Two determinants of dynamic adaptive learning for magnitudes and probabilities

Cedric Foucault[1,2,*] and Florent Meyniel[1,*]

1. Cognitive Neuroimaging Unit, NeuroSpin (INSERM-CEA), University of Paris-Saclay, 91191 Gif-sur-Yvette, France

2. Sorbonne University, Doctoral College, F-75005 Paris, France.

* Corresponding authors: cedric.foucault@gmail.com and florent.meyniel@cea.fr

Classification

Biological Sciences/Psychological and Cognitive Sciences, Social Sciences/Psychological and Cognitive Sciences

Keywords

Learning; probability; uncertainty; dynamic environment; continuous behavior.

Abstract

Humans face a dynamic world that requires them to constantly update their knowledge. Each observed outcome should influence their knowledge to a varying degree depending on whether it arises from a stochastic fluctuation or an environmental change. Thus, humans should dynamically adapt their learning rate based on each outcome. Although crucial for characterizing the learning process, these dynamic adjustments have only been investigated empirically in magnitude learning. Another important type of learning is probability learning. The latter differs from the former in that individual outcomes are much less informative and a single one is insufficient to distinguish environmental changes from stochasticity. Do humans dynamically adapt their learning rate for probabilities? What determinants drive their dynamic adjustments in magnