibits a similar smile shape, in the HD condition. **B.** Repetition propensity, i.e., proportion of repetition trials, as a function of the run-length. **C.** Standard deviation of the responses of the subjects (dashed lines) and of the optimal model (solid lines), and standard deviation of the optimal, Bayesian posterior distribution (long dashes), as a function of the run-length. The optimal model is deterministic and, thus, exhibits no variability in its responses. The optimal posterior distribution, however, has a positive standard deviation which decreases with the run-length, in the HI condition, and exhibits a smile shape, in the HD condition.

The optimal model captures qualitative trends in learning rate and repetition propensity

Equipped with the optimal model for our inference task, we compare its output to experimental data. For short run-lengths ($\tau < 8$), the learning rates in both HI and HD conditions decrease as a function of the run-length, and the HD learning rates are lower than their HI counterparts. They increase, however, at longer run-lengths ($\tau \geq 8$) and ultimately exceed the HI learning rates; these, by contrast, decrease monotonically (Fig. 5A, solid line). These trends are similar to those observed in behavioral data (Fig. 5A, dashed line). Hence, the modulation of the subjects' learning rates with the temporal statistics of the stimuli, and over the course of inference, is consistent, at least qualitatively, with that of a Bayesian observer.

Although a Bayesian observer can, in principle, hold a continuous posterior distribution, we discretize, instead, the posterior, in order to reproduce the experimental condition of a pixelated screen. This discretization allows for repetitions. The repetition propensity of the optimal model varies with the run-length: it increases with τ in both HI and HD conditions, and decreases in the HD condition for long run-lengths, a pattern also found in experimental data (Fig. 5B).

Hence, the optimal model captures the qualitative trends in learning rate and repetition propensity present in the responses of the subjects. Quantitative differences, however, remain. The learning rates of the subjects, averaged over both HI and HD conditions, are 43% higher than the average learning rate in the optimal model, and the average repetition propensity of the subjects is 9 percentage points higher than that in the optimal model.

Relation between human response variability and the Bayesian posterior

The optimal model captures qualitatively the modulations of learning rate and repetition propensity in subjects, but it is deterministic (at each trial, the optimal estimate is a deterministic function of past stimuli) and, as such, it does not capture the variability

inherent to the behavior of subjects. The modulations of the behavioral variability as a function of the run-length and of the temporal structure of the signal (Fig. 4C) is a sign that the variability evolves as the inference process unfolds. The standard deviation of the optimal Bayesian posterior decreases with the run-length, in the HI condition: following a change point, the posterior becomes narrower as new stimuli are observed. In the HD condition, the standard deviation of the posterior exhibits a 'smile shape' as a function of the run-length: it decreases until the run-length reaches 5, then increases for larger run-lengths (Fig. 5C). This behavior is similar to that of the standard deviation of the responses of the subjects. In fact, the standard deviation of the Bayesian posterior and that of subjects' responses across trials are significantly correlated, both in the HI condition (Pearson's $r = .53$, $p < .0001$) and in the HD condition ($r = .25$, $p < .0001$). In other words, when the Bayesian posterior is wide there is more variability in the responses of subjects, and vice-versa (Fig. 6A).

Turning to higher moments of the distribution of subjects' responses, we find that the skewness of this distribution appears, also, to grow in proportion to the skewness of the Bayesian posterior (Fig. 6B). The correlation between these two quantities is positive and significant in the two conditions (HI: $r = .21$, $p < .0001$; HD: $r = .14$, $p < .0001$. These results are not driven by the boundedness of the response domain, which could have artificially skewed the distribution of response; see Supplementary Fig. B3.). Thus, not only the width, but also the asymmetry in the distribution of subjects' responses is correlated with that of the Bayesian posterior. These observations support the hypothesis that the behavioral variability in the data is at least in part related to the underlying inference and decision processes.

In what follows, we introduce an array of suboptimal models, with the aim of resolving the qualitative and quantitative discrepancies between the behavior of the optimal model and that of the subjects. In particular, we formulate several stochastic models which include possible sources of behavioral variability. Two scenarios are consistent with the

modulations of the magnitude and asymmetry of the variability with the width and skewness of the Bayesian posterior: stochasticity in the inference step (i.e., in the computation of the posterior) and stochasticity in the response-selection step (i.e., in the computation of an estimate from the posterior). The models we examine below cover both these scenarios.

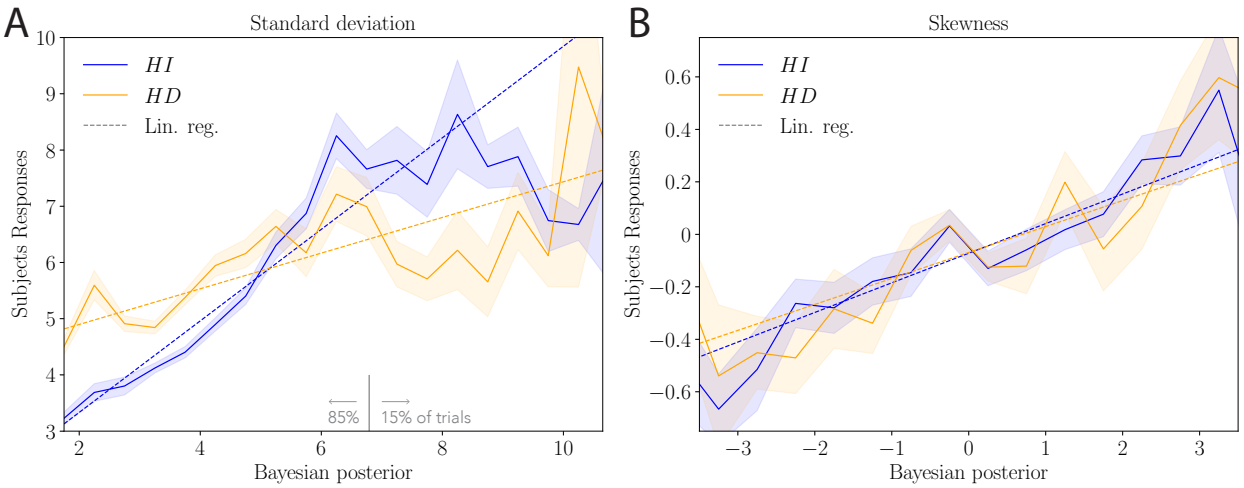

Figure 6. **Both width and skewness of the distribution of subjects' responses are correlated with those of the Bayesian posterior.** Empirical standard deviation **(A)** and skewness **(B)** of subjects' responses as a function of the standard deviation and skewness of the Bayesian posterior, in the HI (blue) and HD (orange) conditions, and linear regressions (ordinary least squares; dashed lines). On 85% of trials, the standard deviation of the Bayesian posterior is lower than 6.8 (vertical grey line). Shaded bands indicate the standard error of the mean.

Suboptimal models reflecting cognitive limitations

In the previous sections, we have examined the learning rate of the subjects, their repetition propensity, and the variability in their responses; comparison of the behaviors of these quantities to that of the Bayesian, optimal model, revealed similarities (namely, the qualitative behaviors of the learning rate and of the repetition propensity) and discrepancies (namely, quantitative differences in these two quantities, and lack of variability in the optimal model). Although the latter call for a non-Bayesian account of human behavior, the former suggest not to abandon the Bayesian approach altogether (in

favor, for instance, of *ad hoc* heuristics). Thus, we choose to examine a family of sub-optimal models obtained from a sequence of deviations away from the Bayesian model, each of which captures potential cognitive limitations hampering the optimal performance.

In the Bayesian model, three ingredients enter the generation of a response upon receiving a stimulus: first, the belief on the structure of the task and on its parameters; second, the inference algorithm which produces a posterior on the basis of stimuli; third, the selection strategy which maps the posterior into a given response. The results presented above, exhibiting the similarity between the standard deviation of the Bayesian posterior and the standard deviation of the responses of the subjects (Fig. 5C), points to a potential departure from the optimal selection strategy, in which the posterior is sampled rather than maximized. This sampling model, which we implement (see below), captures qualitatively the modulated variability of responses; sizable discrepancies in the three quantities we examine, however, remain (see Methods). Hence, we turn to the other ingredients of the estimation process, and we undertake a systematic analysis of the effects on behavior of an array of deviations away from optimality.

Below, we provide a conceptual presentation of the resulting models; we fit them to experimental data, and comment on what the best-fitting models suggest about human inference and estimation processes. For the detailed mathematical descriptions of the models, and an analysis of the ways in which their predictions depart from the optimal behavior, we refer the reader to the Methods section.

Models with erroneous beliefs on the statistics of the signal. Our first model challenges the assumption, made in the optimal model, of a perfectly faithful representation of the set of parameters governing the statistics of the signal. Although subjects were exposed in training phases to blocs of stimuli in which the state, s_t, was made visible, they may have learned the parameters of the generative model incorrectly. We explore this possibility, and, here, we focus on the change probability, $q(\tau)$, which governs the dynamics of the state. (We found that altering the value of this parameter had

a stronger impact on behavior than altering the values of any of the other parameters.) In the HD condition, $q(\tau)$ is a sigmoid function shaped by two parameters: its slope, $\lambda = 1$, which characterizes 'how suddenly' change points become likely, as a function of τ; and the average duration of inter-change-points intervals, $T = 10$. In the HI condition, $q = 0.1$ is constant; it can also be interpreted as an extreme case of a sigmoid in which $\lambda = 0$ and $T = 1/q = 10$. We implement a suboptimal model, referred to as ***IncorrectQ***, in which these two quantities, λ and T, are treated as free parameters, thus allowing for a broad array of different beliefs in the temporal structure of the signal (Fig. 7A).

Models with limited memory. Aside from operating with an inexact representation of the generative model, human subjects may use a suboptimal form of inference. In the HD condition, the optimal model maintains 'in memory' a probability distribution over the entire (s, τ)-space (see Eq. (2)), thus keeping track of a rapidly increasing number of possible histories, each characterized by a sequence of run-lengths. Such a process entails a large computational and memory load. We explore suboptimal models that alleviate this load by truncating the number of possible scenarios stored in memory; this is achieved through various schemes of approximations of the posterior distribution. More specifically, in the following three suboptimal models, the true (marginal) probability of the run-lengths, $p_t(\tau|x_{1:t})$, is replaced by an approximate probability distribution.

A first, simple way of approximating the marginal distribution of the run-lengths is to consider only its mean, i.e., to replace it by a Kronecker delta which takes the value 1 at an estimate of the expected value of the run-lengths. Nassar et al. (2010) introduce a suboptimal model based on this idea, some details of which depend on the specifics of the task; we implement a generalization of this model, adapted to the parameters of our task. We call it the τ***Mean*** model. While the optimal marginal distribution of the run-lengths, $p_t(\tau|x_{1:t})$, spans the whole range of possible values of the run-length, it is approximated, in the τ*Mean* model, by a delta function parameterized by a single value, which we call the 'approximate expected run-length' and which we denote by $\bar{\tau}_t$. Upon the observation of a

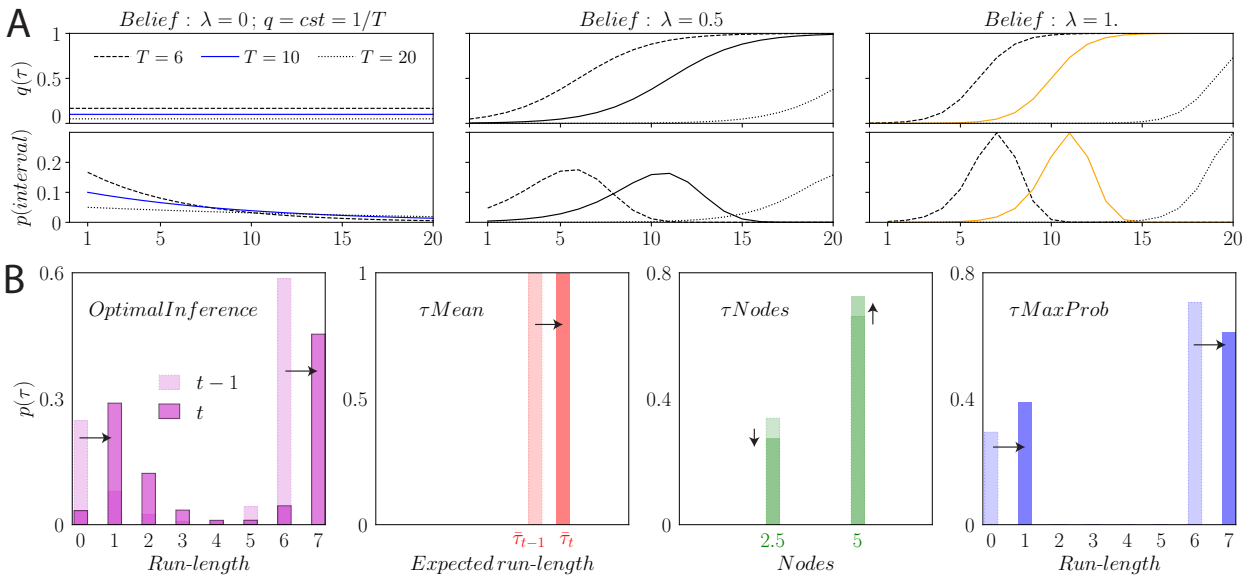

Figure 7. **Illustration of the erroneous beliefs in the *IncorrectQ* model and of the approximations made in the *τMean*, *τNodes*, and *τMaxProb* models. A.** Change probability, $q(\tau)$, as a function of the run-length (first row), and distribution of intervals between two consecutive change points (second row), for various beliefs on the parameters of the change probability: the slope, λ, and the average duration of intervals, T. For a vanishing slope ($\lambda = 0$), the change probability is constant and equal to $1/T$ (first panel). With $T = 10$ this corresponds to the HI condition (blue lines). For a positive slope ($\lambda > 0$), the change probability increases with the run-length (i.e., a change-point becomes more probable as the time since the last change-point increases), and the distribution of intervals between two successive change-points is peaked. The HD condition (orange lines) corresponds to $\lambda = 1$ and $T = 10$. **B.** Schematic illustration of the marginal distribution of the run-length, $p(\tau)$, in each model considered. The *OptimalInference* model assigns a probability to each possible value of the run-length, τ, and optimally updates this distribution upon receiving stimuli (first panel). The *τMean* model uses a single run-length which tracks the inferred expected value, $\bar{\tau}_t$ (second panel). The *τNodes* model holds in memory a limited number, N_τ, of fixed hypotheses on τ ("nodes"), and updates a probability distribution over these nodes; $N_\tau = 2$ in this example (third panel). The *τMaxProb* model reduces the marginal distribution by discarding less likely run-lengths; in this example, 2 run-lengths are stored in memory at any given time (fourth panel).

new stimulus, x_{t+1}, the updated approximate expected run-length, $\bar{\tau}_{t+1}$, is computed as a weighted average between two values of the run-length, 0 and $\bar{\tau}_t + 1$, which correspond to the two possible scenarios: with and without a change point at trial $t + 1$. Each scenario is weighted according to the probability of a change point at trial $t + 1$, given the stimulus, x_{t+1}. This model has no free parameter (Fig. 7B, second panel).

In a second limited-memory model, contrary to the $\tau Mean$ model just presented, the support of the distribution of the run-lengths is not confined to a single value. This model generalizes the one introduced by Wilson et al. (2013). In this model, the marginal distribution of the run-lengths, $p_t(\tau|x_{1:t})$, is approximated by another discrete distribution defined over a limited set of constant values, called 'nodes' (Fig. 7B, third panel). We call this model $\tau\boldsymbol{Nodes}$. A difference with the previous model ($\tau Mean$) is that the support of the distribution is fixed, i.e., the set of nodes remains constant as time unfolds, whereas in the $\tau Mean$ model the single point of support, $\bar{\tau}_t$, depends on the stimuli received. The details of the implementation of this algorithm, and, in particular, of how the approximate marginal distribution of the run-lengths is updated upon receiving a new stimulus, are provided in Methods. The model is parameterized by the number of nodes, N_τ, and the values of the nodes. We implement it with up to five nodes.

The two models just presented are drawn from the literature. We propose a third suboptimal model that relieves the memory load in the inference process. We also approximate, in this model, the marginal distribution of the run-lengths, $p_t(\tau|x_{1:t})$, by another, discrete distribution. We call N_τ the size of the support of our approximate distribution, i.e., the number of values of the run-length at which the approximate distribution does not vanish. A simple way to approximate $p_t(\tau|x_{1:t})$ is to identify the N_τ most likely run-lengths, and set the probabilities of the other run-lengths to zero. More precisely, if, at trial t, the run-length takes a given value, τ_t, then, upon the observation of a new stimulus, at trial $t+1$ it can only take one of two values: 0 (if there is a change point) or $\tau_t + 1$ (if there is no change point). Hence, if the approximate marginal distribution of the run-lengths at trial t is non-vanishing for N_τ values, then the updated distribution is non-vanishing for $N_\tau + 1$ values. We approximate this latter distribution by identifying the most unlikely run-length, $\arg\min p_{t+1}(\tau|x_{1:t+1})$, setting its probability to zero, and renormalizing the distribution. In other words, at each step, the N_τ most likely run-lengths are retained while the least likely run-length is eliminated. We call this

algorithm τ***MaxProb*** (Fig. 7B, fourth panel). It is parameterized by the size of the support of the marginal distribution, N_τ, which can be understood as the number of 'memory slots' in the model.

A model with limited run-length memory through sampling-based inference. The five models considered hitherto (*OptimalInference*, *IncorrectQ*, τ*Mean*, τ*Nodes*, and τ*MaxProb*) are deterministic: a given sequence of stimuli implies a given sequence of responses, in marked contrast with the variability exhibited in the responses of subjects. To account for this experimental observation, we suggest several models in which stochasticity is introduced in the generation of a response. Response stochasticity can stem from the inference step, the response-selection step, or both. We present, first, a model with stochasticity in the inference step.

This model, which we call τ***Sample***, is a stochastic version of the τ*MaxProb* model: instead of retaining deterministically the N_τ most likely run-lengths at each trial, the τ*Sample* model samples N_τ run-lengths using the marginal distribution of the run-lengths, $p_t(\tau|x_{1:t})$. More precisely, if at trial $t+1$ the marginal distribution of the run-lengths, $p_{t+1}(\tau|x_{1:t+1})$, is non-vanishing for $N_\tau + 1$ values, then a run-length is sampled from the distribution $\left[1 - p_{t+1}(\tau|x_{1:t+1})\right]/z_{t+1}$, where z_{t+1} is a normalization factor, and the probability of this run-length is set to zero (Fig. 8). In other words, while the τ*MaxProb* model eliminates the least likely run-length deterministically, the τ*Sample* model eliminates one run-length stochastically, in such a fashion that less probable run-lengths are more likely to be eliminated. The τ*Sample* model has one parameter, N_τ, the size of the support of the marginal distribution of the run-lengths.

Stochastic inference model with sampling in time and in state space: the particle filter. Although the τ*Mean*, τ*Nodes*, τ*MaxProb*, and τ*Sample* models introduced above relieve the memory load by prescribing a form of truncation on the set of run-lengths, inference in these models is still executed on a continuous state space labeled

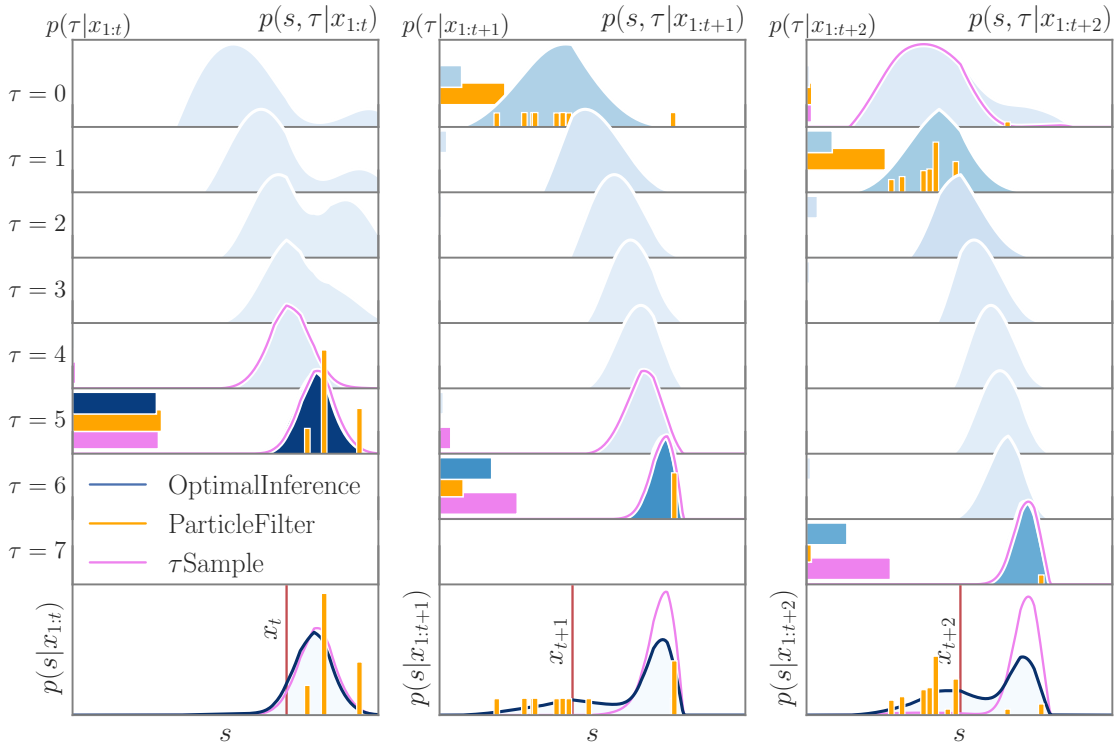

Figure 8. **Posterior density over three successive trials for the *OptimalInference* model, the τ*Sample* model with $N_\tau = 2$, and the *ParticleFilter* model with ten particles.** The three panels correspond to the three successive trials. Each row except the last one corresponds to a different run-length, τ. In these rows, the *horizontal* bars show the marginal probability of the run-length, $p(\tau|x_{1:t})$. The posterior (i.e., the joint distribution of the run-length and the state, $p(s,\tau|x_{1:t})$) is shown as a function of the state, s, for the *OptimalInference* model (blue shaded curve), the τ*Sample* model (pink line), and the *ParticleFilter* model (orange vertical bars). The marginal probability of the run-length, $p(\tau|x_{1:t})$, for the *OptimalInference* model, is additionally reflected in the hue of the curve (darker means higher probability). For the *ParticleFilter* model, the heights of the bars are proportional to the weights of the particles. When the state, s, of two or more particles coincide, a single bar is shown with a height proportional to the sum of the weights. The last row shows the marginal distributions of the states, $p(s|x_{1:t}) = \sum_\tau p(s,\tau|x_{1:t})$, along with the location of the stimulus at each trial (red vertical line). At trial t (left panel), the probability of the run-length $\tau = 5$ dominates in the three models. In the τ*Sample* model, it vanishes at run-lengths from 0 to 3, and it is very small for $\tau = 4$. In the *ParticleFilter* model, the run-lengths of the ten particles are all 5, and thus the probability of all other run-lengths is zero. At trial $t + 1$ (middle panel), upon observation of the new stimulus, x_{t+1}, the marginal probability of the vanishing run-length ($\tau = 0$), which corresponds to a 'change-point' scenario, becomes appreciable in the *OptimalInference* model (top row). The probability of the run-length $\tau = 6$ (a 'no change-point scenario') is however higher. As a result, a 'bump' appears in the marginal distribution of the state, around the new stimulus (bottom row). *(Continued on the following page.)*

Figure 8. In the *τSample* model, the optimal update of the posterior results in a non-vanishing probability for three run-lengths ($\tau = 0$, 5, and 6), more than the number of 'memory slots' available ($N_\tau = 2$). One run-length is thus randomly chosen, and its marginal probability is set to zero; in the particular instantiation of the model presented here, the run-length $\tau = 0$ is chosen, and thus the resulting marginal probability of run-length is non-vanishing for $\tau = 5$ and 6 only. In the *ParticleFilter* model, the stochastic update of the particles results in seven particles adopting a vanishing run-length, and the probability of a 'change-point' scenario ($\tau = 0$) becomes higher than that of the 'no change-point' scenario ($\tau = 6$) supported by the remaining three particles. The various marginal distributions of the states obtained in these three models (bottom row) illustrate how the *τSample* model and the *ParticleFilter* model approximate the optimal posterior: the *τSample* model assigns a negligible probability to a set of states whose probability is substantial under the *OptimalInference* model, while the *ParticleFilter* yields a coarse approximation reduced to a support of ten states (as opposed to a continuous distribution).

by s (or, more precisely, on a discrete space with resolution as fine as a pixel). Much as subjects may retain only a compressed representation of probabilities along the τ axis, it is conceivable that they may not maintain a full probability function over the 1089-pixel-wide s axis, as they carry out the behavioral task. Instead, they may infer using a coarser spatial representation, in order to reduce their memory and computational loads. Monte Carlo algorithms perform such approximations by way of randomly sampling the spatial distribution; sequential Monte Carlo methods, or 'particle filters', were developed in the 1990s to address Hidden Markov Models, a class of hidden-state problems within which falls our inference task (Arulampalam, Maskell, Gordon, & Clapp, 2002; Doucet & Johansen, 2008; Gordon, Salmond, & Smith, 1993). Particle filters approximate a distribution by a weighted sum of delta functions. In our case, a *particle i at trial t* is a triplet, (s_t^i, τ_t^i, w_t^i), composed of a state, a run-length, and a weight; a particle filter with N_P particles approximates the posterior, $p_t(s, \tau | x_{1:t})$, by the distribution

$$\tilde{p}_t(s, \tau | x_{1:t}) = \sum_{i=1}^{N_P} w_t^i \delta(s - s_t^i) \delta_{\tau, \tau_t^i}, \tag{3}$$

where $\delta(s - s_t^i)$ is a Dirac delta function, and δ_{τ,τ_t^i} a Kronecker delta. In other words, a distribution over the (s, τ) space is replaced by a (possibly small) number of points, or

samples, in that space, along with their probability weights.

To obtain the approximate posterior at trial $t+1$ upon the observation of a new stimulus, x_{t+1}, we note, first, that the Bayesian update (Eq. (2)) of the approximate posterior, $\tilde{p}_t(s, \tau | x_{1:t})$, is a mixture (a weighted sum) of the N_P Bayesian updates of each single particle (i.e., Eq. (2) with the prior, $p_t(s, \tau | x_{1:t})$, replaced, for each particle, by $\delta(s - s_t^i)\delta_{\tau,\tau_t^i}$). Then, we sample independently each component of the mixture (i.e., each Bayesian update of a particle), to obtain stochastically the updated particles, $(s_{t+1}^i, \tau_{t+1}^i)$, and to each particle is assigned the weight of the corresponding component in the mixture. The details of the procedure just sketched, in particular the derivation of the mixture and of its weights, and how we handle the difficulties arising in practical applications of the particle filter algorithm, can be found in Methods. This model, which we call ***ParticleFilter***, has a single free parameter: the number of particles, N_P (Fig. 8).

Models with variability originating in the response-selection step. The *τSample* and *ParticleFilter* models presented above reduce the dimensionality of the inference problem by pruning stochastically the posterior, in the inference step. But, as we pointed out, the behavior of the standard deviation of the responses of the subjects, as compared to that of the width of the Bayesian posterior (Fig. 5C), hints at a more straightforward mechanism at the origin of response variability. The model we now introduce features stochasticity not in the inference step, but rather in the response-selection step of an otherwise optimal model. In this model, the response is sampled from the marginal posterior on the states, $p_t(s|x_{1:t})$, i.e., the response, $\hat{s}_t$, is a random variable whose density is the posterior. This contrasts with the optimal response-selection strategy, which maximizes the expected score based on the Bayesian posterior, and which was implemented in all the models presented above. Henceforth, we denote the optimal, deterministic response-selection strategy by ***Max***, and the suboptimal, stochastic strategy just introduced by ***Sampling***. It has no free parameter.

Another source of variability in the response-selection step might originate in a limited

motor precision, in the execution of the task. To model this motor variability, in some implementations of our models we include a fixed, additive, Gaussian noise, parameterized by its standard deviation, σ_m, to obtain the final estimate. Both this motor noise and the *Sampling* strategy entail stochasticity in response selection. The former, however, has a fixed variance, σ_m^2, while the variance of the latter depends on the posterior which varies over the course of inference (Fig. 5C). When we include motor noise in the *Max* or in the *Sampling* strategies, we refer to these as ***NoisyMax*** and ***NoisySampling***, respectively.

In sum, we have described four response-selection strategies (*Max*, *Sampling*, *NoisyMax*, and *NoisySampling*), and seven inference strategies, of which five are deterministic (*OptimalInference*, *IncorrectQ*, τ*Mean*, τ*Nodes*, and τ*MaxProb*) and two are stochastic (τ*Sample* and *ParticleFilter*). We can combine any inference strategy with any response-selection strategy: thus, we have at hand $4\times7 = 28$ different models, 27 of which are suboptimal, obtained from pairings of the inference and selection strategies. We label each of the 28 models by the combination of the two names referring to the two steps in the process, e.g., *ParticleFilter+Sampling*.

Fitting models to experimental data favors sample-based inference

The 27 suboptimal models introduced in the previous section yield a range of discrepancies from the optimal behavior. The ways in which each deviation from the optimal model impacts behavior is examined in Methods; here, we ask how well these models account for the behavior of human subjects. Whereas the optimal model, *OptimalInference+Max*, computes the Bayesian posterior (*OptimalInference*) and selects the maximizing response (*Max*), the suboptimal models mimic cognitive limitations that may prevent the brain from reaching optimality: incorrect belief in the temporal structure of the signal (*IncorrectQ*), compressed representation of the Bayesian posterior, either deterministically (τ*Mean*, τ*Nodes*, and τ*MaxProb*) or stochastically (τ*Sample* and *ParticleFilter*), and noise introduced in the response-selection step, with a width either

Table 1. **Model fitting favors the *ParticleFilter* inference strategy with NoisyMax response selection.** Ratios of the normalized mean squared error (NMSE) in each model to that of the best-fitting model, *ParticleFilter+NoisyMax*. Each model is a combination of an inference strategy (columns) with a response-selection strategy (rows). The second best model, also a *ParticleFilter* but with a *NoisySampling* response-selection strategy, yields an NMSE 46% higher than the best-fitting model.

scaling with that of the posterior (*Sampling*), or constant (*NoisyMax*), or a combination of the two (*NoisySampling*).

To evaluate the ability of each of these models to account for human behavior, we compare quantitatively their respective outputs with the responses of human subjects. For the three quantities we examine (the learning rate, the repetition propensity, and the standard deviation of the responses), we compute the normalized mean squared error (NMSE) (sometimes referred to as the 'Fraction of Variance Unexplained' in the context of linear regressions). It is defined, for a given model and for each quantity, as the ratio of the mean squared error in the model output as compared to data, and the variance of the quantity under scrutiny in the behavioral data. We fit each of our models to human data, using the average of the three NMSEs as our error measure. (We note that the

OptimalInference inference strategy is a special case of all the other inference strategies, except τ*Mean*, thus its NMSE cannot be lower than that of these strategies. Likewise, the *Max* and *Sampling* response-selection strategies are special cases of the *NoisyMax* and *NoisySampling* strategies, respectively.)

We find that the five best-fitting models make use of stochastic compression in the inference step, in either the τ*Sample* approximation or the *ParticleFilter* approximation (Table 1). These models all reproduce the qualitative trends in the behavior of subjects with respect to our three measures: for the learning rate and the standard deviation, the 'smile shape' of the HD curve, which crosses a decreasing HI curve; for the repetition propensity, conversely, an inverted U shape of the HD curve which crosses an increasing HI curve (Fig. 9, results from the τ*Sample+Max* and *ParticleFilter+Sampling* models are not shown, but the corresponding curves are similar).

The τ*Sample* and *ParticleFilter* strategies have one or two parameters, depending on whether they include motor noise or not. Other models, including all models with a deterministic inference step, have an error at least 30% higher than the best five models (and 2.45 times higher than the best model), despite the fact that other strategies come with up to five parameters (Table 1). The best-fitting model is *ParticleFilter+NoisyMax* with $N_P = 9$ particles. The fitted standard deviation, σ_m, of the Gaussian motor noise is approximately equal to 0.77 pixels; as a consequence, in about half of the trials, the noise component is within the width of a pixel, and thus has no impact. The second best model also follows a *ParticleFilter* inference strategy, with $N_P = 14$, combined with a *NoisySampling* response selection (with $\sigma_m = 0.70$ pixels).

The third and fourth best-fitting models use the τ*Sample* inference strategy, with $N_\tau = 1$, and the *NoisyMax* (with $\sigma_m = 0.45$ pixels, for the third one) and the *Max* (for the fourth one) selection strategies. At any given trial, these two models retain only a single assumption, τ_t, on the run-length. Upon receiving a new stimulus, x_{t+1}, a model subject computes $p_{change} = p_{t+1}(\tau = 0|x_{1:t+1})$ and $1 - p_{change} = p_{t+1}(\tau = \tau_t + 1|x_{1:t+1})$, and decides

whether there was a change-point by sampling this simple, Bernoulli distribution. This sampling process, over a marginalization of the posterior, is similar to that in the particle filter model, which samples over the full (s, τ)-dependent posterior. As a consequence of sampling, the *τSample* strategy also exhibits variability, which behaves in a fashion similar to the variability in the *ParticleFilter* strategy (Fig. 9, bottom right). As for response selection, we note that with the *Sampling* and *NoisySampling* selection strategies (instead of the *Max* and *NoisyMax* strategies), these models do not perform as well, and result in errors larger by 86% (*Sampling* vs. *Max*) and 96% (*NoisySampling* vs. *NoisyMax*). In fact, for all the seven inference models, the *NoisyMax* response-selection strategy results in errors lower or equal (but more often, lower) than the other three selection strategies (*Max*, *Sampling* and *NoisySampling*) (Table 1). This suggests that the variability in human responses does not originate from a posterior-sampling strategy in the response-selection step, but, rather, from an intrinsically stochastic inference process. In order to seek further validation of this finding, we explore, below, a generalization of the *Sampling* strategy.

Robustness of the results. To substantiate the picture that emerges from the results summarized above, we perform two supplementary analyses. First, we investigate whether a generalized *Sampling* strategy yields smaller errors than the *NoisyMax* strategy. Second, we consider our choice of fitting-performance measure (the average of the NMSEs in the three quantities we examine), and we check for the robustness of model fitting to changes in the relative weights of each quantity in the fitting performance measure.

Sampling from the posterior function is only one of many possible sampling strategies for response selection. Furthermore, in practice sampling may be difficult to tease apart from maximizing a perturbed posterior function. Acerbi, Vijayakumar, and Wolpert (2014) argue that, for some forms of random perturbations of a posterior probability density, maximizing the randomly perturbed function yields similar results to sampling from a modified posterior density function obtained as a power of the correct posterior: $p_\kappa(s) \propto p(s|x_{1:t})^\kappa$. To establish the equivalence, the exponent, κ, is chosen as inversely

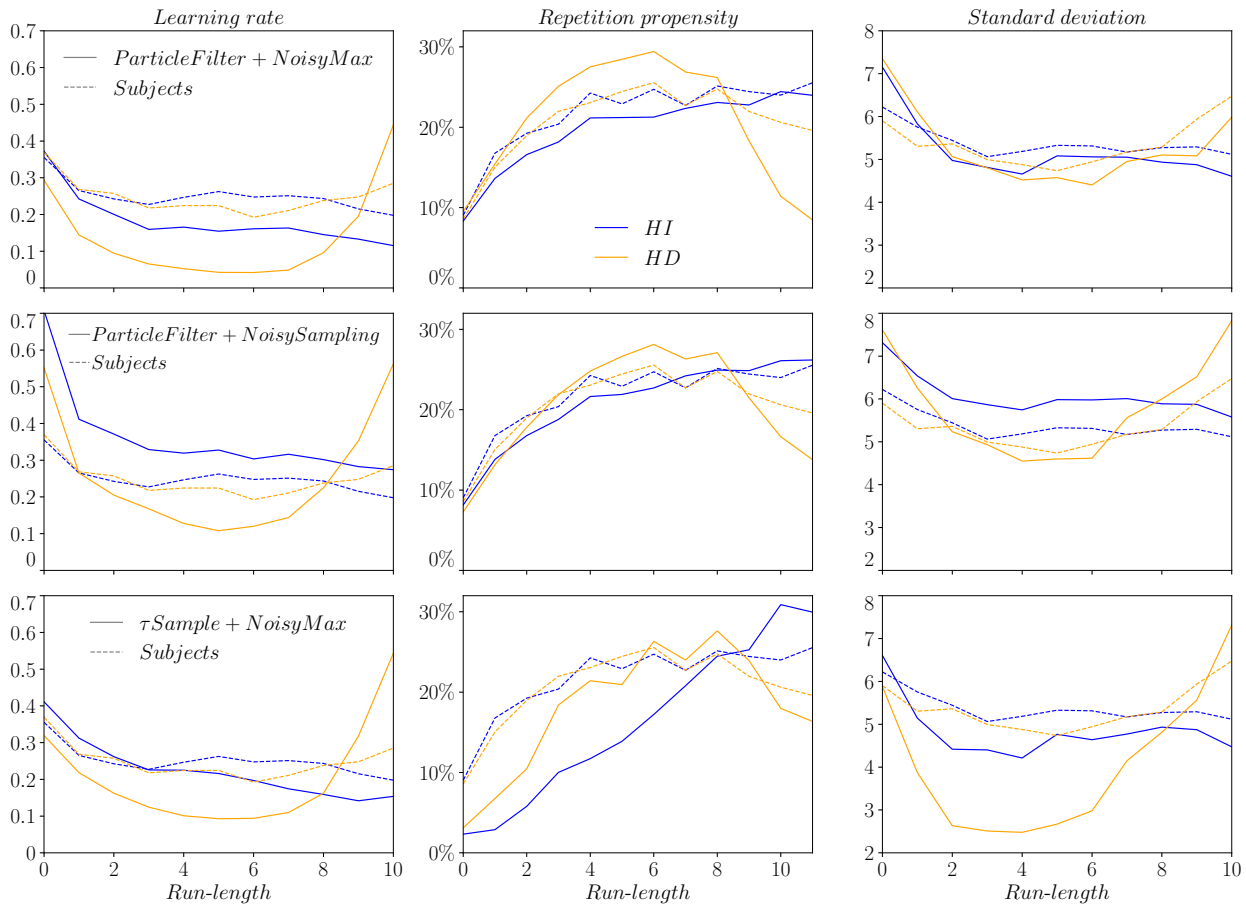

Figure 9. **Behavior of the three best-fitting models.** In HI (blue curves) and HD (yellow curves) conditions, average learning rate (first column), repetition propensity (second column), and standard deviation of responses (third column), as a function of run-length, for the subjects (solid lines) and the three best-fitting models (dashed lines): *ParticleFilter+NoisyMax* (first row), *ParticleFilter+NoisySampling* (second row), and *τSample+NoisyMax* (third row).

related to the magnitude of the perturbing noise. Sampling from the modified posterior yields a behavior that interpolates between posterior sampling (for $\kappa = 1$) and maximizing (for $\kappa \to \infty$); it yields a family of softmax operations over the posterior (Vul, 2011; Yu & Huang, 2014). Another interpretation of this sampling strategy is proposed by Battaglia et al. (2011): in the case of an integer κ and a Gaussian posterior, the mean of κ samples drawn from the posterior is a Gaussian random variable, with a standard deviation equal to that of the posterior scaled by $1/\sqrt{\kappa}$; i.e., a distribution equal to the posterior raised to the power κ, and normalized. Hence, in the Gaussian case, sampling from the

exponentiated posterior can be interpreted as drawing κ samples from the unexponentiated posterior, and taking the mean.

We implement this strategy of response selection by sampling a modified posterior, which we denote κ*Sampling*. We find that it performs better than the *Sampling* strategy, as expected since the *Sampling* strategy is a special case of the κ*Sampling* (with the parameter, κ, set to unity). However, in the case of all seven inference models, the κ*Sampling* strategy, which has one free parameter, performs worse than the *NoisyMax* strategy, which has, also, a single parameter (Fig. 10B). Hence, a random, additive perturbation of the maximization strategy remains a better account of human behavior than a posterior-sampling strategy.

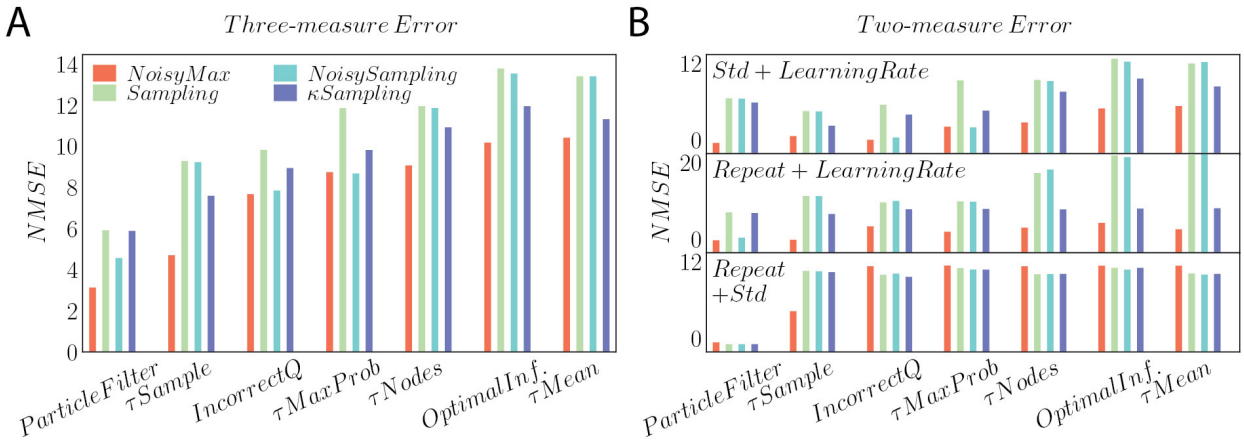

Figure 10. **Model fitting is robust to the measure used for model comparison.** Normalized Mean Squared Error of fitted models. **A.** NMSE of models fitted to subjects data, averaged over the three measures (learning rate, repetition propensity, and standard deviation of responses), grouped by inference models. **B.** NMSE between fitted models and subjects data, averaged over two out of the three measures.

Our results, which suggest that the variability in the responses of subjects originate in the inference step rather than in the response-selection step, rely upon the fitting performance measure used for model comparison. We chose a measure that weighted equally the three NMSEs (on the learning rate, repetition propensity, and standard deviation), so as to obtain a model performing well on all fronts, but that choice was

arbitrary. Hence, one may be concerned, for instance, that the goodness-of-fit of the *ParticleFilter* model be due to our choice of weighing errors. As a control, we computed the three 'two-measure errors', each excluding one of the three measures and averaging the errors in the two remaining ones. We found that, regardless of the choice of the combination, the relative order of the models in terms of performance stays identical, with only a few exceptions. Most importantly, the *ParticleFilter* remains, in all three cases of error combinations, the best-fitting model (Fig. 10B).

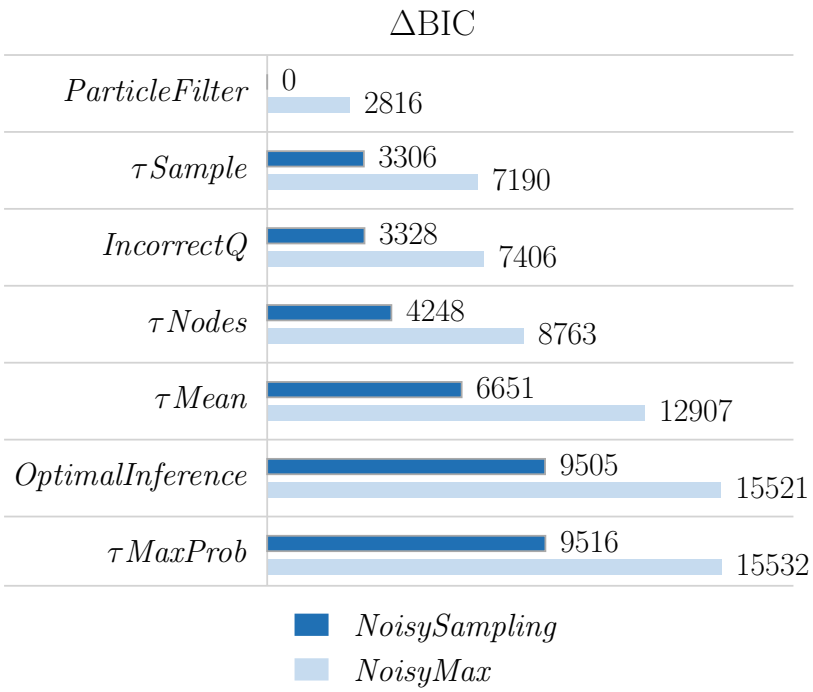

Figure 11. **Bayesian model selection favors the *ParticleFilter* inference strategy.** Difference between the BIC of each model and that of the best-fitting model, for the models combining one of the seven inference strategies with the $NoisyMax$ or the $NoisySampling$ response-selection strategies. The two best-fitting models make use of the *ParticleFilter* inference strategy.

Bayesian model selection also favors sample-based inference

Another concern regarding the choice of the NMSE as our performance measure for model comparison is that it does not take into account the number of parameters in the

models. A standard method to fit and compare models is to maximize the log-likelihood of each model and compute its Bayesian Information Criterion (BIC), which includes a penalty as a function of the number of parameters in the model (Schwarz, 1978). In several of our models (and in many models in the literature), the responses in successive trials, conditioned on the stimuli presented to the subject, are independent; as a result, the log-likelihood over all trials is the sum of the log-likelihoods for each trial, taken separately. This obtains for all the models in which the inference strategy is deterministic (*OptimalInference*, *IncorrectQ*, τ*Mean*, τ*Nodes*, and τ*MaxProb*). It does not apply, however, for the models with stochastic inference strategies (τ*Sample* and *ParticleFilter*): in these models, successive responses, conditional on observed stimuli, are *not* independent as they depend on the realization of the stochastic process that governs inference. To compute the BIC, it is therefore necessary to compute, first, the distribution of the possible realizations of the stochastic inference process. The difficulty, here, lies in the fact that the space of these realizations grows exponentially with the number of trials in an experimental run.

In the context of our task, in which the subjects undergo 1000 trials in a run, an exact computation of the BIC is prohibitive. In order to circumvent this problem, we propose to approximate the log-likelihood of a model by way of a Monte-Carlo estimation of the log-likelihoods of short sequences of responses. This approach limits the computational load of the estimation while taking into account the sequential dependence of responses. We report, here, the results of this estimation scheme using short sequences of 10 successive trials, but in our investigations we repeated the calculations for different choices, which yielded comparable results. We detail the procedure in Methods. Here, we mention that, even though models with temporal correlation such as the particle filter have been used to capture cognitive processes, to the best of our knowledge Bayesian model selection using the BIC has not been applied to them, except in the case of a binary categorization task (Lloyd, Sanborn, Leslie, & Lewandowsky, 2019). The approximate approach we

propose may thus be of use beyond the confines of the specifics of our problem.

In the models that do not feature a (Gaussian) motor noise, some responses of the subjects have a vanishing probability, and thus these models have an infinite BIC. Hence, we look at the BICs of the models equipped with the *Noisy* or the *NoisySampling* response-selection strategy. We find that the three best-fitting models involve a stochastic approximation of the Bayesian inference: the two best-fitting models make use of the *ParticleFilter* inference strategy, and the third best-fitting model has a τ*Sample* inference strategy (Fig. 11). We note that with the NMSE metric the three best-fitting models were also the two *ParticleFilter* models followed by a τ*Sample* model. Thus, both model-fitting approaches suggest that human inference evolves according to a stochastic compression of the posterior. The best-fitting model is *ParticleFilter+NoisySampling*, with $N_P = 8$ particles, and its BIC is smaller than that of the second best-fitting model, the *ParticleFilter+NoisyMax* model with $N_P = 4$ particles, by $2,816$. This result is consistent with the best-fitting numbers of particles obtained when minimizing the NMSE, which were also relatively modest, although slightly larger ($N_P = 14$ with *NoisySampling* and $N_P = 9$ with *NoisyMax*).

Taken together, our results suggest that variability in human behavior, at least in the context of our task, is dictated primarily by stochasticity in the inference step — i.e., in the manipulation and update of probabilities — rather than by 'output noise' such as stochasticity in the response-selection step or motor noise. This view agrees with the conclusion of a recent study of a cue combination task (Drugowitsch, Wyart, Devauchelle, & Koechlin, 2016); its authors argue that a "dominant fraction" of human choice suboptimality arises from random fluctuations in the inference step.

Discussion

Summary

This study investigates the behavior of human subjects in an online inference task, and examines mechanisms that can account for behavioral trends found in experimental data. An important aspect of this task is that it makes use of both a history-independent (HI) condition with no temporal structure, and a history-dependent (HD) condition in which a hidden state is almost periodical and, hence, highly structured in time (Fig. 1D). We find that subjects display different behaviors in the two conditions, adapting their learning rate to the temporal structure of the hidden state. We also note a propensity in subjects to repeat their response in consecutive trials; this repetition propensity increases with the run-length, and in the HD condition drops again for larger run-lengths. Moreover, we observe that subjects exhibit a greater variability in their responses shortly after a change point, in both conditions, and at long run-lengths in the HD condition, i.e., the *variability* in behavior also depends on the temporal statistics of the stimuli.

The distinctive behaviors of the learning rate and the repetition propensity in the HI and HD conditions are reproduced qualitatively by a Bayesian model of inference which yields optimal updates of the probability density of the hidden state. As for the variability in subjects' responses, we find that its behavior is similar to that of the standard deviation of the Bayesian posterior. We therefore use the Bayesian model as a starting point to elaborate variant models which can account for the trends exhibited in human responses. We find that the variability in human behavior, and its dynamics, can be reproduced by supoptimal models in which inference is executed in a stochastic manner. Specifically, the τ*Sample* and the *ParticleFilter* models alter the optimal inference step by maintaining an approximate version of the posterior, by means of random sampling. This alteration of the optimal model at once introduces variability in the behavior and relieves the memory capacity, through sampling either in the 'time dimension' (in the τ*Sample* model) or in the 'time and space dimensions' (in the *ParticleFilter* model).

The behavioral patterns that arise in our task in the HI condition are also found in other experiments. Gallistel et al., 2014 and Khaw et al., 2017 both conducted an online inference task, with change points that occurred with constant probability (similarly to our HI condition). We examined the responses of their subjects in the context of the respective tasks, and we observed very similar behavioral trends: the learning rate decreases as a function of the run-length, while the repetition propensity increases. As for the empirical standard deviation of subjects' responses, we note that in these studies some subjects were presented several times with the same sequence of stimuli, in different sessions, thus allowing for the examination of the variability of responses within subjects. We find, here also, that the within-subject standard deviation of responses shows the same modulations than the standard deviation of the Bayesian posterior (Supplementary Fig. B5).

Temporal structures in nature and their behavioral and neural counterparts

In order to make appropriate decisions in relation to their environment, humans and animals must infer the state of the surrounding world on the basis of the sensory signals they receive. If these signals are noisy and if the environment is changing, their inference task is complicated by the fact that a new stimulus may reflect either noise or a change in the underlying state. However, if events in the world present some kind of temporal structure, such as in our HD signal, it is possible to use this structure to refine one's inference. Conversely, if events follow a Poisson process, as in the HI signal, their occurrences present no particular temporal structure, and what just happened conveys no information on what is likely to happen next. Hence, there is a fundamental difference between the HI and HD conditions, which impacts the inference of an optimal observer.

Many natural events are not Poisson-distributed in time, and exhibit strong regularities. Nunes Amaral et al. (2004), Nakamura et al. (2007), Nakamura et al. (2008), and Anteneodo and Chialvo (2009) have recorded the motor activity of both rodents and human subjects over the course of several days. In both species, they found that the time

intervals between motion events were distributed as a power law, a distribution characterized by a long tail, leading to bursts, or clusters, of events followed by long waiting epochs. The durations of motion episodes also exhibited heavy tails. These kinds of distribution are incompatible with Poisson processes, which yield exponentially distributed inter-event epochs. Moreover, both rodent and human activity exhibited long-range autocorrelations, another feature that cannot be explained by a Poisson process. A particular form of autocorrelation is periodicity, which occurs in a wide range of phenomena. In the context of human motor behavior, walking is a highly rhythmical natural activity (Griffin et al., 2000; Hausdorff et al., 1995). More complex patterns exist (neither clustered nor periodic), such as in human speech which presents a variety of temporal structures, whether at the level of syllables, stresses, or pauses (Campione & Véronis, 2002; Low et al., 2000; Ramus et al., 1999). In all these examples, natural mechanisms produce series of temporally structured events. The ubiquity of history-dependent statistics of events in nature begs for explorations of inference mechanisms in their presence. For the purposes of our experiment, we chose an idealized temporal signal that combined several advantages: it featured a prominent form of history dependence, approximate periodicity; it was not easily distinguishable from the other, history-independent signal used in the task; and it was amenable to modeling.

In the case of studies of perception and decision-making, in both humans and animals, history-dependent signals have been used widely. In a number of experiments (Ghose & Maunsell, 2002; Janssen & Shadlen, 2005; Jazayeri & Shadlen, 2010; Li & Dudman, 2013; Miyazaki et al., 2005), a first event (a sensory cue, or a motor action such as a lever press) is followed by a second event, such as the delivery of a reward, or a 'go' signal triggering the next behavior. The time elapsed between these two events – the 'reward delay' or the 'waiting time' – is randomized and sampled from distributions that, depending on the studies, vary in mean, variance, or shape. For instance, both Janssen and Shadlen (2005) and Ghose and Maunsell (2002) use unimodal and bimodal temporal distributions.

Because of the stochasticity of the waiting time, the probability of occurence of the second event varies with time, similarly to the probability of a change point in our HD condition; these studies explore whether variations of this probability are used by human and animal subjects. Janssen and Shadlen (2005)'s recordings from the V4 cortical area in rhesus monkey indicate that, for both unimodal and bimodal waiting times distributions, the attentional modulation of sensory neurons varies consistently with the event probability. Ghose and Maunsell (2002) note that the reaction times of macaques are inversely related to the event probability, for both unimodal and bimodal distributions, and that the activity of neurons in the lateral intraparietal (LIP) area is correlated with the evolution of this probability over time. Li and Dudman (2013) manipulate another aspect of the distribution of reward delays: between blocks of trials, the standard deviation of this distribution is varied, while the mean is left unchanged. Mice, in this situation, are shown to adapt their waiting times to this variability of reward delays, consistently with a probabilistic inference model of reward timing.

Akin to the tasks just outlined are 'ready-set-go time-reproduction tasks', in which subjects are asked to estimate the random time interval between 'ready' and 'set' cues, and to reproduce it immediately afterwards. Miyazaki et al. (2005) and Jazayeri and Shadlen (2010) show that human subjects combine optimally the cue (consisting in the perceived ready-set interval) with their prior on the interval length. Different priors are learned in training runs: they differ by the variances of the interval distributions (Miyazaki et al., 2005) or by their means (Jazayeri & Shadlen, 2010). In both cases, subjects integrate the prior in a fashion consistent with Bayesian inference. Adopting a different approach, ten Oever et al. (2014) show that attentional resources can be dynamically allocated to points in time at which input is expected: when asked to detect auditory stimuli (beeps) of low intensity embedded in a continuous white noise, human subjects perform better when detecting periodic beeps rather than random beeps, suggesting that they are able to identify the temporal regularity and use it in their detection process.

In all these studies, the event of interest has a probability of occurrence that varies with time. The resulting temporal structure in the signal appears to be captured by human and animal subjects, and reflected in behavior and in its neural correlate. Various probability distributions used in the reported tasks can be compared directly to our HD sigmoid-shaped change probability, with adjusted parameters. In line with these studies, our results confirm that human subjects adapt their behavior depending on the temporal structure of stimuli. Additionally, we provide a comparison between two different conditions, a HD condition akin to a 'jittered periodic' process, and the Poisson, HI condition; the latter produces a memoryless process. Importantly, it plays the role of a benchmark from the point of view of probability theory: in discrete time it yields a geometric distribution, and in continuous time it yields an exponential distribution; both distributions maximize the entropy, subject to the constraint of a fixed event rate. In this study, we compared a specific, temporally structured HD condition to this benchmark, HI condition.

Online Bayesian inference

Our first observation is that the average learning rate of subjects and their repetition propensity are captured by a Bayesian model. The Bayesian paradigm has been viewed as an extension of logic that enables reasoning with propositions whose truth or falsity is uncertain (Cox, 1946; Jaynes, 2003). In cognitive science, it has successfully accounted for a wide range of observations, including cue combination in humans (Battaglia et al., 2003; Battaglia et al., 2011; Ernst & Banks, 2002; Hillis et al., 2004; Jacobs, 1999; Knill, 2003, 2007; van Beers et al., 1999), sensorimotor control (Berniker & Kording, 2011; Körding & Wolpert, 2004, 2006), integration of temporal statistics (Ghose & Maunsell, 2002; Janssen & Shadlen, 2005; Jazayeri & Shadlen, 2010; Li & Dudman, 2013; Miyazaki et al., 2005), perceptual multistability (Gershman et al., 2012; Moreno-Bote et al., 2011; Sundareswara & Schrater, 2008), and various aspects of cognition (Blaisdell, Sawa, Leising, & Waldmann,

2006; Goodman et al., 2008; Griffiths & Tenenbaum, 2006, 2011; Stocker & Simoncelli, 2008).

The literature on Bayesian *online* inference in cognition, where belief is updated iteratively as a function of incoming information, is growing; examples can be found in word segmentation (Pearl, Goldwater, & Steyvers, 2011), sentence processing (Levy, Reali, & Griffiths, 2008), conditioning (Daw & Courville, 2008), as well as in the change-point literature (Gallistel et al., 2014; Glaze et al., 2018; Glaze et al., 2015; Khaw et al., 2017; Nassar et al., 2010; Piet et al., 2018; Wilson et al., 2010, 2013). In change-point tasks, subjects are presented with a long sequence of consecutive inference problems (1000 of them, in our case). Each trial is a slightly different task, in which one has to handle the uncertainty resulting from the belief distribution, from the signal likelihood, and from the possibility of a change point. The latter, in the HD condition, bears the added complexity of a change-point probability, $q(\tau)$, that depends on the time of the last change point. How these uncertainties are handled determines the behavior, and in particular to what extent an observer reacts to a new stimulus: either shift the estimate towards it, or not move at all. We quantify this response through the learning rate and the repetition propensity, and we find that the ideal Bayesian observer and the subjects obey similar trends (Fig. 5A-B).

The success of the Bayesian paradigm, however, is limited, and comes with three shortcomings. First, subjects do not behave quantitatively like the ideal Bayesian observer, and hence there remains unexplained suboptimality. Second, we find variability in the responses of subjects (Fig. 5C), an observation incompatible with optimal Bayesian inference. Third, inference problems in the real world are complex and high-dimensional, rendering Bayesian reasoning computationally heavy and memory-intensive. This suggests that humans use approximations when carrying out inference and estimation tasks (Gershman & Beck, 2016; Gershman, Horvitz, & Tenenbaum, 2015; Sanborn & Chater, 2016). These three observations call for the investigation of alternatives to the optimal Bayesian paradigm. Our study explores several scenarios.

Sampling versus noisy maximization

Although the behavior of the subjects and of the Bayesian model differ in that the former exhibits variability while the latter is deterministic, the temporal modulation of the human variability follows a similar course to that of the standard deviation of the Bayesian posterior (Fig. 5C), and both the standard deviation and the skewness of the distribution of subjects' responses are correlated with those of the Bayesian posterior (Fig. 6). Therefore, it is natural to propose that response selection operates by sampling the Bayesian posterior instead of maximizing the expected score. Decision by posterior sampling, or 'probability matching', has been suggested by other decision-making experiments (Denison, Bonawitz, Gopnik, & Griffiths, 2013; Herrnstein, 1961; Koehler & James, 2009) and, more recently, perceptual experiments (Battaglia et al., 2011; Moreno-Bote et al., 2011; Wozny, Beierholm, & Shams, 2010). Although close to optimal in some specific paradigms (Kaufmann, Korda, & Munos, 2012), sampling is suboptimal in the context of our behavioral task. When fitting models to human data, we observe that the *ParticleFilter+Sampling* and the *ParticleFilter+NoisySampling* models yield larger NMSE than the *ParticleFilter+NoisyMax* model (the best-fitting model), by 90% and 48%, respectively. More generally, we observe that for each of our seven inference strategies, the *NoisyMax* response-selection strategy results in better fits (lower NMSE) than the *Sampling* and the *NoisySampling* strategies (Fig. 10A). Relaxing the *Sampling* model by allowing the posterior to be exponentiated before sampling, as in the κ*Sampling* strategy, does not yield better fits than the *NoisyMax* strategy either. We conclude that posterior sampling accounts less successfully for our data than a simple perturbation of the optimal maximization strategy by an additive, fixed-width, Gaussian noise.

Alternative models of response selection: 'rational inattention' and *ad hoc* repetition probability

Aside from posterior sampling and motor noise, the so-called 'rational inattention' approach which has been gaining grounds in economics (Sims, 2003, 2011; Woodford, 2009), suggests a different account of the variability of responses in decision-making tasks. In complement to the study of Bayesian and approximate Bayesian models, we have examined models inspired by that approach. We summarize, here, our results, and provide a more detailed discussion in Methods. Rational-inattention models posit the existence of a cognitive cost which prevents subjects from making optimal decisions. In a standard formulation of the approach, this cost is assumed to be proportional to the mutual information between a subject's mental representation (of quantities relevant to produce responses) and the external variables relevant to the decision (here, the sequence of presented stimuli). The subject optimizes the 'information structure', i.e., the distribution of the mental representation conditional on the observed stimuli, under the cognitive cost. The optimal distribution of responses depends on both the (prior) distribution of stimuli and the form of the reward function. We implement this model and compute its BIC (see Methods for details). We find that it is much larger (by 29,656) than the BIC of the *OptimalInference+NoisyMax* model, which itself has a larger BIC than most of our other models (Fig. 11). Hence, a direct application of a rational inattention approach does not provide a better account of behavioral data than the addition of a Gaussian noise in response selection following optimal inference.

Faced with a similar issue, Khaw et al., 2017 introduced a variant of the rational-inattention model that is particularly relevant to our study, as it applies to a sequential inference task with (history-independent) change points. In this variant, the response selection is split into a two-stage decision process: first, the subject decides whether to repeat the previous response; second, only if the decision is made not to repeat, then the subject chooses the location of a new response; and both decision stages are

subject to cognitive costs. This presents a difference with the models that we have analyzed so far, in that an *ad hoc* probability of repetition is included explicitly in the model, whereas our approach, instead, was to study deviations from optimality that resulted from deterministic or stochastic approximations of a Bayesian scheme.

We have analyzed our data using a model similar to the one proposed by Khaw et al., 2017. In order to evaluate the relevance of a rational-inattention information structure, we have also studied a model that include a two-stage decision process, but does not involve cognitive costs. Specifically, this model combines the *OptimalInference* strategy with a strategy of response selection in which at each trial the model subject chooses, with fixed probability, whether to repeat the previous response. The probability of a repetition is constant, in this last model, whereas in the rational-inattention model it depends on the stimulus history and on the location of the slider at the beginning of the trial.

These two models yield a BIC lower than that of our previously best-fitting model, suggesting that a two-stage response-selection process is worth considering as a candidate mechanism for sequential decision-making. The previously best-fitting model, however, makes use of the *ParticleFilter* inference strategy, whereas the models just presented rely on the *OptimalInference* strategy. Hence, we implement an array of models that combine the same two-stage response-selection processes with, instead, the *ParticleFilter* inference strategy. With this inference strategy, the model in which the subject chooses whether to repeat with a fixed repetition probability results in a lower BIC (by 1,398) than the model in which the repetition probability is governed by a rational-inattention cognitive cost. Moreover, as in the other analyses conducted above, the models with a *ParticleFilter* inference strategy all yield substantially lower BICs than their counterparts that make use of the *OptimalInference* strategy. It appears, thus, that while the introduction of an explicit repetition probability in models improves their explanatory power, deriving this repetition probability from a simple form of cognitive cost does not provide a better account of behavioral data than positing a fixed repetition probability.

An explicit repetition probability alone is insufficient to capture human inference in our task. Instead, stochastic compression of beliefs, as illustrated by stochastic pruning or particle filtering, results in a closer match with experimental observations. (We provide details on the rational-inattention models, the fixed-repetition-probability models, and their BICs in Methods.)

Memory load and stochastic pruning during inference

After rejecting the rational inattention and the sampling hypotheses for response selection, we are left with an unexplained modulation of the variability – specifically, the relation between the magnitude of behavioral variability and the width of the Bayesian posterior at successive times (Fig. 4). Noisy maximization, which makes use of an additive random perturbation with fixed variance, leads to behavioral variability with constant variance; it is thus insufficient to explain the experimental observations. If modulated variability does not originate in the response-selection step, it must derive from the inference step. Out of the 28 models we consider, the five best-fitting models implement either the *ParticleFilter* or the τ*Sample* inference strategy, both of which rely on sampling *during inference*. These strategies capture the trends in the variability of human responses, in both HI and HD conditions (Fig. 9).

Both these inference strategies reduce the memory load in the inference problem by stochastically trimming the posterior, in a fashion akin to the 'pruning' model proposed by Huys et al. (2012) in the context of a decision-tree task. In their decision model, the evaluation of a possible sequence of decisions is more likely to be curtailed (thus alleviating the dimensionality of the problem) if it appears to have a low value. Similarly, the *ParticleFilter* and the τ*Sample* inference strategies ignore with a higher probability possible run-lengths and states that are less likely to be correct. Furthermore, we note that the τ*MaxProb* inference strategy also relies on pruning unlikely run-lengths, but it deterministically eliminates the most unlikely, in contrast to its stochastic counterpart,

τ*Sample* — which yields a better fit of the data. In explore-exploit problems, 'Thompson sampling' (Thompson, 1933) refers to a strategy in which one 'explores' by randomly choosing an action with the probability that this action maximizes the reward (instead of deterministically choosing the action most likely to maximize the reward). Several studies have reported that the responses of human subjects in explore-exploit tasks appeared consistent with Thompson sampling (Gershman, 2018; Schulz, Konstantinidis, & Speekenbrink, 2015; Speekenbrink & Konstantinidis, 2015). In this perspective, the stochastic pruning of the posterior in our best-fitting models appears as an exploration strategy, deployed during inference. Beyond the specifics of the pruning or exploration mechanisms, the main conceptual point, here, is that behavioral biases may result from an approximation to a Bayesian scheme that relieves memory load. A similar picture has been advanced in the context of prediction tasks, where 'over-reaction' – effectively a biased, enhanced learning rate – results from the compression of information stored in memory (Afrouzi, Kwon, & Ma, 2020; Azeredo da Silveira & Woodford, 2019; Neligh, 2019).

Aside from stochasticity in the inference step and in the response-selection step, noise in the sensory observation is a possible alternate account of behavioral variability, discussed, among others, by Stocker and Simoncelli (2006) and Drugowitsch et al. (2016). The design of our experiment, however, minimized perceptual ambiguity: the stimulus we presented to the subject at each trial, in our task, was a white dot that clearly contrasted with the background, and which remained on the screen until the subject responded (the subject was thus free to look at it for as long as he or she wished). By contrast, in the two studies just mentioned, the stimuli consisted of low-contrast gratings presented for one second or less. We presume that our experimental design limited perceptual noise.

It would nevertheless be interesting to know whether and how perceptual noise may contribute to behavioral variability in a sequential experiment, and how it may couple with stochasticity in inference. A possibility is that the magnitude of perceptual noise is constant throughout the task, in which case it would be expected to contribute an equal

amount of variability at all run-lengths; if so, it would not account for the modulations of variability that we record in our task. Another possibility is that perceptual noise itself adapts dynamically during the task. Under this hypothesis, we speculate that the magnitude of perceptual noise would decrease if uncertainty increases; if so, it would result in an effect on behavioral variability *opposite* to the observed effect. Although we cannot exclude that an effect of this nature is at play, it does not appear to offset completely the modulations of the variability which can be understood in terms of an approximate Bayesian inference. In sum, in the present setting of the experiments, a natural explanation of the behavioral variability in terms of perceptual ambiguity seems unlikely.

Incorrect belief about the temporal statistics of the signal

Comparing the behaviors of the best-fitting model and that of the subjects, we note that there remain discrepancies between the two, particularly at long run-lengths in the HD condition. The increase in the subjects' learning rate, and the reduction in their repetition propensity, at these run-lengths and in this condition, are not as sharp as those of the best-fitting model (Fig. 9). A candidate explanation of these deviations is that the subjects hold an inexact belief on the shape of the change probability, $q(\tau)$, as a function of the run-length. The analysis of the *IncorrectQ* inference strategy, in Methods, examines the case of a model subject who believes that the change probability increases more slowly ($\lambda < 1$) than it actually does and that the average interval length is greater ($T > 10$) than it actually is in our task. The learning rate of this model subject does not increase as abruptly, and the repetition propensity does not decrease as quickly, as those of the best-fitting model, similarly to the behavior of actual subjects (Fig. 12, middle panels). Moreover, among the five deterministic inference strategies considered, *IncorrectQ* is the best-fitting strategy, regardless of the response-selection it is combined with, and with both model-comparison measures — NMSE (Table 1) and BIC (Fig. 11). This suggests that, aside from the stochastic compression of the posterior, the subjects' deviations from

optimality may also result, to some extent, from an incorrect belief in the temporal structure of the signal.

Inference through sample-based representations of probability

The *ParticleFilter* strategy is noteworthy in a number of respects. First, it is our best-fitting model. Second, it constitutes a generic approach to inference; it was reported to account successfully for other inference and learning behaviors, such as category learning (Sanborn, Griffiths, & Navarro, 2006, 2010), conditioning in pigeons (Daw & Courville, 2008), sentence processing (Levy et al., 2008), hidden state inference (Brown & Steyvers, 2009), and visual tracking of multiple objects (Vul, Frank, Tenenbaum, & Alvarez, 2009). Third, out of all the models we consider, it is by far the less demanding on memory: with nine particles, one needs to store 27 numbers (for s, τ, and the weight of each particle) in memory. As a comparison, the optimal model stores a discretized probability distribution over the (s, τ) space, which amounts to about 16000 numbers (the optimal posterior could be well approximated with less memory-intensive methods, but this would require further hypotheses.) Previous uses of particle-filter methods in the context of a variety of cognitive tasks yielded best-fitting numbers of particles which ranged from one to several hundreds: from one to 400 particles, with a mean of 56, in Brown and Steyvers (2009); 130 particles (but 70 when subjects simultaneously perform a distractor task) in Thaker, Tenenbaum, and Gershman (2017); around 20 particles in Levy et al. (2008); 20 particles in Glaze et al. (2018); and as few as one particle in Daw and Courville (2008) and Sanborn et al. (2010). In addition, Bramley, Dayan, Griffiths, and Lagnado (2017) consider only the case of a single particle. We note that our analysis of the *ParticleFilter* inference strategy, detailed in Methods, reveals that a model with just one particle fails to reproduce the decreasing learning rates, in the HI condition, and the smile shape of the learning rates, in the HD condition, while models with two or more particles do capture these behavioral trends (Fig. 15B). Hence, in contrast to the last three studies cited, we find that a particle-filter model

with a single particle is qualitatively inconsistent with the behavior of human subjects, at least in the context of our task.

A fourth aspect of particle filters is that they provide a natural interpretation of the high repetition propensity observed in subjects (Fig. 5B). As the support of the probability distribution is reduced to $N_P = 9$ points on the (s, τ) plane, there is a fair chance that the posterior-maximizing particle at trial t, (s_t, τ_t), remains the posterior-maximizing particle at trial $t + 1$. Hence, the response s_t is likely to be repeated. In a similar spirit, particle filters have been shown to account for order effects in category learning (Sanborn et al., 2010) and observations about online sentence comprehension (such as the processing of 'garden-path sentences' (Levy et al., 2008)).

The success of particle filters, also known as Sequential Monte Carlo method, in accounting for human behavior in an online inference task adds to a growing literature on sample-based representations in cognitive processes (Gershman et al., 2012; Goodman et al., 2008; Moreno-Bote et al., 2011; Vul et al., 2014). Monte Carlo methods, which approximate probability distributions with sets of samples, constitute a major element of a family of techniques used in machine learning to address a wide range of problems (inference, optimization, numerical integration, etc); they have also been put forth as candidate cognitive algorithms (Gershman & Beck, 2016; Sanborn, 2015). Moreover, they account for a range of cognitive biases in the laboratory, such as base-rate neglect, conjunction fallacy, and the unpacking effect, as well as for human performance in complex, real-world tasks, and specific observations such as response variability and autocorrelation in perception and reasoning tasks (Gershman et al., 2012; Sanborn & Chater, 2016). At the implementation level, sample-based representations are well suited to learning in neural networks (Fiser, Berkes, Orbán, & Lengyel, 2010). Here, the variability in neural activity can be interpreted in terms of sampling-based representations of probability (Buesing, Bill, Nessler, & Maass, 2011; Fiser et al., 2010; Gershman et al., 2012; Hoyer & Hyvärinen, 2003), and a number of neural network models performing probability sampling have been

proposed (Aitchison & Lengyel, 2016; Hennequin, Aitchison, & Lengyel, 2014; Moreno-Bote et al., 2011; Savin & Denève, 2014; Shi & Griffiths, 2009).

Methods

Details of the behavioral task

The computer-based task was programmed and run with Psychopy (Peirce, 2009). In this task, white dots appeared on a horizontal line in the middle of a grey screen. Subjects were told that these white dots were snowballs thrown by a hidden person, the 'enemy' (also located on the horizontal line). The horizontal location of a snowball was the stimulus, x_t, and the position of the hidden person was the state, s_t. The state space was arbitrarily chosen to be [0, 300]; this scale did not appear on the screen. By clicking with a mouse (whose pointer moved on the horizontal axis only), subjects could indicate where they thought the hidden person was (i.e., give their estimate, $\hat{s}_t$, of the state). The time of response was not constrained. A green dot provided a visual feedback of the location of the click. After 100ms, a new white dot appeared, starting the next trial (Fig. 1A-B). If a subject's 'shot' was within a fixed distance around the state (the radius of the enemy), the subject was rewarded with 1 point. If the shot was 'outside the enemy' but within a distance equal to twice the enemy radius, the reward was 0.25 point (Fig. 1E). Otherwise, the reward was zero. Subjects were not informed of the reward immediately after each shot, as this would have provided additional information on the location of the state. The total score was given every 100 trials, to allow for an assessment of average self-performance and to foster motivation.

Subjects

We ran the computer-based task on 30 paid subjects; all gave informed consent. The study was approved by Princeton University's Institutional Review Board for Human Subjects. The sample size was determined so as to be comparable to that used in similar

experiments (Nassar et al., 2012; Nassar et al., 2010). Four subjects performed significantly worse than the other ones: their average error, defined as the absolute difference between their estimate and the state, $|\hat{s}_t - s_t|$, was 10.4 (standard deviation (s.d.): 0.93), while the average error of the other 26 subjects was 6.5 (s.d.: 0.62). Because of this difference of more than 5 standard deviations, these four subjects were excluded from the analyses. Hence, a total of 26 subjects were included in the analyses. Our conclusion remain unchanged if all 30 subjects are included in the analyses (see Supplementary Fig. B4).

Details of the signal

The stimulus, x_t, was generated around the state, s_t, according to the likelihood probability, $g(x_t|s_t)$, which was chosen to be triangular, centered at s_t, and of half-width 20. The state, s_t, was piecewise-constant with respect to time, i.e., constant in the absence of a change point. In the HI condition, the probability of a change point, q_t, was constant and equal to 10%. In the HD condition, q_t depended on the run-length, τ_t, defined as the number of trials since the last change point, and had a sigmoid shape: $q_t(\tau_t) = 1/(1 + e^{-(\tau_t - 10)})$. At $\tau_t = 0$ (i.e., immediately after a change point), the probability of another change point was very small. Six trials after a change point it was still small, less than 2%, before growing appreciably (50% at$\tau_t = 10$, 95% at $\tau_t = 13$). This led to more regular intervals between change points than in the HI condition, with a change point roughly every 10 trials (Fig. 1C-D). The average number of trials between two change points in both conditions was 10. At a change point, the state randomly jumped to a new state, s_{t+1}, according to the state transition probability, $a(s_{t+1}|s_t)$. This distribution was chosen to be bimodal, symmetric, and centered at s_t (two triangles of half-width 20 each, centered at $s_t \pm d$, where $d = 25$, Fig. 1E). This prevented new states to be too close or too far from the previous state, which would have made change-point detection too difficult or too obvious.

Training runs

All subjects did the task in both HI and HD conditions. 14 started with the HD task and 16 started with the HI task (no significant difference were found in results between these two groups). Subjects were not told the specificity of each situation. An explanatory text indicated that there were "differences" between each condition but no further indications were given. Each condition started with a series of explanations and tutorial runs. In a first tutorial run, the enemy (i.e., the state) was visible and moved according to the current (HI or HD) condition, and successive snowballs appeared without any action from the user (as in passive video viewing). In a second run, the enemy was still visible and subjects had to click at each trial, after which the next snowball would appear. This run was a very simple version of the actual task, because subjects were seeing the state. In a third run, the half-width of the triangular likelihood, $g(x_t|s_t)$, was 10, i.e., half the value it took in the actual task. In this run, the state was not visible, except after a change point: in the occurrence of a change point, the position of the state before the change point was shown, along with the shots of the subject since the previous change point. This run had two goals: first, to emphasize the timing of change points, and, second, to allow for self-performance assessment and to illustrate that a strategy consisting in 'following the white dots', i.e., clicking on the stimuli, was inefficient. A fourth tutorial run was an 'easy' version of the actual task: the state was always hidden, but the likelihood, g, had a half-width of 10. A fifth and last tutorial run reproduced the third run, but with the likelihood, g, with half-width 20. During the task, 15 subjects (7 amongst the HD-first group and 8 amongst the HI-first group) were also shown the positions of past stimuli, as white dots with decreasing contrast, gradually merging with the grey background (Fig. 1). The other 15 subjects were not shown past stimuli. No significant differences were found in data between the two groups. The number of stimuli presented in the tutorial runs totaled 297 for each condition. During the actual task there were 1000 trials in each condition, leading to a total of 2000 data points per subjects.

Empirical run-length

As subjects did not know the true run-length, τ, we computed an empirical run-length, $\tilde{\tau}$, based on the responses of subjects. Whereas the true run-length is defined as the number of trials since the last change point, the empirical run-length is defined as the number of trials since the last 'large correction'; a large correction is defined as a correction with absolute value larger than the 90th percentile of corrections. This percentile level is chosen in relation to the average frequency of change points, 1 for every 10 trials, in both HI and HD conditions. In some occasions, a subject "misses" a change point: the run-length and the empirical run-length, consequently, differ. For instance, $\tilde{\tau} = 10$, while $\tau = 0$ or 1. In such a case, because the change point did occur, the subject experiences a large surprise and is thus likely to subsequently opt for a large correction, i.e., to increase the learning rate. In the HD condition, because of the temporal statistics of change points, this situation is more likely to occur at empirical run-lengths around 10. Hence, this effect could bias the learning rates to higher values at these empirical run-lengths, in this condition. This effect, however, does not originate in the inference process, but rather in the temporal statistics of the HD signal. In other words, even an observer whose inference algorithm is not adapted to the HD condition would have higher learning rates at empirical run-lengths around 10. In the results presented, we removed all trials with a (true) run-length of 0 or 1, in order to avoid this artifact.

Regression of the learning rate on run-lengths

In order to provide statistical evidence of the smile shape of the learning rates in the HD condition (and the absence of a smile shape in the HI condition), we regress the learning rate on the run-lengths, with a quadratic term. For the HD condition, we find that the coefficient for the quadratic term is positive and significantly different from zero (0.0046; p-value = 1e-11). For the HI condition, this coefficient is smaller and we cannot reject at a significance level of 5% the null hypothesis that it is zero (0.0013; p-value = 0.068).

Moreover, the difference between the two quadratic coefficients is statistically significant (F-test p-value = 5.7e-4). The coefficient for the linear term is significantly negative (p-value < 1e-2).

Bayesian update equation

We derive the Bayesian update equation for a general case with $q = q(s_t, \tau_t)$, $p(x_t|s_t, \tau_t) = g(x_t|s_t, \tau_t)$, and $p(s_{t+1}|\tau_{t+1} = 0, s_t, \tau_t) = a(s_{t+1}|s_t, \tau_t)$, which includes the case used in our task with $q = q(\tau_t)$, $g = g(x_t|s_t)$, and $a = a(s_{t+1}|s_t)$.

Our goal is to obtain an update rule for the posterior, $p_t(s, \tau|x_{1:t})$ upon the observation of a new stimulus, x_{t+1}. Bayes' rule yields

$$p_{t+1}(s, \tau|x_{1:t+1}) = \frac{1}{Z_{t+1}} g(x_{t+1}|s, \tau) p_{t+1}(s, \tau|x_{1:t}), \tag{4}$$

where $Z_{t+1} = p_{t+1}(x_{t+1}|x_{1:t})$ is a normalization constant. The third term in this product can be written as

$$p_{t+1}(s, \tau|x_{1:t}) = \sum_{\tau_t} \int_{s_t} p_{t+1}(s, \tau|s_t, \tau_t) p_t(s_t, \tau_t|x_{1:t}) \mathrm{d}s_t. \tag{5}$$

The transition probability, $p_{t+1}(s, \tau|s_t, \tau_t)$, is determined by q and a. An absence of change point occurs with probability $1 - q(s_t, \tau_t)$, and in such a case a state (s_t, τ_t) evolves into the state $(s_{t+1} = s_t, \tau_{t+1} = \tau_t + 1)$. In the case of a change point, an event which occurs with probability $q(s_t, \tau_t)$, possible states at $t + 1$ have the form $(s_{t+1}, \tau_{t+1} = 0)$. Hence the transition probability from (s_t, τ_t) to (s, τ) at $t + 1$ is given by

$$p_{t+1}(s, \tau|s_t, \tau_t) = \mathbb{1}_{\tau=0}\, q(s_t, \tau_t) a(s_t, \tau_t, s) + \mathbb{1}_{\tau=\tau_t+1, s=s_t}(1 - q(s_t, \tau_t)). \tag{6}$$

Combining Equations (4), (5), and (6), we obtain the Bayesian update equation, as

$$p_{t+1}(s,\tau|x_{1:t+1}) = \frac{1}{Z_{t+1}} g(x_{t+1}|s,\tau) \Bigg[\mathbb{1}_{\tau=0} \sum_{\tau_t} \int_{s_t} q(s_t,\tau_t) a(s_t,\tau_t,s) p_t(s_t,\tau_t|x_{1:t}) \mathrm{d}s_t + \mathbb{1}_{\tau>0} \left(1 - q(s,\tau-1)\right) p_t(s,\tau-1|x_{1:t}) \Bigg]. \tag{7}$$

In the special case with $q = q(\tau_t)$, $a = a(s_{t+1}|s_t)$, and $g = g(x_t|s_t)$, we obtain the slightly simpler Eq. (2). In addition, we note that, in the HI condition, the change probability is constant, $q(\tau) = q$; in this condition, we can marginalize over the variable τ to obtain a closed recursion over the state posterior, as

$$p_{t+1}(s|x_{1:t+1}) = \frac{1}{Z_{t+1}} g(x_{t+1}|s) \left[q \int_{s_t} a(s|s_t) p_t(s_t|x_{1:t}) \mathrm{d}s_t + (1-q) p_t(s|x_{1:t}) \right]. \tag{8}$$

Derivation of the suboptimal models and analysis of their behaviors

***IncorrectQ* model.** In the *IncorrectQ* model, the two quantities governing the shape of $q(\tau)$, λ and T, are treated as free parameters, and we explore how varying these parameters impacts behavior, as compared to the *OptimalInference* model.

Keeping T constant at 10, and varying λ from 0 (HI condition) to 1 (HD condition), we find that the learning rate as a function of the run-length gradually morphs from the HI, monotonically decreasing curve, to the HD, non-monotonic and 'smile-shaped' curve (Fig. 12B). A similar behavior obtains at any fixed value of T, with the difference that the minimum of the HD curve is shifted to smaller τ for smaller T, and to larger τ for larger T. In other words, for fixed T, a higher value of λ, i.e., a sharper slope of the change probability, leads to a higher learning rate at run-lengths comparable to T. (Note that, for $T = 20$, the minimum of the learning rate occurs at run-lengths larger than 10, hence the non-monotonicity is not apparent in Fig. 12B.) Conversely, for a fixed $\lambda > 0$, the minimum

of the learning rate occurs at a run-length comparable to T which, precisely, determines when change points become likely. Finally, for $\lambda = 0$, the change probability is constant and there is no increase in the learning rates; these are, however, slightly higher for smaller T, because in that case the change probability, $q = 1/T$, is larger, so a new stimulus is more likely to be interpreted as stemming from a change point.

A subtlety that any analysis has to grapple with is that the statistics of responses depend not only on the inference process, but also, of course, on the statistics of the stimuli. To tease the two effects apart, for each *IncorrectQ* model subjects (with differing values of λ and of T) we computed the response behavior in presence of either signal: the HI signal, characterized by a constant change probability, $q = 0.1$, and the HD signal, characterized by a change probability, $q(\tau)$, varying with the run-length as a sigmoid with parameters $\lambda = 1$ and $T = 10$. We note that the impact on behavior of changing the signal is modest, as compared to the impact of changing the model subject's beliefs (Fig. 12B). This indicates that the discrepancy in human behavior in the HI and HD conditions does not originate primarily from the statistics of the signals, but rather from the different beliefs on the temporal statistics of the signals, held by the subjects.

Paralleling the behavior of learning rates, the repetition propensity in the HD condition peaks earlier or later depending on the value of T, and its shallowness depend on the value of λ (Fig. 12C). A belief in a shorter average inter-change-point interval, T, leads to a smaller repetition propensity: assuming frequent change points enhances the frequency of changes in one's estimate.

Human subjects correctly believe that q is not constant in the HD condition, and they use this belief in their inference process, but they may hold an inexact representation of the shape of $q(\tau)$ (Fig. 12). This, however, is not sufficient to capture data quantitatively: subjects exhibit both higher learning rates and more frequent repetitions than in the optimal model (Fig. 5), an observation that cannot be explained by manipulating λ and T in the *IncorrectQ* model; in the latter, high learning rates are accompanied by lower

repetition propensity, and vice versa. Thus, and letting alone the issue of variability, an erroneous belief on the change probability, $q(\tau)$, is insufficient to model experimental data.

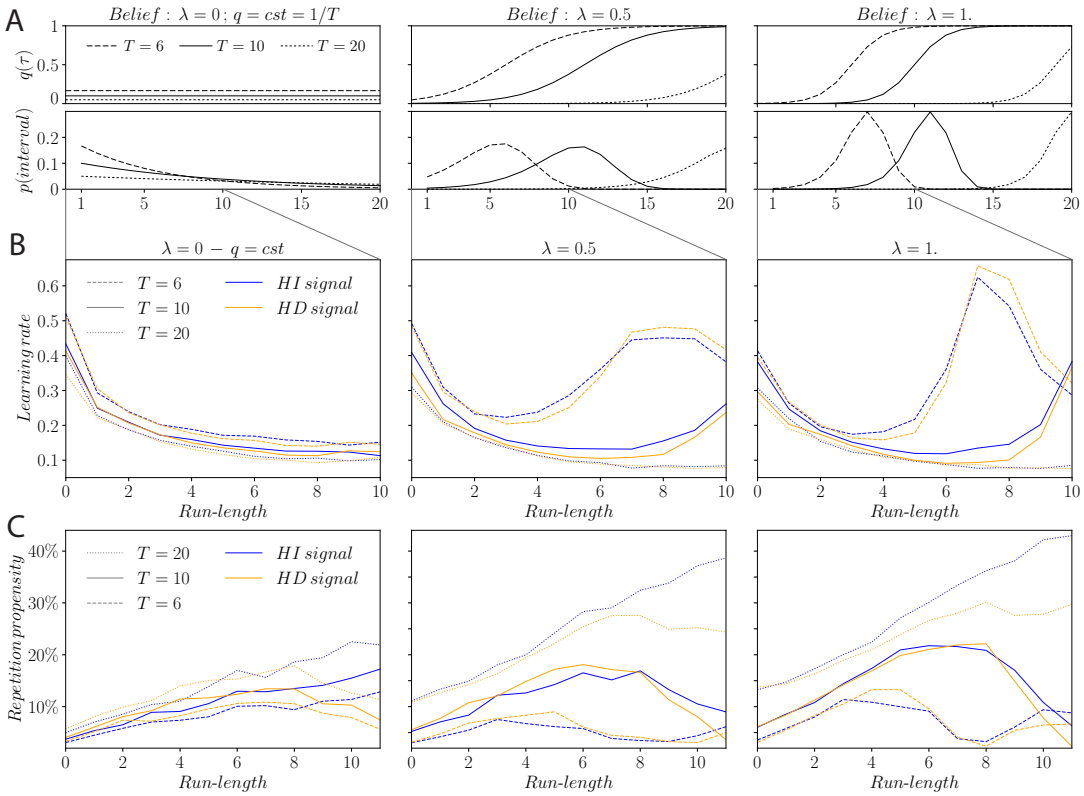

Figure 12. **Illustration of the *IncorrectQ* model with various beliefs on the shape of the change probability. A.** Examples of beliefs in *IncorrectQ* models. Change probability, $q(\tau)$ (top line), and resulting inter-change-point interval distribution (bottom line), for constant q (left column), sigmoid-shaped $q(\tau)$ with slope parameter $\lambda = 0.5$ (middle) and $\lambda = 1$ (right); and for average interval length, T, of 6 (dashed line), 10 (solid), and 20 (dotted). The 'true' HI signals used in the task correspond to $\lambda = 0$, $T = 10$, while the HD signals correspond to $\lambda = 1$, $T = 10$. **B, C.** Average learning rate (B) and repetition propensity (C) as a function of the run-length, in *IncorrectQ* models performing optimal inference with various beliefs on the change probability, $q(\tau)$, and presented with HI signals (blue) and HD signals (orange).

τ*Mean* model.

Derivation. This model is a generalization of the model introduced by Nassar et al. (2010). The approximate joint probability of the state and the run-length, which we denote by $\tilde{p}_t(s, \tau | x_{1:t})$, is assumed, in this model, to vanish at all values of the run-length, except for one, which we call the 'approximate expected run-length' and which we denote by $\bar{\tau}_t$.

Hence,

$$\tilde{p}_t(s, \tau | x_{1:t}) = \delta_{\tau, \bar{\tau}_t} \, \tilde{p}_t(s, \bar{\tau}_t | x_{1:t}), \tag{9}$$

where $\delta_{\tau,\bar{\tau}_t}$ is the Kronecker delta. As in the optimal model (see Eq. (2)), we use Baye's rule and the parameters of the task to derive the update equation, as

$$\begin{aligned} p_{t+1}(s, \tau | x_{1:t+1}) = \frac{1}{Z_{t+1}} g(x_{t+1}|s) \Bigg[& \mathbb{1}_{\tau=0} \, q(\bar{\tau}_t) \int_{s_t} a(s|s_t) p_t(s_t, \bar{\tau}_t | x_{1:t}) \mathrm{d}s_t \\ & + \mathbb{1}_{\tau=\bar{\tau}_t+1} \, (1 - q(\bar{\tau}_t)) p_t(s, \bar{\tau}_t | x_{1:t}) \Bigg]. \end{aligned} \tag{10}$$

This distribution is non-vanishing for two values of the run-length, 0 and $\bar{\tau}_t + 1$, which correspond to the two possible scenarios: with and without a change point at trial t. We use this distribution to compute the approximate expected run-length at trial $t+1$, $\bar{\tau}_{t+1}$, and the approximate posterior at trial $t+1$, $\tilde{p}_{t+1}(s, \tau | x_{1:t+1})$. First, we obtain the probability of a change point at trial $t+1$,

$$\Omega_{t+1} \equiv p_{t+1}(\tau = 0 | x_{1:t+1}) = \frac{1}{Z_{t+1}} q(\bar{\tau}_t) \int_s g(x_{t+1}|s) \int_{s_t} a(s|s_t) p_t(s_t, \bar{\tau}_t | x_{1:t}) \mathrm{d}s_t \mathrm{d}s, \tag{11}$$

and we use it to compute the approximate expected run-length at trial $t+1$:

$$\bar{\tau}_{t+1} = \Omega_{t+1} \cdot 0 + (1 - \Omega_{t+1})(\bar{\tau}_t + 1). \tag{12}$$

Second, we approximate the posterior (Eq. (10)) by marginalizing it over the run-lengths, and multiplying the result by a Kronecker delta which takes the value 1 at $\bar{\tau}_{t+1}$:

$$\begin{aligned} p_{t+1}(s | x_{1:t+1}) = \sum_\tau p_{t+1}(s, \tau | x_{1:t+1}) = p_{t+1}(s, \tau = 0 | x_{1:t+1}) + p_{t+1}(s, \tau = \bar{\tau}_t + 1 | x_{1:t+1}), \\ \tilde{p}_{t+1}(s, \tau | x_{1:t+1}) = \delta_{\tau, \bar{\tau}_{t+1}} p_{t+1}(s | x_{1:t+1}). \end{aligned} \tag{13}$$

This model has no parameter.

Behavior. While the optimal marginal distribution of the run-lengths, $p_t(\tau|x_{1:t})$, spans the whole range of possible values of the run-length, it is approximated, in the τ*Mean* model, by a delta function on a single value, $\bar{\tau}_t$. In the HI condition, the change probability, q, does not depend on the run-length, τ; the approximate joint distribution evaluated at $\tau = \bar{\tau}_t$, $\tilde{p}_t(s, \bar{\tau}_t|x_{1:t})$, is equal to the optimal posterior distribution on the state, $p_t(s|x_{1:t})$. (Compare Eq. (8) to the combination of Eqs. (10) and (13).) As a result, the τ*Mean* model computes the optimal posterior on the state, and, thus, the responses in this model are the same as those in the optimal model, i.e., the τ*Mean* model is optimal in the HI condition. In the HD condition, by contrast, the change probability, $q(\tau)$, depends on the run-length. The τ*Mean* model evaluates this function at only one run-length, $\bar{\tau}_t$, an approximation of the mean run-length; as compared to the optimal model, it fails to capture fully the consequences of the dependence of the change probability on the run-length (Fig. 1D). The learning rates in this model are higher than the optimal ones for short run-lengths, and lower than the optimal ones for long run-lengths (Fig. 13B); and the repetition propensities are lower than the optimal ones for short run-lengths, and higher than the optimal ones for long run-lengths (Fig. 13C).

τ*Nodes* model.

Derivation. This model generalizes the one introduced by Wilson et al. (2013). In this paper, the authors interpret a change-point setting similar to ours as a 'message-passing' graph where run-lengths are nodes, weighted by their marginal probability $p_t(\tau|x_{1:t})$, edges are characterized by the change probability, q, and 'messages' are passed along edges from one node to another. More precisely, we compute the marginal probability of the run-length, using Eqs. (5) and (6):

$$p_{t+1}(\tau|x_{1:t}) = \begin{cases} \sum_{\tau_t} q(\tau_t) p_t(\tau_t|x_{1:t}) & \text{if } \tau = 0 \\ (1 - q(\tau - 1)) p_t(\tau - 1|x_{1:t}) & \text{otherwise} \end{cases} . \tag{14}$$

Hence, at trial $t + 1$, the weight of a node τ (i.e., the marginal probability of the

corresponding run-length) is equal, if $\tau = 0$, to a sum of the marginal probabilities of all nodes at trial t, τ_t, weighted by their corresponding change probabilities, $q(\tau_t)$; and, if $\tau > 0$, it is the probability of the node $\tau - 1$, at trial t, weighted by the probability that there was no change, $1 - q(\tau - 1)$. Taking a different view, one can reformulate these weighted sums of probabilities as transfers of probability masses, as follows. Each node τ sends two 'messages': a 'no-change-point' message is sent to node $\tau + 1$ so as to set its weight to $(1 - q(\tau))p_t(\tau|x_{1:t})$, and a 'change-point' message is sent to node $\tau = 0$ to increase its probability by $q(\tau)p_t(\tau|x_{1:t})$. This is the message-passing algorithm. Wilson et al. (2013) assume a change probability, q, that is constant; a likelihood, g, which belongs to the exponential family; and a state transition probability, $a(s_{t+1})$, that does not depend on the previous state, s_t, and which is a conjugate prior of the likelihood. They show that in this case each node can be seen as implementing a delta-rule, and the optimal Bayesian model amounts to the weighted sum of these delta-rules.

The authors then 'reduce' this model by removing nodes and accordingly revise the message-passing algorithm and each node's update rule. We only focus on the aspects of the model that will be used in our *τNodes* implementation. The set of new nodes comprises 'virtual' run-lengths $l \in \{l^0, l^1, ..., l^N\}$. A node l^i, with $i \neq 0$, now sends three messages: one to l^0, one to the next node l^{i+1}, and one to itself. The 'change-point' message remains the same as in the previous algorithm, i.e., the quantity $q(l^i)p_t(l^i|x_{1:t})$ is sent to l^0 (i.e., this quantity is added to the probability of this node). The 'no-change-point' message is now split in two, one message being sent to the next node, l^{i+1}, and the other one being a self-passing message (i.e., sent to itself, l^i). The authors seek the relative weight $w(l^i)$ assigned to the self-passing message which gives an average run-length increase of 1 (i.e. $\mathbb{E}(l_{t+1}|l_t^i, \text{no change}) = l^i + 1$). They find $w(l^i) = \frac{l^{i+1} - l^i - 1}{l^{i+1} - l^i}$ for $i \neq N$ and $w(l^N) = 1$. The next node, l^{i+1}, hence receives the message $(1 - w(l^i))(1 - q(l^i))p_t(l^i|x_{1:t})$. With the assumptions mentioned above on q, g, and a, the model can again be understood as 'a mixture of delta-rules'.

We implement these ideas in our τ*Nodes* model. Instead of $p_t(s, \tau | x_{1:t})$, we consider the probability distribution $p_t(s, l | x_{1:t})$ and apply the same Bayesian and marginalization equations used in Eqs. (4) and (5). The main difference of the new model resides in the transition probability, $p_{t+1}(s, l | s_t, l_t)$, which becomes

$$\begin{aligned} p_{t+1}(s, l | s_t, l_t) = \mathbb{1}_{l=l^0}\, q(s_t, l_t) a(s_t, \tau_t, s) \\ + \mathbb{1}_{l=l_t, s=s_t}(1 - q(s_t, l_t)) w(l_t) \\ + \mathbb{1}_{l=l_t+1, s=s_t}(1 - q(s_t, l_t))(1 - w(l_t)). \end{aligned} \tag{15}$$

Combining Eq. (15) to Eqs. (4) and (5) adapted with l, we obtain an update equation similar to the full Bayesian update equation (Eq. (7)), with an additional term corresponding to the ability of nodes for self-passing messages. The model is parameterized by the number of nodes, N_τ, but also by the values of the nodes. We chose the possible values of the nodes to be in the set {0, 2.5, 5, 7.5, 10, 12.5}. When fitting the model for a given N_τ, all the models corresponding to every possible choice of N_τ nodes within these values, were computed, and the best-fitting one was chosen.

Behavior. In the HI condition, the model computes the optimal posterior on the state, $p_t(s | x_{1:t})$. Thus, as for the τ*Mean* model above, the responses in the τ*Nodes* model are the same as those in the optimal model, in the HI condition. In the HD condition, the greater the number of nodes, the more faithfully the model approximates optimal behavior (Fig. 13B,C). The learning rates are higher than the optimal ones for short run-lengths, and lower than the optimal ones for long run-lengths (Fig. 13B). The repetition propensities are higher than the optimal ones for long run-lengths; for short run-lengths, they are appreciably closer to the optimal ones, in the model with five nodes, than in the model with one node (Fig. 13C).

τ*MaxProb* model.

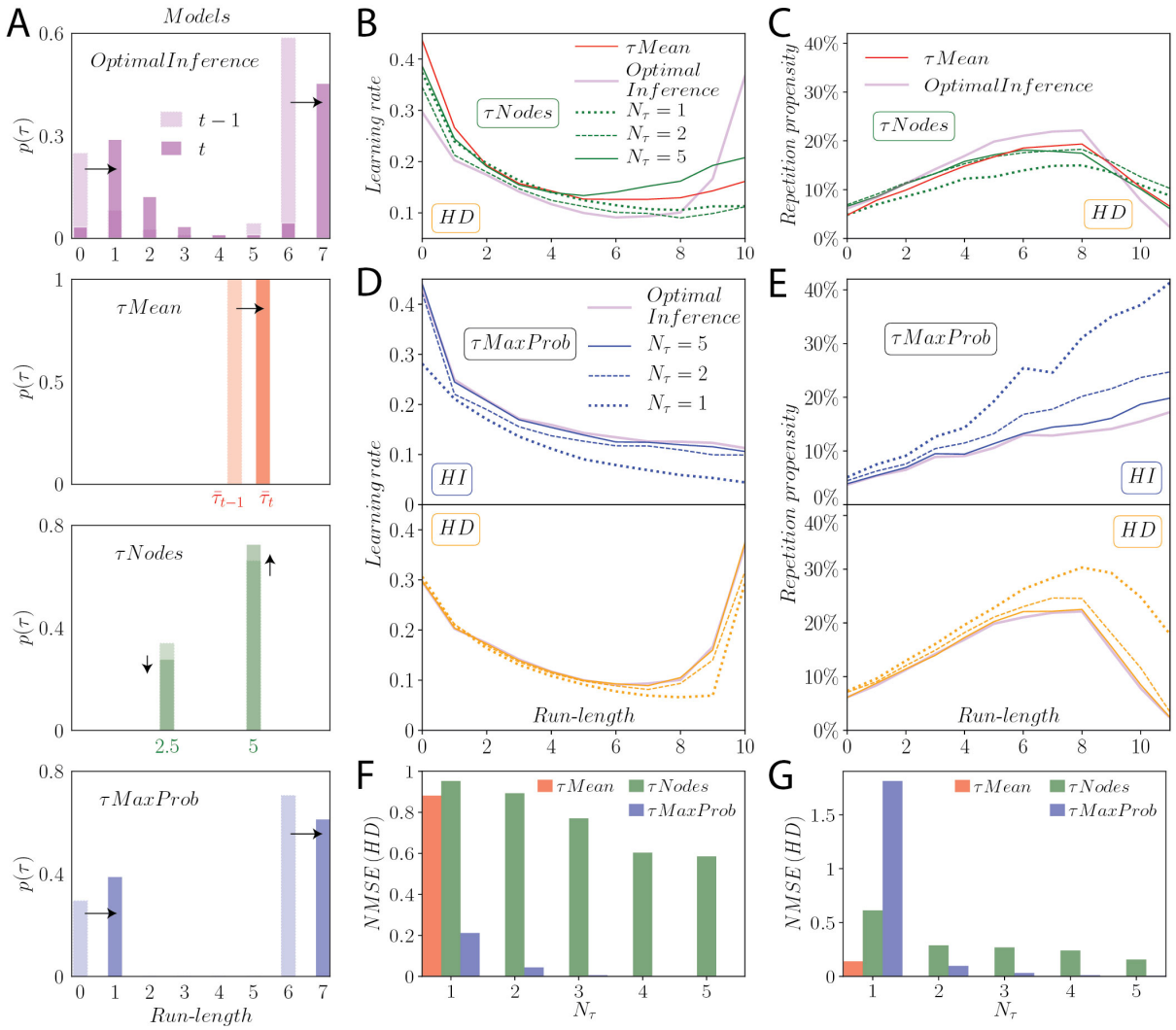

Figure 13. **Behavior of the limited-memory models, as compared to the *OptimalInference* model. A.** Schematic illustration of the marginal distribution of the run-length, $p(\tau)$, in each model considered. The *OptimalInference* model assigns a probability to each possible value of the run-length, τ, and optimally updates that distribution upon receiving stimuli (first panel). The τ*Mean* model uses a single run-length which tracks the inferred expected value, $\langle\tau\rangle$ (second panel). The τ*Nodes* model holds in memory a limited number, N_τ, of fixed hypotheses on τ ("nodes"), and updates a probability distribution over these nodes; $N_\tau = 2$ in this example (third panel). The τ*MaxProb* model reduces the marginal distribution by discarding less likely run-lengths; in this example, 2 run-lengths are stored in memory at any given time (fourth panel). **B, C.** Average learning rate (B), and repetition propensity (C), as functions of the run-length, in the *OptimalInference* model, the τ*Mean* model, and the τ*Nodes* model with $N_\tau = 1$, 2, and 5, in the HD condition. The HI condition is not displayed as the τ*Mean* and τ*Nodes* models do not differ from the *OptimalInference* model in this condition. *(Continued on the following page.)*

Figure 13. **D, E.** Average learning rate (D), and repetition propensity (E), as functions of the run-length, in the *OptimalInference* model and the τ*MaxProb* model with $N_\tau = 1$, 2, and 5, in the HI condition (top panels) and in the HD condition (bottom panels). **F,G.** Normalized Mean Squared Error relative to the learning rate (F) and the repetition propensity (G), as compared to the *OptimalInference* model, for the τ*Mean* model, which has no free parameter, and for the τ*Nodes* and τ*MaxProb* models with $N_\tau = 1$ to 5.

Derivation. In the τ*MaxProb* model, we assume that we have, at trial t, an approximation of the joint distribution of the state and the run-length, which we denote by $\tilde{p}_t(s, \tau | x_{1:t})$, and we assume that the approximate marginal distribution of the run-lengths, $\tilde{p}_t(\tau | x_{1:t})$, is non-vanishing for no more than N_τ values of the run-length. Upon receiving a new stimulus, we perform a Bayesian update of the approximate joint distribution, $\tilde{p}_t(s, \tau | x_{1:t})$, as in Eq. (2), and obtain the posterior, $p_{t+1}(s, \tau | x_{1:t+1})$, from which we derive the marginal distribution of the run-lengths, $p_{t+1}(\tau | x_{1:t+1})$. If, at trial t, the run-length takes a given value, τ_t, then, at trial $t+1$, it can only take one of two values: 0 (if there is a change point) or $\tau_t + 1$ (if there is no change point). Hence, if the marginal distribution at trial t, $\tilde{p}_t(\tau | x_{1:t})$, is non-vanishing for at most N_τ values, as we assume, then the updated distribution, $p_{t+1}(\tau | x_{1:t+1})$, is non-vanishing for at most $N_\tau + 1$ values. In the case that this distribution is non-vanishing for *less* than $N_\tau + 1$ values, we do not perform further approximations, at this stage. In the other, more generic case, i.e., if $N_\tau + 1$ values of the run-length have a non-zero probability, then we identify the most unlikely run-length, $\tau^* = \arg\min p_{t+1}(\tau | x_{1:t+1})$, and we approximate the posterior as

$$\tilde{p}_{t+1}(s, \tau | x_{1:t+1}) = \begin{cases} 0 & \text{if } \tau = \tau^* \\ \frac{1}{Z} p_{t+1}(s, \tau | x_{1:t+1}) & \text{if } \tau \neq \tau^* \end{cases}, \tag{16}$$

where Z is a normalization constant equal to $1 - p_{t+1}(\tau^* | x_{1:t+1})$.

Behavior. Even with $N_\tau = 1$ (one memory slot), the τ*MaxProb* model captures qualitatively the optimal, high learning rates for large run-lengths, in the HD condition (Fig. 13D, second panel, dotted line). However, in other situations (HD condition for

shorter run-lengths, and HI condition for all run-lengths), change points are not likely ($q < 0.5$). Hence, in most cases, a vanishing run-length, i.e., the hypothesis of a change point, minimizes the marginal distribution, $p_{t+1}(\tau|x_{1:t+1})$, and its probability vanishes in our approximation: $\tilde{p}_{t+1}(\tau = 0|x_{1:t+1}) = 0$. In other words, change points tend to go by undetected. Consequently, suppressed learning rates and enhanced repetition propensity obtain in a model with a single memory slot (Fig. 13D, E).

To compare our suboptimal models to the *OptimalInference* model, we compute their normalized mean squared errors (NMSE) with regard to the responses of the optimal model (as opposed to the responses of the human subjects, as we do in the main text). With a NMSE for learning rates at 0.21, the τ*MaxProb* model with $N_\tau = 1$ is closer to optimality than the τ*Mean* model (NMSE of 0.88) and the τ*Nodes* model with one node (NMSE of 0.95), in the HD condition (Fig. 13F). The high repetition propensity of the τ*MaxProb* model, however, leads to a larger error for this measure (NMSE of 1.81), as compared to the τ*Mean* (0.14) and τ*Nodes* (0.61) models (Fig. 13G). Adding a second memory slot allows for a better approximation of the marginal distribution, $p_t(\tau|x_{1:t})$, in the τ*MaxProb* model, as demonstrated by its close-to-optimal behavior with $N_\tau = 2$, both in terms of learning rates (NMSE of 0.043; compare to the τ*Nodes* model: 0.89) and repetition propensity (0.097; compare to the τ*Nodes* model: 0.29) (Fig. 13F, G).

τ*Sample* model. The τ*Sample* model is identical to the τ*MaxProb* model, except that the run-length τ^* is chosen randomly, i.e., sampled from the distribution $\left[1 - p_{t+1}(\tau|x_{1:t+1})\right]/z_{t+1}$, where z_{t+1} is a normalization factor. The stochastic nature of the update rule on the probable run-lengths influences the learning rate and the repetition propensity. In the case $N_\tau = 1$, there is, at trial t, a single run-length, τ_t, with non-vanishing probability. At trial $t + 1$, the model subject chooses randomly between the no-change-point scenario, with $\tau_{t+1} = \tau_t + 1$, and the change-point scenario, with $\tau_{t+1} = 0$. Hence, the model can incur 'false positives' (a change-point scenario is opted for in the absence of a true change point) and 'false negatives' (a true change point goes undetected

by the model subject), and these occur stochastically. In most trials, the change-point scenario is less likely than the no-change-point scenario; in the τ*MaxProb* model, the former would be eliminated, but it occurs with some probability in the τ*Sample* model, leading to false positives, which induce higher learning rates. Similarly, average learning rates of the τ*Sample* model are higher than the optimal ones (Fig. 14A). The false negatives, in which change points go undetected, result, as in the τ*MaxProb* model, in higher repetition propensities (Fig. 14C). With increasing memory capacity, N_τ, the behavior of the model approaches optimality, as reflected in the decrease of the NMSEs for the learning rates (Fig. 14B) and the repetition propensities (Fig. 14D).

A qualitatively new aspect brought in by the τ*Sample* model is the stochasticity in the inference step, which is reflected in behavioral variability and measured by the standard deviation of the responses of a model subject. Quantitatively, false negatives have a large impact on the behavioral variability. A model subject can however correct for a false negative during the few trials that follow a true change point, i.e., at short run-lengths. This occurs randomly, in the τ*Sample* model, resulting in variability in responses. At longer run-lengths, the posterior probability of a change point, $p_t(\tau = 0|x_{1:t})$, is dominated by the shape of the change probability, $q(\tau)$, rather than by the observed evidence. In the HI condition, q is constant; hence, the variability reaches a plateau for run-lengths larger than about 2 (Fig. 14E, top panel). In the HD condition, as $q(\tau)$ is an increasing function of the run-length, the variability increases for run-lengths larger than 5, resulting in the 'smile shape' of the curve (Fig. 14E, bottom panel). As the parameter N_τ is increased, the behavior of the model approaches optimality, and, correspondingly, the standard deviation of the responses of the model subject decreases (Fig. 14F).

ParticleFilter.

Derivation. The *ParticleFilter* approximates the posterior by a weighted sum of delta functions (Eq. (3)). To obtain the approximate posterior at trial $t + 1$, upon receiving a new observation, x_{t+1}, we start by writing the Bayesian update (Eq. (2)) of the

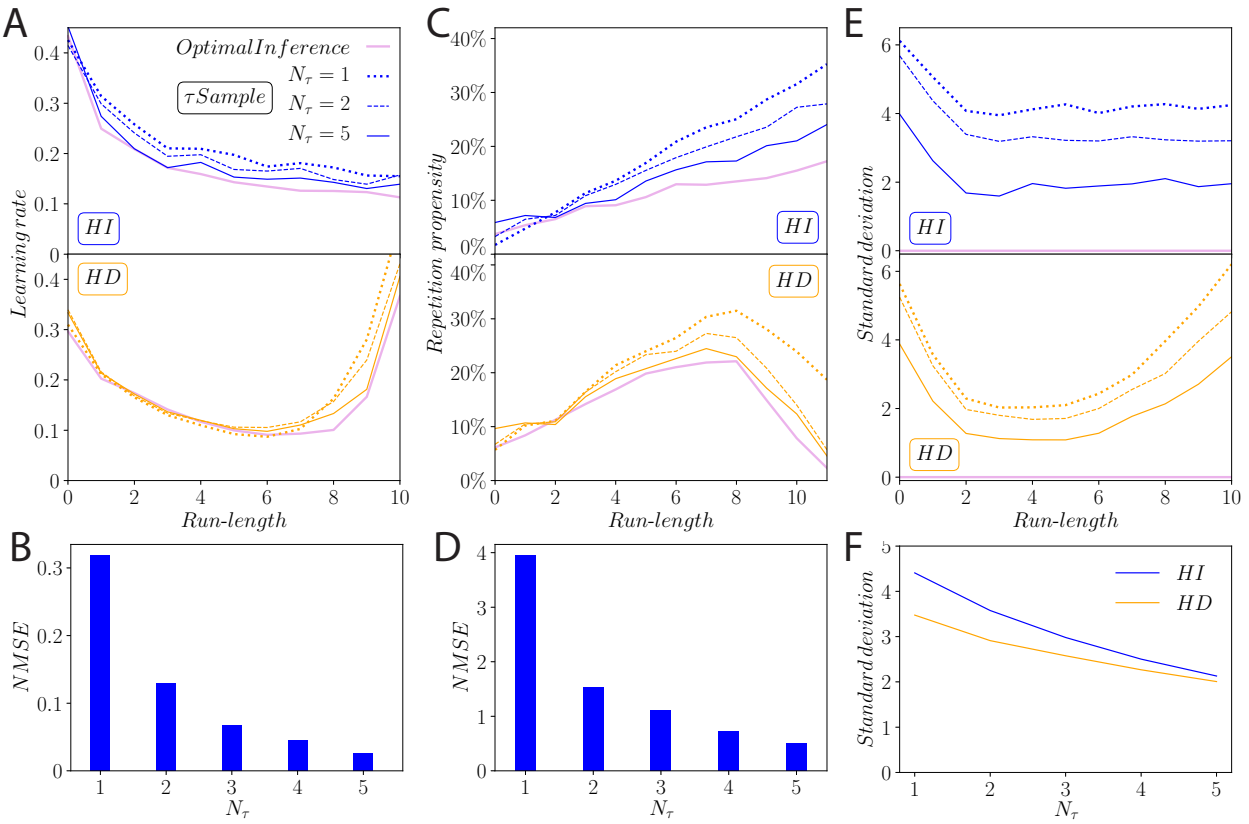

Figure 14. **Behavior of the *τSample* model, as compared to the *OptimalInference* model. A, C, E.** Average learning rate (A), repetition propensity (C), and standard deviations of responses (E), as a function of run-length, in the *OptimalInference* model and the *τSample* model with $N_\tau = 1$, 2, and 5, in the HI condition (top panels) and in the HD condition (bottom panels). **B, D.** Normalized Mean Squared Error on learning rates (B), and on repetition propensity (D), as compared to the *OptimalInference* model, for the *τSample* model, with $N_\tau = 1$ to 5. **F.** Standard deviation of the responses of the *τSample* model, as a function of N_τ.

approximate posterior at trial t, $\tilde{p}_t(s, \tau | x_{1:t})$, as

$$p_{t+1}(s, \tau | x_{1:t+1}) = \frac{1}{Z_{t+1}} g(x_{t+1}|s) \sum_{\tau_t} \int_{s_t} p_{t+1}(s, \tau | s_t, \tau_t) \tilde{p}_t(s, \tau | x_{1:t}) \mathrm{d}s_t, \tag{17}$$

with the transition probability, $p_{t+1}(s, \tau | s_t, \tau_t)$, defined in Eq. (6). Injecting the expression of the approximate posterior at trial t (Eq. (3)), we can rewrite the Bayesian update as a

sum of N_P functions:

$$p_{t+1}(s,\tau|x_{1:t+1}) = \frac{1}{Z_{t+1}} \sum_{i=1}^{N_P} w_t^i g(x_{t+1}|s) p_{t+1}(s,\tau|s_t^i,\tau_t^i), \tag{18}$$

where

$$p_{t+1}(s,\tau|s_t^i,\tau_t^i) = \sum_{\tau_t} \int_{s_t} p_{t+1}(s,\tau|s_t,\tau_t)\delta(s-s_t^i)\delta_{\tau,\tau_t^i}\mathrm{d}s_t. \tag{19}$$

The interpretation of this form becomes apparent if we introduce, for each particle, a probability distribution over (s,τ), defined as

$$\pi_{t+1}(s,\tau|s_t^i,\tau_t^i,x_{t+1}) \equiv \frac{g(x_{t+1}|s)p_{t+1}(s,\tau|s_t^i,\tau_t^i)}{p(x_{t+1}|s_t^i,\tau_t^i)}, \tag{20}$$

where the denominator is obtained by normalization,

$$p(x_{t+1}|s_t^i,\tau_t^i) = \sum_{\tau} \int_s g(x_{t+1}|s) p_{t+1}(s,\tau|s_t^i,\tau_t^i)\mathrm{d}s. \tag{21}$$

The distribution $\pi_{t+1}(s,\tau|s_t^i,\tau_t^i,x_{t+1})$ is none other than the Bayesian update of a single particle (i.e., Eq. (17) with the approximate prior, $\tilde{p}_t(s,\tau|x_{1:t})$, replaced by $\delta(s-s_t^i)\delta_{\tau,\tau_t^i}$), and the full Bayesian update is a weighted sum of these N_P functions:

$$p_{t+1}(s,\tau|x_{1:t+1}) = \sum_{i=1}^{N_P} \frac{w_t^i p(x_{t+1}|s_t^i,\tau_t^i)}{Z_{t+1}} \pi_{t+1}(s,\tau|s_t^i,\tau_t^i,x_{t+1}). \tag{22}$$

To complete the definition of the particle filter, we have to formulate a prescription for selecting the N_P particles at trial $t+1$. Following the literature, instead of sampling the full Bayesian update, $p_{t+1}(s,\tau|x_{1:t+1})$, we sample independently each component of the mixture, $\pi_{t+1}(s,\tau|s_t^i,\tau_t^i,x_{t+1})$, to obtain the updated particles, (s_{t+1}^i,τ_{t+1}^i). To each sample, i.e., to each particle, is assigned the weight of the corresponding component in the mixture, $w_{t+1}^i = w_t^i p(x_{t+1}|s_t^i,\tau_t^i)/Z_{t+1}$. In the rare cases in which $p(x_{t+1}|s_t^i,\tau_t^i) = 0$, i.e., if new data invalidate particle i, and thus, $\pi_{t+1}(s,\tau|s_t^i,\tau_t^i,x_{t+1}) = 0$, we resample a new particle i from

the other particles.

In practical applications of particle filters, there exists a 'weight degeneracy' risk, whereby the weight of one particle may overwhelm the combined weight of the others. A common method to mitigate this shortcoming is called 'resampling'. It is a stochastic method in which the particles with high weights are likely to be duplicated, while the particles with low weights are likely to be eliminated. To achieve this, we use the N_P-dimensional categorical distribution parameterized by the N_P weights of the particles, i.e., $p(j) = w_t^j$. We sample this distribution N_P times, and obtain, thus, a set of N_P indexes, $\{j_i\}_{i=1}^{N_P}$. We use those to define the new N_P particles: for each particle i, we replace (s_t^i, τ_t^i) by $(s_t^{j_i}, \tau_t^{j_i})$, and we set all the weights to $1/N_P$. In other words, the set of particles is randomly sampled with replacement, N_P times. Particles with low weights are unlikely to survive this scheme, as compared to particles with high weights. For the sake of simplicity, we resample at each trial.

A possible consequence of resampling is the 'sample impoverishment' problem, i.e., the loss of particle diversity (all particles end up bearing the same state). A common procedure in the particle-filter literature that addresses this problem is 'particle rejuvenation', which increases the variability of particles by 'jittering' their parameters. Here, however, this issue is mitigated naturally by the structure of our problem, as a new state is sampled from the distribution $a(s|s_t)$ every time a new particle carries a change-point run-length ($\tau = 0$), thus renewing the set of particles. In addition, introducing a rejuvenation step implies choosing arbitrarily a transition kernel and an acceptance rule for candidate particles (usually, the Metropolis-Hastings rule is adopted). Many kernels used in the literature come with additional parameters. While the rejuvenation method would be an interesting addition to our model, the performance of our implementation of the particle-filter model does not warrant the introduction of this new layer of complexity.

Behavior. With a single particle ($N_P = 1$), the posterior is reduced to a unique sample, (s_t^1, τ_t^1), and, thus, the model subject has access to a single 'hypothesis' on the

change probability, $q(\tau_t^1)$. The particle can then evolve in one of two ways: either it opts, with this probability, for a change-point scenario, in which the new stimulus, x_{t+1}, is the only information available on the new state, along with the prior transition probability, and thus the learning rate is close to 1, or, with probability $1 - q(\tau_t^1)$, a no-change-point scenario is opted for, and the particle stays put ($s_{t+1}^1 = s_t^1$), i.e., the learning rate vanishes. As a result, when averaged over several instantiations of the particle filter, the behavior of the learning rate as a function of run-length resembles that of the change probability, $q(\tau)$, i.e., constant in the HI condition, and increasing in the HD condition (Fig. 15B), a behavior qualitatively different from either that of the *OptimalInference* model or the human responses. But it is sufficient to add no more than a second particle for the model to capture the main trends in the learning rate (a decreasing learning rate in the HI condition and smile shape in the HD condition). The NMSE drops sharply from 2.2 for $N_P = 1$ to less than 0.7 for $N_P = 2$. As additional particles are included in the model, the latter approaches optimality (the NMSE becomes less than 0.1 for $N_P \geq 70$) (Fig. 15C).

As mentioned, sampling in the particle filter induces variability in behavior: two particle filters receiving the same sequence of observations do not respond with the same sequence of estimates. Since the stochasticity stems from the sampling of an (approximate) posterior, the resulting variability scales with the width of the posterior. As measured by the standard deviation of responses, it decreases with the run-length, in the HI condition. In the HD condition, it decreases at short run-lengths before increasing at longer run-lengths (Fig. 15F). This behavior reproduces, at least qualitatively, that of the subjects (compare to Fig. 5C). The greater the number of particles in a particle filter, the closer the latter approximates the *OptimalInference* model; the standard deviation of the responses is a decreasing function of the number of particles (Fig. 15G).

Since it operates on a low-dimensional spatial representation, the particle filter naturally predicts a higher repetition propensity than the *OptimalInference* model does. More specifically, the posterior is non-vanishing for only a finite (possibly small) set of values at

each trial, and it is more likely than in the optimal case that the subject model's estimate remains unchanged following stimulus presentation. This effect is quantitatively appreciable, and leads to repetition propensities which are multiples of those in the *OptimalInference* model. Again, the repetition propensities decrease toward their optimal values as the number of particles, N_P, increases. The corresponding NMSE drops from 202 for $N_P = 1$ to 24 for $N_P = 150$ (Fig. 15D, E).

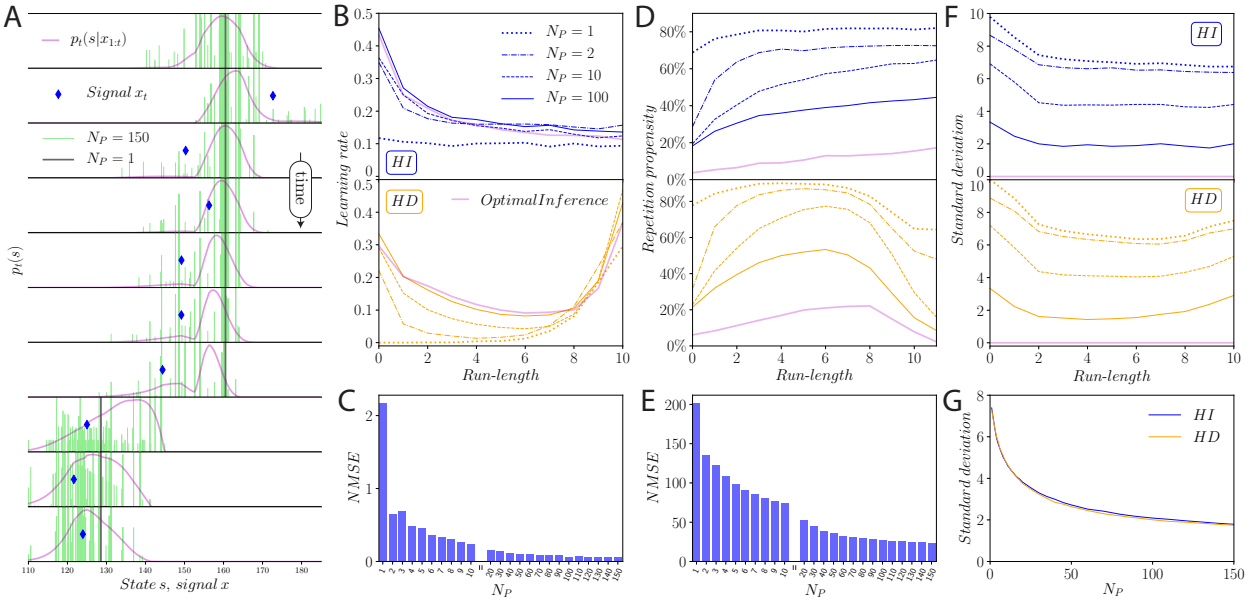

Figure 15. **Illustration of the *ParticleFilter* model and its behavior. A.** Distribution of particles during inference, compared with the optimal posterior distribution, for particle filters with N_P =1 and 150. Each particle is a point on the (s, τ) plane, equipped with a weight. Only the spatial components s are represented here, as vertical bars (grey for $N_P = 1$, green for $N_P = 150$). Bars heights are proportional to the corresponding weights, but some are truncated due to the choice of scale, which emphasizes weight diversity. Upon receiving a new stimulus, x_{t+1} (blue), a particle i is updated by sampling $p_{t+1}(s, \tau | s_t^i, \tau_t^i, x_{t+1})$. This may or may not involve a change point, in which case $s_{t+1}^i \neq s_t^i$. **B, D, F.** Average learning rate (B), repetition propensity (D), and standard deviations of responses (F), as a function of run-length, in the *OptimalInference* model and the *ParticleFilter* model with $N_P = 1$, 2, 10, and 100, in the HI condition (top panels) and in the HD condition (bottom panels). **C, E.** Normalized Mean Squared Error on learning rates (C), and on repetition propensity (E), as compared to the *OptimalInference* model, for the *ParticleFilter* model, with $N_P = 1$ to 150. **G.** Standard deviation of the responses of the *ParticleFilter* model, as a function of number of particles, N_P.

***Sampling* model.** In the *Sampling* model, instead of using the Bayesian posterior to maximize its expected score, a model subject samples its response from the marginal posterior on the states, $p_t(s|x_{1:t})$. In spite of this suboptimal selection rule, the average learning rate as a function of the run-length has a behavior similar to the optimal one (decreasing in the HI condition, smile-shaped in the HD condition), albeit with higher average values (Fig. 16A). The repetition propensity also behaves similarly to the optimal one, but is suppressed in magnitude due to sampling (Fig. 16B). Finally, as expected by construction, the *Sampling* model leads to behavioral variability, and the amplitude of the latter scales with the width of the posterior distribution (Fig. 16C).

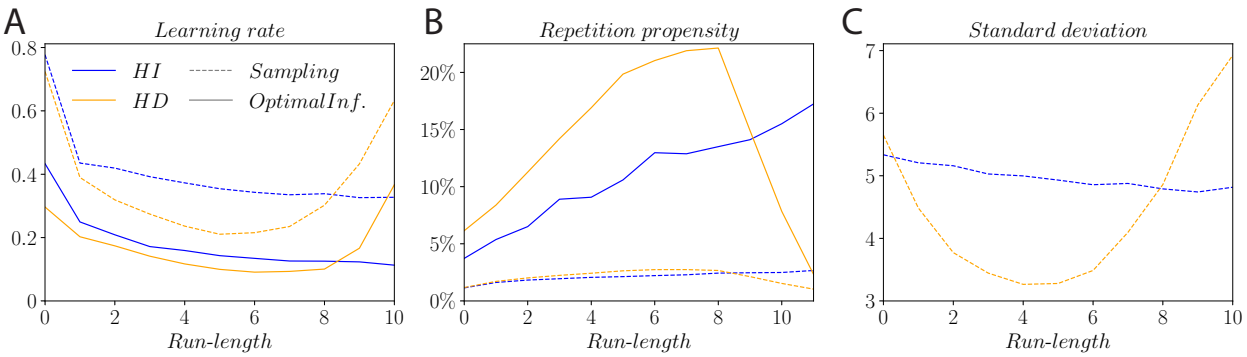

Figure 16. **Behavior of the *Sampling* model, as compared to the *OptimalInference* model. A,B.** Average learning rate (A), and repetition propensity (B), in the *Sampling* model (dashed lines) and the *OptimalInference* model (solid lines), as a function of run-length, in the HI and HD conditions. **C.** Standard deviations of responses, as a function of run-length, of the *Sampling* model in the HI and HD conditions. The *OptimalInference* model exhibits no variability.

Normalized Mean Squared Error. This section provides some details on the Normalized Mean Squared Error we use to compare the results of the various models to the *OptimalInference* model and to human data. Let $y_i(\tau)$ be the value at run-length τ of the quantity of interest i (learning rate, repetition propensity, or standard deviation of responses), as observed in data or as resulting from the optimal model, and $\hat{y}_i(\tau)$ the value resulting from a suboptimal model. The mean squared error is $MSE(\hat{y}_i) = \frac{1}{n}\sum_\tau(\hat{y}_i(\tau) - y_i(\tau))^2$, where n is the number of run-lengths. We want to be

able to compare the errors for different quantities of interest. By dividing the MSE by the variance of y_i, we obtain the Normalized Mean Squared Error, which is translation-invariant and scale-invariant:

$$NMSE_i = \frac{MSE(\hat{y}_i)}{\mathrm{Var}[y_i]} = \frac{\sum_\tau (\hat{y}_i(\tau) - y_i(\tau))^2}{\sum_\tau (\bar{y}_i - y_i(\tau))^2}, \tag{23}$$

where $\bar{y}_i$ is the average $y_i(\tau)$. For model fitting, we then use the average of this quantity over the three quantities of interest ("three-error measure") or over two of them ("two-error measure").

Approximate formulation of the Bayesian Information Criterion. For a given subject in a given condition (HI or HD), we denote the probability of a sequence of T responses, $\hat{s}_{1:T}$, by $p(\hat{s}_{1:T})$. In the models with deterministic inference, the joint probability of responses is the product of the probabilities of each of the responses in the successive trials:

$$p(\hat{s}_{1:T}) = \prod_{t=1}^{T} p(\hat{s}_t). \tag{24}$$

This independence condition does not hold for the models with stochastic inference. We describe, here, how we overcome this issue in the case of the *ParticleFilter* model, which is the most involved model we consider and the most costly computationally. We denote by N the number of particles, and we define the 'internal state' of a particle filter at time t as the joint states of its N particles, each defined by a location, s_t, a run-length, τ_t, and a weight, w_t. We denote the internal state by σ_t, and the sequence of the internal states throughout a run with T trials by $\sigma_{1:T}$. The probability of a subject's sequence of responses is expressed in terms of the probability of the sequences of internal states, as

$$p(\hat{s}_{1:T}) = \sum_{\sigma_{1:T}} p(\hat{s}_{1:T}|\sigma_{1:T}) p(\sigma_{1:T}). \tag{25}$$

We note that conditional on the internal state, the responses are independent:

$$p(\hat{s}_{1:T}|\sigma_{1:T}) = \prod_{t=1}^{T} p(\hat{s}_t|\sigma_t). \tag{26}$$

In other words, a *single* realization of the particle filter can be thought of as a model with deterministic inference. To compute the probability of responses, however, we must determine the distribution of the realizations of the internal states of the particle filters, $p(\sigma_{1:T})$. The support of this distribution is the Cartesian product of the NT internal states of the particle filter, and, hence, its size grows exponentially with NT. Estimating a probability distribution over this space seems computationally intractable; we can, however, carry out an approximate calculation of the probability distribution.

Our method of approximation relies on a Monte-Carlo estimation: we run $M = 500$ simulations of the inference model (*ParticleFilter* or τ*Sample*), and consequently we obtain M points $\sigma_{1:T}$ in the space of possible sequences of internal states. For each realization of the internal state, we know the probability distribution of the response, conditional on the state, $p(\hat{s}_t|\sigma_t)$. A Monte-Carlo approximation of the probability of a sequence of T responses, $\tilde{p}(\hat{s}_{1:T})$, is then obtained as

$$\tilde{p}(\hat{s}_{1:T}) = \frac{1}{M} \sum_{\sigma_{1:T}} p(\hat{s}_{1:T}|\sigma_{1:T}), \tag{27}$$

i.e.,

$$\tilde{p}(\hat{s}_{1:T}) = \frac{1}{M} \sum_{\sigma_{1:T}} \prod_{t=1}^{T} p(\hat{s}_t|\sigma_t), \tag{28}$$

where the sum is taken over the M sampled sequences of internal states. This empirical approximation is satisfactory provided M is sufficiently large, so as to overcome the exponential growth of the number of possible sequences.

For any sampled sequence of internal states, $\sigma_{1:T}$, it is extremely likely that at least one response, $\hat{s}_t$, has vanishingly small probability given the corresponding internal state, σ_t,

i.e., $p(\hat{s}_t|\sigma_t) \approx 0$. In other words, given a sequence of responses, $\hat{s}_{1:T}$, it is extremely unlikely that any one of M sequences of internal states produces $\hat{s}_{1:T}$, i.e., it is prohibitively improbable that any one of M sequences of internal states account *simultaneously* for 1000 responses. Thus, if we carry out the MonteCarlo approximation naively, we underestimate severely the likelihood of the data. (We emphasize that, in that case, the low value of the likelihood does not reflect an inherent inability of the model to account for the data, but rather the poor sampling of the internal states in the model.)

We can circumvent this practical problem by dividing our experimental runs into shorter sequences: while our sample is too small for obtaining a useful approximation of the density of possible sequences of 1000 successive internal states, we can treat this density over *shorter sequences*. In the extreme case, we can consider the (Monte-Carlo-approximated) likelihood of just *one* response, $\hat{s}_t$, at a trial t:

$$\tilde{p}(\hat{s}_t) = \frac{1}{M} \sum_{\sigma_{1:T}} p(\hat{s}_t|\sigma_t). \tag{29}$$

Here, the M samples, $\sigma_{1:T}$, are used to estimate a one-dimensional density, instead of a 1000-dimensional density. The joint likelihood of all responses can then be approximated as the product of the likelihoods of each response, i.e.,

$$\tilde{p}(\hat{s}_{1:T}) \approx \prod_{t=1}^{T} \tilde{p}(\hat{s}_t). \tag{30}$$

This approximation, however, in effect makes the crude assumption that successive responses are independent, and thus neglects the sequential dependence in responses that a model may predict. For instance, this approximation vastly underestimates the likelihood of a model that correctly predicts an appreciable probability of repetitions.

In order to obtain an approximation of the likelihood that takes into account the sequential dependence of responses, and which can be computed on the basis of M samples,

we choose to compute the likelihood of responses over sequences of 10 successive trials,

$$\tilde{p}(\hat{s}_{t:t+9}) = \frac{1}{M} \sum_{\sigma_{1:T}} p(\hat{s}_{t:t+9}|\sigma_{t:t+9}). \tag{31}$$

We fit the models by evaluating how well they reproduce the responses in all the 10-trial-long sequences. More precisely, we associate to each model a BIC calculated as

$$\text{BIC} = -2 \ln \left[\prod_{t=1,11,\ldots} \tilde{p}(\hat{s}_{t:t+9}) \right] + k \ln n, \tag{32}$$

where k is the number of parameters in the model under consideration, and n is the number of data points. The specific choice of sequences with 10 trials is arbitrary; in our analyses, we repeated the calculations for different choices, which yielded comparable results. We chose to illustrate this choice as it corresponds to sequences no shorter than the mean inter-change-point duration, that optimizes the precision of the approximation.

We note that our approximation of the *ParticleFilter* model's BIC could be interpreted as a (possibly still approximate) calculation of the BIC of a different model. Specifically, the latter would include a particular form of a particle-rejuvenation procedure in which all the particles are replaced, every 10 trials, by as many new particles, each randomly sampled from the distribution of possible particles, at these trials. The rejuvenation kernel would thus be independent of the rejuvenated particle (a possibility considered in the particle-filter literature, see Chopin, 2002), and the acceptance rate would be equal to 1, each 10 trials, and to 0 at other trials; thus, it would also be independent of the particle (which is at odds with the usual form of the rejuvenation procedures introduced in the literature).

Rational-inattention models and models with fixed repetition probability. The rational-inattention models we present are inspired by the model introduced by Khaw et al., 2017. We adopt their notation for the new quantities introduced here to describe the response-selection process. The major new ingredient in this model is that the subject,

after having observed a sequence of stimuli, $x_{1:t}$, is assumed to choose, first, whether to adjust or to repeat the current estimate (the 'repetition variable'); the repetition variable is a Bernoulli random variable parameterized by the probability of adjusting, denoted by $\Lambda_t(x_{1:t}, \hat{s}_{t-1})$ (and the probability of a repetition is thus $1 - \Lambda_t(x_{1:t}, \hat{s}_{t-1})$). Second, conditional on adjusting, the subject randomly chooses a new estimate (the 'location variable'), sampled from a distribution which we denote by $\mu_t(\hat{s}_t|x_{1:t})$. Thus, the model subject's distribution of responses conditional on the observed stimuli and on the preceding response is

$$p(\hat{s}_t|x_{1:t}, \hat{s}_{t-1}) = (1 - \Lambda_t(x_{1:t}, \hat{s}_{t-1}))\delta(\hat{s}_t - \hat{s}_{t-1}) + \Lambda_t(x_{1:t}, \hat{s}_{t-1})\mu_t(\hat{s}_t|x_{1:t}), \tag{33}$$

where δ is the Dirac delta function.

Following the rational-inattention approach, we assume that the distribution of responses conditional on the observed stimuli and on the preceding response, $p(\hat{s}_t|x_{1:t}, \hat{s}_{t-1})$, maximizes, under a constraint detailed below, the expected reward. Although the response at trial t, $\hat{s}_t$, affects the rewards in subsequent trials, through the ensuing distributions of responses (Eq. (33)), for the sake of calculations simplicity we will carry out a 'greedy' optimization in the model. Specifically, we will assume that the distribution of responses conditional on the observed stimuli and on the preceding response, $p(\hat{s}_t|x_{1:t}, \hat{s}_{t-1})$, is obtained by considering only the immediate reward in expectation over the state, the response, and the sequence of past stimuli, i.e., the quantity

$$\bar{R} \equiv \int \cdots \int p(x_{1:t}) \int p(\hat{s}_t|x_{1:t}, \hat{s}_{t-1}) \int p_t(s|x_{1:t}) R(\hat{s}, s) \mathrm{d}s \mathrm{d}\hat{s}_t \mathrm{d}x_1 \ldots \mathrm{d}x_t, \tag{34}$$

where $R(\hat{s}, s)$ is the reward obtained if the estimate is $\hat{s}$ and the correct state is s.

In the absence of a constraint, the optimal response is obtained by maximizing the expected reward implied by the sequence of past stimuli, $\int p_t(s|x_{1:t}) R(\hat{s}, s) \mathrm{d}s$, which we denote by $r(\hat{s}_t|x_{1:t})$. We assume, however, that it is costly for the subject to choose with

precision the repetition variable and the location variable, and this hampers the ability to obtain this optimal estimate. Following Khaw et al., 2017, we assume that the repetition variable (distributed according to $1 - \Lambda_t(x_{1:t}, \hat{s}_{t-1})$), and the location variable in trials in which the estimate is not repeated (distributed according to $\mu_t(\hat{s}_t|x_{1:t})$), each bear a cognitive cost proportional to measures of the amount of information on the sequence of stimuli involved in choosing the repetition variable and the location variable, respectively, defined as

$$I_1 = \int \cdots \int p(x_{1:t}) D_{KL}(\Lambda_t(x_{1:t}, \hat{s}_{t-1}) || \tilde{\Lambda}) \mathrm{d}x_1 \ldots \mathrm{d}x_t \tag{35}$$

$$\text{and } I_2 = \int \cdots \int p(x_{1:t}) \Lambda_t(x_{1:t}, \hat{s}_{t-1}) D_{KL}(\mu_t(.|x_{1:t}) || \tilde{\mu}) \mathrm{d}x_1 \ldots \mathrm{d}x_t, \tag{36}$$

where $\tilde{\Lambda}$ and $\tilde{\mu}$ are the unconditional (not conditional on $x_{1:t}$) probability of adjusting and the distribution of estimates, respectively. The distributions $\Lambda_t(x_{1:t}, \hat{s}_{t-1})$ and $\mu_t(\hat{s}_t|x_{1:t})$ are obtained as those that maximize the quantity

$$\bar{R} - \psi_1 I_1 - \psi_2 I_2 \tag{37}$$

which expresses a trade-off between expected reward and cognitive costs, where ψ_1 and ψ_2 are numerical coefficients specifying the strength of the information-theoretic costs.

The solution to the optimization problem just posed is given by

$$\mu_t(\hat{s}_t|x_{1:t}) = \frac{1}{Z(x_{1:t})} \tilde{\mu}(\hat{s}_t) \exp(\psi_2^{-1} r(\hat{s}_t|x_{1:t})), \tag{38}$$

and

$$\ln \frac{\Lambda_t(x_{1:t}, \hat{s}_{t-1})}{1 - \Lambda_t(x_{1:t}, \hat{s}_{t-1})} = \ln \frac{\tilde{\Lambda}}{1 - \tilde{\Lambda}} + \psi_1^{-1}(\psi_2 \ln Z(x_{1:t}) - r(\hat{s}_{t-1}|x_{1:t})), \tag{39}$$

where

$$Z(x_{1:t}) = \int \tilde{\mu}(\hat{s}) \exp(\psi_2^{-1} r(\hat{s}|x_{1:t})) \mathrm{d}\hat{s}. \tag{40}$$

The unconditional distribution of estimates, $\tilde{\mu}(\hat{s})$, can be approximated by a uniform distribution on the space of responses; for the sake of simplicity, we use this approximation in our calculations.

We implement four variants of this model and compute their BICs (Table 2). In a first variant, no cost weighs on the repetition variable ($\psi_1 = 0$), but a cognitive cost prevents the model subject from choosing the optimal response ($\psi_2 \neq 0$, Table 2, second row). By contrast, in the second variant of the model the repetition variable is subject to a cost ($\psi_1 \neq 0$), while the location variable is not ($\psi_2 = 0$); the latter follows, however, a *NoisyMax* response-selection strategy (Table 2, third row). (If the location variable, instead, were optimal, the model would assign a vanishing probability to most of the subjects' responses, and consequently would yield an infinite BIC.) In a third variant of the model, there are attentional costs weighing on both the repetition and the location variable ($\psi_1 \neq 0$ and $\psi_2 \neq 0$, Table 2, fifth row). In the fourth variant of the model, the location variable is subject to a cognitive cost ($\psi_2 \neq 0$); the repetition variable, however, is not derived in a rational-inattention approach, but instead is random and governed by a fixed repetition probability (the first variant, above, is a special case of this fourth variant of the model, corresponding to a repetition probability set to zero; Table 2, sixth row). For the sake of comparison, we implement a fifth model that combines a fixed repetition probability with a *NoisyMax* strategy for the location variable (this model does not feature any cognitive cost; Table 2, fourth row).

The five models just presented make use of the *OptimalInference* strategy. We implement, in addition, five other models in which the repetition variable and the location variable are chosen as in these five models, but the *OptimalInference* strategy is replaced by the *ParticleFilter* inference strategy (Table 2, last column).

Repetition variable	**Location variable**	**BIC**	
		OptimalInf.	***ParticleFilter***
-	*NoisyMax*	495,063	482,358
-	Rational Inattention	524,719	486,217
Rational Inattention	*NoisyMax*	445,189	439,195
Fixed repetition probability	*NoisyMax*	445,354	437,797
Rational Inattention	Rational Inattention	471,995	446,187
Fixed repetition probability	Rational Inattention	472,810	446,848

Table 2. **BICs of the rational-inattention models and the models with fixed repetition probability, combined with the *OptimalInference* and the *ParticleFilter* inference strategies.**

Acknowledgements

This work was supported by the CNRS through UMR8550, the Global Scholar Program at Princeton University, and the Visiting Faculty Program at the Weizmann Institute of Science. A.P.C. was supported by a PhD fellowship of the Fondation Pierre-Gilles de Gennes pour la Recherche. This work was granted access to the HPC resources of MesoPSL financed by the Region Ile de France and the project Equip@Meso (reference ANR-10-EQPX-29-01) of the program Investissements d'Avenir supervised by the Agence Nationale pour la Recherche.

References

Acerbi, L., Vijayakumar, S., & Wolpert, D. M. (2014). On the Origins of Suboptimality in Human Probabilistic Inference - S2 - Noisy Probabilistic Inference. *PLoS Computational Biology*, *10*(6), 1–5.

Adams, R. P., & MacKay, D. J. C. (2007). Bayesian Online Changepoint Detection. *ArXiv e-prints*, 1–7. arXiv: 0710.3742. Retrieved from http://arxiv.org/abs/0710.3742

Afrouzi, H., Kwon, S. Y., & Ma, Y. (2020). *A Model of Costly Recall*.

Ahn, S., & Fessler, J. (2003). Standard Errors of Mean, Variance, and Standard Deviation Estimators. *EECS Department, University of Michigan*, 1–2.

Aitchison, L., & Lengyel, M. (2016). The Hamiltonian Brain: Efficient Probabilistic Inference with Excitatory-Inhibitory Neural Circuit Dynamics. *PLoS Computational Biology*, *12*(12), 1–24. doi:10.1371/journal.pcbi.1005186. arXiv: 1407.0973

Anteneodo, C., & Chialvo, D. R. (2009). Unraveling the fluctuations of animal motor activity. *Chaos*, *19*, 1–7. doi:10.1063/1.3211189

Arulampalam, M., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, *50*(2), 174–188. doi:10.1109/78.978374

Azeredo da Silveira, R., & Woodford, M. (2019). Noisy Memory and Over-Reaction to News. *AEA Papers and Proceedings*, *109*, 557–561. doi:10.1257/pandp.20191049

Battaglia, P. W., Jacobs, R. a., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, *20*(7), 1391. doi:10.1364/JOSAA.20.001391

Battaglia, P. W., Kersten, D., & Schrater, P. R. (2011). How haptic size sensations improve distance perception. *PLoS Computational Biology*, *7*(6), e1002080. doi:10.1371/journal.pcbi.1002080

Berniker, M., & Kording, K. (2011). Bayesian approaches to sensory integration for motor control. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(4), 419–428. doi:10.1002/wcs.125

Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science*, *311*(5763), 1020–1022. doi:10.1126/science.1121872. arXiv: arXiv:1011.1669v3

Bramley, N. R., Dayan, P., Griffiths, T. L., & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*, *124*(3), 301–338. doi:10.1037/rev0000061

Brown, S. D., & Steyvers, M. (2009). Detecting and predicting changes. *Cognitive Psychology*, *58*(1), 49–67. doi:10.1016/j.cogpsych.2008.09.002

Buesing, L., Bill, J., Nessler, B., & Maass, W. (2011). Neural dynamics as sampling: A model for stochastic computation in recurrent networks of spiking neurons. *PLoS Computational Biology*, *7*(11), e1002211. doi:10.1371/journal.pcbi.1002211

Campione, E., & Véronis, J. (2002). A Large-Scale Multilingual Study of Silent Pause Duration. *Speech Prosody 2002*, 199–202. doi:10.1.1.12.561

Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika*, *89*(3), 539–552. doi:10.1093/biomet/89.3.539

Cox, R. T. (1946). Probability, Frequency and Reasonable Expectation. *American Journal of Physics*, *14*(1), 1. doi:10.1119/1.1990764. arXiv: arXiv:1011.1669v3

Daw, N. D., & Courville, A. C. (2008). The pigeon as particle filter. *Advances in neural information processing systems*, 369–376. doi:10.1.1.73.3246

Denison, S., Bonawitz, E., Gopnik, A., & Griffiths, T. L. (2013). Rational variability in children's causal inferences: The Sampling Hypothesis. *Cognition*, *126*(2), 280–300. doi:10.1016/j.cognition.2012.10.010

Doucet, A., & Johansen, A. M. (2008). A Tutorial on Particle Filtering and Smoothing: Fifteen years later. In D. Crisan & B. Rozovsky (Eds.), *The oxford handbook of nonlinear filtering* (pp. 1–39). Oxford University Press.

Drugowitsch, J., Wyart, V., Devauchelle, A.-D., & Koechlin, E. (2016). Computational Precision of Mental Inference as Critical Source of Human Choice Suboptimality. *Neuron*, *92*(6), 1398–1411. doi:10.1016/j.neuron.2016.11.005

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. doi:10.1038/415429a

Fearnhead, P., & Liu, Z. (2007). On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*(4), 589–605. doi:10.1111/j.1467-9868.2007.00601.x

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences*, *14*(3), 119–130. doi:10.1016/j.tics.2010.01.003. arXiv: NIHMS150003

Gallistel, C. R., Krishan, M., Liu, Y., Miller, R., & Latham, P. E. (2014). The perception of probability. *Psychological Review*, *121*(1), 96–123. doi:10.1037/a0035232

Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition*, *173*(December 2017), 34–42. doi:10.1016/j.cognition.2017.12.014

Gershman, S. J., & Beck, J. M. (2016). Complex Probabilistic Inference : From Cognition to Neural Computation. In *Computational models of brain and behavior* (pp. 1–17). doi:10.1002/9781119159193.ch33

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278. doi:10.1126/science.aac6076. arXiv: arXiv:1011.1669v3

Gershman, S. J., Vul, E., & Tenenbaum, J. B. (2012). Multistability and Perceptual Inference. *Neural Computation*, *24*(1), 1–24.

Ghose, G. M., & Maunsell, J. H. R. (2002). Attentional modulation in visual cortex depends on task timing. *Nature*, *419*(6907), 616–620. doi:10.1038/nature01057

Glaze, C. M., Filipowicz, A. L., Kable, J. W., Balasubramanian, V., & Gold, J. I. (2018). A bias-variance trade-off governs individual differences in on-line learning in an unpredictable environment. *Nature Human Behaviour*, *2*(3), 213–224. doi:10.1038/s41562-018-0297-4

Glaze, C. M., Kable, J. W., & Gold, J. I. (2015). Normative evidence accumulation in unpredictable environments. *eLife*, *4*(AUGUST2015), 1–27. doi:10.7554/eLife.08825

Goodman, N., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A Rational Analysis of Rule-Based Concept Learning. *Cognitive Science: A Multidisciplinary Journal*, *32*(1), 108–154. doi:10.1080/03640210701802071

Gordon, N., Salmond, D., & Smith, A. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F Radar and Signal Processing*, *140*(2), 107. doi:10.1049/ip-f-2.1993.0015

Griffin, L., West, D. J., & West, B. J. (2000). Random stride intervals with memory. *Journal of Biological Physics*, *26*(3), 185–202. doi:10.1023/A:1010322406831

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal Predictions in Everyday Cognition. *Psychological Science*, *17*(9), 767–773. doi:10.1111/j.1467-9280.2006.01780.x

Griffiths, T. L., & Tenenbaum, J. B. (2011). Predicting the future as Bayesian inference: People combine prior knowledge with observations when estimating duration and extent. *Journal of Experimental Psychology: General*, *140*(4), 725–743. doi:10.1037/a0024899

Hausdorff, J. M., Peng, C. K., Ladin, Z., Wei, J. Y., & Goldberger, a. L. (1995). Is walking a random walk? Evidence for long-range correlations in stride interval of human gait. *Journal of applied physiology (Bethesda, Md. : 1985)*, *78*(1), 349–358.

Hennequin, G., Aitchison, L., & Lengyel, M. (2014). Fast Sampling-Based Inference in Balanced Neuronal Networks. *Advances in Neural Information Processing Systems*, 1–9.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, *4*(3), 267–272. doi:10.1901/jeab.1961.4-267

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, *4*(12), 1. doi:10.1167/4.12.1

Holm, S. (1978). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70. Retrieved from https://www.jstor.org/stable/4615733

Hoyer, P. O., & Hyvärinen, A. (2003). Interpreting neural response variability as monte carlo sampling of the posterior. *Advances in neural information processing systems*, (1), 293–300. doi:10.1.1.71.1731

Huys, Q. J. M., Eshel, N., Nions, E. O., Sheridan, L., Dayan, P., & Jonathan, P. (2012). Bonsai Trees in Your Head : How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees. *PLOS Computational Biology*, *8*(3). doi:10.1371/journal.pcbi.1002410

Jacobs, R. a. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*(21), 3621–3629. doi:10.1016/S0042-6989(99)00088-7

Janssen, P., & Shadlen, M. N. (2005). A representation of the hazard rate of elapsed time in macaque area LIP. *Nature Neuroscience*, *8*(2), 234–241. doi:10.1038/nn1386

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*(8), 1020–1026. doi:10.1038/nn.2590

Kaufmann, E., Korda, N., & Munos, R. (2012). Thompson Sampling: An Asymptotically Optimal Finite Time Analysis. *Algorithmic Learning Theory*, (1), 15. arXiv: 1205.4217. Retrieved from http://arxiv.org/abs/1205.4217

Khaw, M. W., Stevens, L., & Woodford, M. (2017). Discrete adjustment to a changing environment : Experimental evidence. *Journal of Monetary Economics*, *91*, 88–103. doi:10.1016/j.jmoneco.2017.09.001

Knill, D. C. (2003). Mixture models and the probabilistic structure of depth cues. *Vision Research*, *43*(7), 831–854. doi:10.1016/S0042-6989(03)00003-8

Knill, D. C. (2007). Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision*, *7*(7), 5. doi:10.1167/7.7.5

Koehler, D. J., & James, G. (2009). Probability matching in choice under uncertainty: Intuition versus deliberation. *Cognition*, *113*(1), 123–127. doi:10.1016/j.cognition.2009.07.003

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247. doi:10.1038/nature02169

Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, *10*(7), 319–326. doi:10.1016/j.tics.2006.05.003

Levy, R., Reali, F., & Griffiths, T. L. (2008). Modeling the effects of memory on human online sentence processing with particle filters. *Proceedings of NIPS 2008*, 1–8. doi:10.1.1.211.8877

Li, Y., & Dudman, J. T. (2013). Mice infer probabilistic models for timing. *Proceedings of the National Academy of Sciences*, *110*(42), 17154–17159. doi:10.1073/pnas.1310666110

Lloyd, K., Sanborn, A., Leslie, D., & Lewandowsky, S. (2019). Why Higher Working Memory Capacity May Help You Learn: Sampling, Search, and Degrees of Approximation. *Cognitive Science*, *43*(12). doi:10.1111/cogs.12805

Low, E., Grabe, E., & Nolan, F. (2000). Quantitative characterisations of speech rhythm. *Language and Speech*, *46*, 377–401.

Miyazaki, M., Nozaki, D., & Nakajima, Y. (2005). Testing Bayesian models of human coincidence timing. *Journal of neurophysiology*, *94*(1), 395–399. doi:10.1152/jn.01168.2004

Morasso, P. (1981). Spatial control of arm movements. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *42*(2), 223–7. doi:10.1007/BF00236911

Moreno-Bote, R., Knill, D. C., & Pouget, A. (2011). Bayesian sampling in visual perception. *Proceedings of the National Academy of Sciences*, *108*(30), 12491–12496. doi:10.1073/pnas.1101430108

Nakamura, T., Kiyono, K., Yoshiuchi, K., Nakahara, R., Struzik, Z. R., & Yamamoto, Y. (2007). Universal Scaling Law in Human Behavioral Organization. *Physical Review Letters*, *138103*(September), 1–4. doi:10.1103/PhysRevLett.99.138103

Nakamura, T., Takumi, T., Takano, A., Aoyagi, N., Yoshiuchi, K., Zbigniew, R., & Yamamoto, Y. (2008). Of Mice and Men — Universality and Breakdown of Behavioral Organization. *PLoS ONE*, *3*(4), 1–8. doi:10.1371/journal.pone.0002050

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040–1046. Retrieved from http://www.nature.com/doifinder/10.1038/nn.3130

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An Approximately Bayesian Delta-Rule Model Explains the Dynamics of Belief Updating in a Changing Environment. *Journal of Neuroscience*, *30*(37), 12366–12378. Retrieved from http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.0822-10.2010

Neligh, N. (2019). *Rational Memory with Decay*.

Nunes Amaral, L. A., Bezerra Soares, D. J., da Silva, L. R., Lucena, L. S., Saito, M., Kumano, H., ... Yamamoto, Y. (2004). Power law temporal auto-correlations in day-long records of human physical activity. *Europhys. Lett. 66*(3), 448–454. doi:10.1209/epl/i2003-10227-7

Pearl, L., Goldwater, S., & Steyvers, M. (2011). Online Learning Mechanisms for Bayesian Models of Word Segmentation. *Research on Language and Computation*, *8*(2-3), 1–26. doi:10.1007/s11168-011-9074-5

Peirce, J. W. (2009). Generating stimuli for neuroscience using psychopy. *Frontiers in Neuroinformatics*, *2*(10). doi:10.3389/neuro.11.010.2008

Piet, A. T., El Hady, A., & Brody, C. D. (2018). Rats adopt the optimal timescale for evidence integration in a dynamic environment. *Nature Communications*, *9*(1), 1–12. doi:10.1038/s41467-018-06561-y

Radillo, A. E., Veliz-Cuba, A., Josić, K., & Kilpatrick, Z. P. (2017). Evidence Accumulation and Change Rate Inference in Dynamic Environments. *Neural Computation*, *29*(6), 1561–1610. doi:10.1162/NECO_a_00957

Radillo, A. E., Veliz-Cuba, A., Josić, K., & Kilpatrick, Z. P. (2019). Performance of normative and approximate evidence accumulation on the dynamic clicks task. (516088), 1–29. arXiv: 1902.01535. Retrieved from http://arxiv.org/abs/1902.01535

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292. doi:10.1016/S0010-0277(99)00058-X

Rigoux, L., & Guigon, E. (2012). A Model of Reward- and Effort-Based Optimal Decision Making and Motor Control. *PLoS Computational Biology*, *8*(10). doi:10.1371/journal.pcbi.1002716

Sanborn, A. N. (2015). Types of approximation for probabilistic cognition: Sampling and variational. *Brain and Cognition*, 8–11. doi:10.1016/j.bandc.2015.06.008

Sanborn, A. N., & Chater, N. (2016). Bayesian Brains without Probabilities. *Trends in Cognitive Sciences*, *20*, 1–11. doi:10.1016/j.tics.2016.10.003

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2006). A More Rational Model of Categorization. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 1–6.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117*(4), 1144–1167. doi:10.1037/a0020511

Savin, C., & Denève, S. (2014). Spatio-temporal representations of uncertainty in spiking neural networks. *Advances in Neural Information Processing Systems*, 1–9.

Schulz, E., Konstantinidis, E., & Speekenbrink, M. (2015). Learning and decisions in contextual multi-armed bandit tasks. *Proceedings of the 37th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society*, *1*, 2122–2127. Retrieved from https://ericschulz.github.io/publications/Schulz2015learning.pdf

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. doi:10.1214/aos/1176344136

Shadmehr, R. (2009). Computational Approaches to Motor Control. In *Encyclopedia of neuroscience* (Vol. 3, pp. 9–17).

Shi, L., & Griffiths, T. L. (2009). Neural implementation of hierarchical Bayesian inference by importance sampling. *Advances in Neural Information Processing Systems 22*, 1669–1677. Retrieved from https://papers.nips.cc/paper/3782-neural-implementation-of-hierarchical-bayesian-inference-by-importance-sampling.pdf

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665–690. doi:10.1016/S0304-3932(03)00029-1

Sims, C. A. (2011). Rational Inattention and Monetary Economics - Handbook of Monetary Economics. *Handbook of Monetary Economics*, *3*(2007), 155–181. doi:10.1016/B978-0-444-53238-1.00004-1

Speekenbrink, M., & Konstantinidis, E. (2015). Uncertainty and exploration in a restless bandit problem. *Topics in Cognitive Science*, *7*(2), 351–367. doi:10.1111/tops.12145

Stocker, A. A., & Simoncelli, E. P. (2006). Sensory adaptation within a Bayesian framework for perception. *Advances in Neural Information Processing Systems*, *18*, 1291–1298. doi:10.1016/j.cub.2008.10.014

Stocker, A. A., & Simoncelli, E. P. (2008). A Bayesian Model of Conditioned Perception. *Advances in neural information processing systems*, *20*(May), 1409–1416.

Sundareswara, R., & Schrater, P. R. (2008). Perceptual multistability predicted by search model for Bayesian decisions. *Journal of vision*, *8*(5), 12.1–19. doi:10.1167/8.5.12

ten Oever, S., Schroeder, C. E., Poeppel, D., van Atteveldt, N., & Zion-Golumbic, E. (2014). Rhythmicity and cross-modal temporal cues facilitate detection. *Neuropsychologia*, *63*, 43–50. doi:10.1016/j.neuropsychologia.2014.08.008

Thaker, P., Tenenbaum, J. B., & Gershman, S. J. (2017). Online learning of symbolic concepts. *Journal of Mathematical Psychology*, *77*, 10–20. doi:10.1016/j.jmp.2017.01.002

Thompson, W. R. (1933). On the Likelihood That One Unknown Probability Exceeds Another In View Of The Evidence Of Two Samples. *Biometrika*, *25*(3), 285–294.

van Beers, R. J., Sittig, A. C., & Denier, J. A. N. J. (1999). Integration of Proprioceptive and Visual Position-Information: An Experimentally Supported Model. *Integration The Vlsi Journal*, *81*(3), 1355–1364.

Veliz-Cuba, A., Kilpatrick, Z. P., & Josić, K. (2016). Stochastic models of evidence accumulation in changing environments. *SIAM Review*, *58*(2), 264–289. doi:10.1137/15M1028443

Vul, E. (2011). Sampling in human cognition - Chapter 1. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *72*(4-B), 2471. Retrieved from https://search.proquest.com/docview/860001857

Vul, E., Frank, M. C., Tenenbaum, J. B., & Alvarez, G. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. *Advances in Neural Information Processing Systems*, *22*, 1–9.

Retrieved from http://books.nips.cc/papers/files/nips22/NIPS2009%7B%5C_%7D0980.pdf

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and Done? Optimal Decisions From Very Few Samples. *Cognitive Science*, *38*(4), 599–637. doi:10.1111/cogs.12101

Wilson, R. C., Nassar, M. R., & Gold, J. I. (2010). Bayesian Online Learning of the Hazard Rate In Change Point Problems. *Neural Computation*, 1–25.

Wilson, R. C., Nassar, M. R., & Gold, J. I. (2013). A Mixture of Delta-Rules Approximation to Bayesian Inference in Change-Point Problems. *PLoS Computational Biology*, *9*(7), e1003150. Retrieved from http://dx.plos.org/10.1371/journal.pcbi.1003150.s003

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, *1*(6), 1–5. doi:10.1007/s13398-014-0173-7.2. arXiv: arXiv:1011.1669v3

Woodford, M. (2009). Information-constrained state-dependent pricing. *Journal of Monetary Economics*, *56*(SUPPL.), S100–S124. doi:10.1016/j.jmoneco.2009.06.014

Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability Matching as a Computational Strategy Used in Perception. *PLoS Computational Biology*, *6*(8), e1000871. doi:10.1371/journal.pcbi.1000871

Yu, A. J., & Huang, H. (2014). Maximizing Masquerading as Matching in Human Visual Search Choice Behavior. *Decision*, *1*(4), 275–287. doi:http://dx.doi.org/10.1037/dec0000013

Appendix A

Supplementary tables: statistical tests

$\tilde{\tau}$	0	1	2	3	4	5	6	7	8	9	10
W	.307	.448	.265	.335	.180	.072*	.016**	.073*	.426	.160	.008***
MWU	.269	.391	.281	.212	.207	.066*	.034**	.022**	.491	.568	.006***
S	.307	.448	.268	.337	.178	.070*	.015**	.074*	.426	.158	.008***
N(HI)	454	521	613	508	461	436	411	360	318	316	240
N(HD)	603	616	707	632	637	521	499	421	387	300	231
Excluding repetitions											
W	.353	.421	.329	.352	.142	.042**	.008***	.151	.391	.202	.034**
MWU	.357	.431	.484	.170	.073*	.025**	.007***	.058*	.343	.590	.050**
S	.352	.421	.332	.353	.140	.041**	.007***	.153	.391	.201	.035**
N(HI)	412	435	495	402	353	333	307	288	237	231	174
N(HD)	553	529	582	497	493	405	379	318	291	226	186

Table A1. ***p-values for the one-sided statistical tests of equality of the means of the learning rates in the HI and HD conditions, for each run-length (Fig. 2C).*** W: Welch's test. MWU: Mann-Whitney's U test. S: Student's test. The N(HI) and N(HD) lines report the number of observations. The second half of the table reports the same quantities when excluding all occurrences of repetitions (Suppl. Fig. B2B).

Condition 1	Condition 2	Avg_1	Avg_2	W	MWU	S	N_1	N_2
HI, $\tilde{\tau} \in [5, 6]$	HI, $\tilde{\tau} \in [9, 10]$	0.255	0.208	.0099***	.0476**	.0119**	847	556
HI, $\tilde{\tau} \in [5, 6]$	HD, $\tilde{\tau} \in [5, 6]$	0.255	0.209	.0055***	.0089***	.0052***	847	1020
HI, $\tilde{\tau} \in [9, 10]$	HD, $\tilde{\tau} \in [9, 10]$	0.208	0.264	.0103**	.0602*	.0100***	556	531
HD, $\tilde{\tau} \in [5, 6]$	HD, $\tilde{\tau} \in [9, 10]$	0.209	0.264	.0072***	.0173**	.0050***	1020	531
Excluding repetitions								
HI, $\tilde{\tau} \in [5, 6]$	HI, $\tilde{\tau} \in [9, 10]$	0.338	0.285	.0203**	.0780*	.0233**	640	405
HI, $\tilde{\tau} \in [5, 6]$	HD, $\tilde{\tau} \in [5, 6]$	0.338	0.272	.0017***	.0009***	.0016***	640	784
HI, $\tilde{\tau} \in [9, 10]$	HD, $\tilde{\tau} \in [9, 10]$	0.285	0.340	.0342**	.1721	.0345**	405	412
HD, $\tilde{\tau} \in [5, 6]$	HD, $\tilde{\tau} \in [9, 10]$	0.272	0.340	.0066***	.0095***	.0047***	784	412

Table A2. ***Learning rates averages under various HI/HD and short/long run-length conditions, and p-values for one-sided statistical tests of equality of the means (Fig. 2B).*** The first two columns indicate the HI/HD and short/long run-length conditions. Learning rates at trials verifying these two conditions have their averages reported in the Avg_1 and Avg_2 columns. Columns W, MWU, and S provide the p-values for the tests of equality of the means between the two conditions. W: Welch's test. MWU: Mann-Whitney's U test. S: Student's test. N_1 and N_2 report the number of observations for each condition. The second half of the table reports the same quantities when excluding all occurrences of repetitions (Suppl. Fig. B2A).

$\tilde{\tau}$	0	1	2	3	4	5	6	7	8	9	10
p	.0266**	.0331**	.0123**	.0550*	.0048***	.0000***	.0387**	.213	.445	.0331**	.0000***
N(HI)	1255	1425	1611	1440	1347	1268	1111	975	883	795	722
N(HD)	1591	1674	1943	1804	1650	1493	1333	1168	983	792	584
Excluding repetitions											
p	.0243**	.0293**	.0556*	.4847	.0094***	.0000***	.4853	.272	.953	.0560*	.0004***
N(HI)	1123	1147	1291	1127	1003	963	818	748	643	583	533
N(HD)	1455	1405	1559	1390	1255	1115	978	890	723	614	458

Table A3. ***p-values for Levene's statistical test of equality of the variances of the corrections in the HI and HD conditions, for each run-length (Fig. 4C).*** The N(HI) and N(HD) lines report the number of observations. The second part of the table reports the same quantities when excluding all occurrences of repetitions (Suppl. Fig. B2).

$\tilde{\tau}$	0	1	2	3	4	5	6	7	8	9	10	11
p	0.561	0.114	0.843	0.245	0.420	0.333	0.636	1.000	0.848	0.194	0.091*	0.005***
N(HI)	2635	2211	1984	1786	1620	1471	1351	1215	1098	978	875	791
N(HD)	2647	2339	2135	1943	1804	1661	1534	1404	1240	1111	946	781

Table A4. ***p-values for Fisher's exact test of independence of the repetition propensities in the HI and HD conditions, at each run-length (Fig. 3B).*** The N(HI) and N(HD) lines report the number of observations.

Appendix B

Supplementary figures

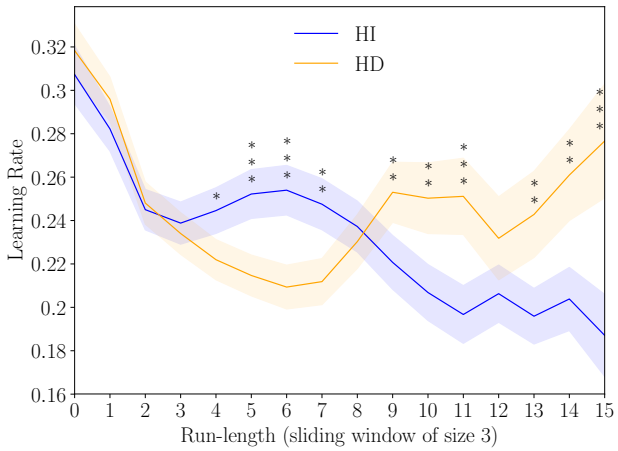

Figure B1. **Learning rates at run-lengths greater than 10.** In order to curb the fluctuations due to the decreasing amount of data at high run-lengths, we use a sliding window of size 3 over the run-lengths (i.e., we pool together the learning rates corresponding to three consecutive run-lengths). The number reported on the x-axis is the center of this sliding window. In the HD condition, the learning rate, after run-length 10, remains large and appears to increase. In the HI condition, the learning rate keeps decreasing after run-length 10. Although presumably the learning rate in this condition should eventually plateau and remain constant, it does not seem that our subjects reach this stage over the run-lengths for which we have sufficient data to allow for analysis (up to 15).

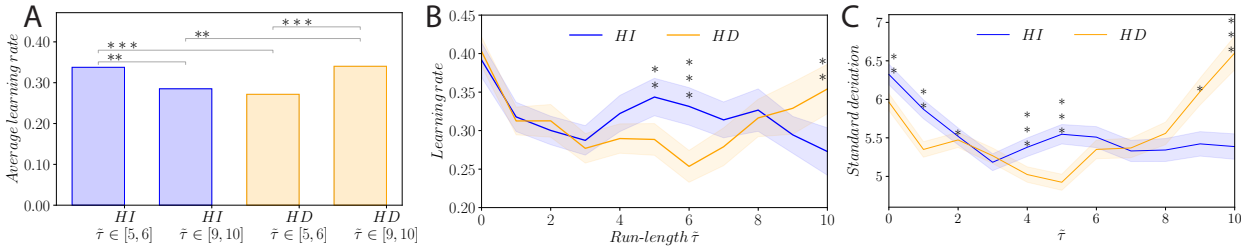

Figure B2. **Human learning rates and standard deviations of responses, excluding all occurrences of repetitions.** (**A,B**) as in Figs. 2B,C; (**C**) as in Fig. 4C.

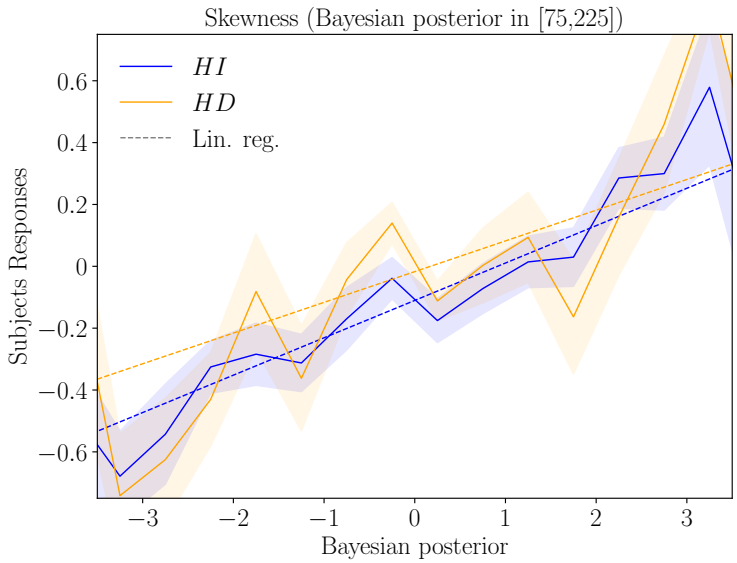

Figure B3. **Skewness in subjects' responses away from the bounds of the response range.** Same analysis of the skewness as in Fig. 6B, but restricted to the trials in which the support of the Bayesian posterior is entirely contained in the middle interval of width half that of the state space (i.e., [75,225], to be compared to the state space, [0,300]). In both conditions, the correlation between the skewness of the Bayesian posterior and the empirical skewness of the subjects' responses is positive and significant (HI: Pearson's r = 0.23, p = 2e-12; HD: r = 0.14, p = 4e-4).

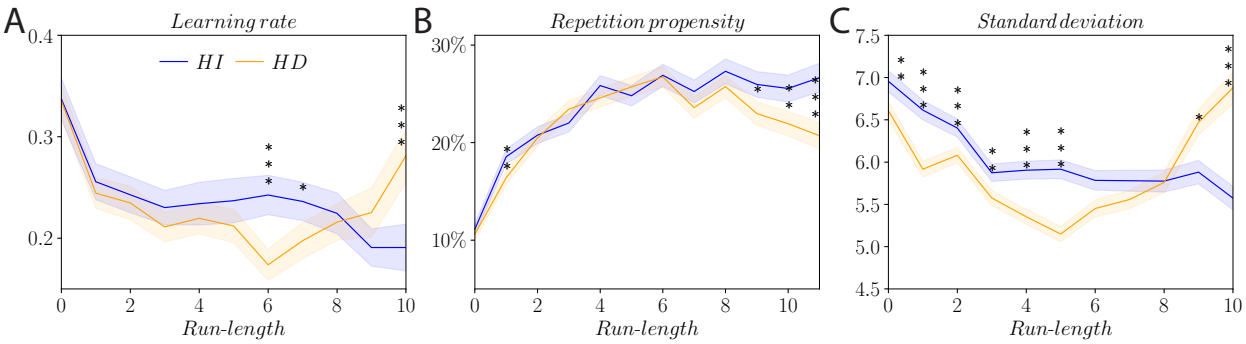

Figure B4. **Human learning rates, repetition propensities, and standard deviations of responses, from the complete contingent of subjects including the four subjects excluded from the analysis presented in the main text (see Methods).** (**A**) as in Fig. 2C; (**B**) as in Fig. 3B; (**C**) as in Fig. 4C.

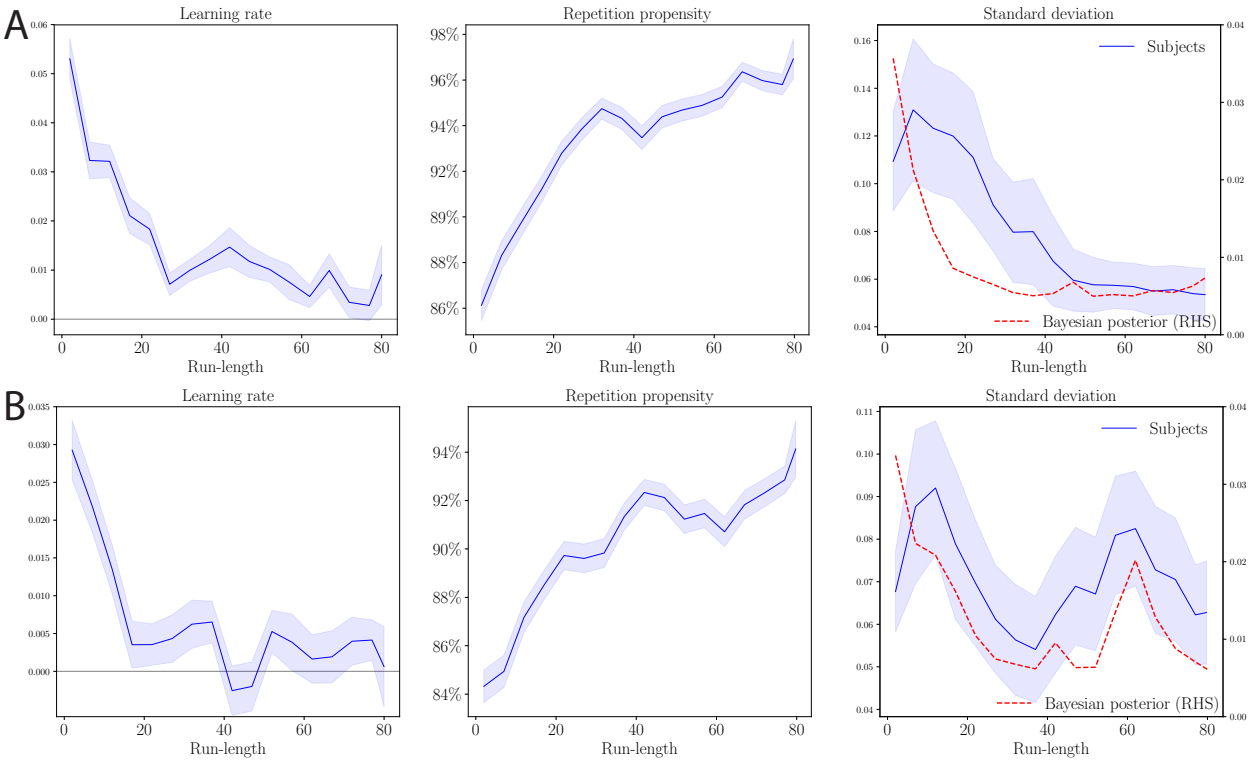

Figure B5. **Behavior of human subjects in the online inference tasks conducted by Gallistel et al., 2014 (A), and Khaw et al., 2017 (B).** *Left panels*: Learning rates as a function of the run-lengths, in trials in which the surprise is greater than .25. *Middle panels*: Repetition propensities. *Right panels*: Standard deviations of responses, averaged across subjects. In these studies, the stimulus is binary and Bernoulli-distributed. The task is to infer the parameter of the Bernoulli distribution, which is subject to change points with a constant probability of 0.5% (to be compared to 10% in the HI condition of our task.) Moreover, some subjects were presented several times with the same sequence of stimuli (in different sessions). It is thus possible to examine the variability of responses within subjects (right panels, across-subject mean of the within-subject standard-deviations). In all panels, in order to mitigate noise, we pool together the responses in windows of five consecutive run-lengths.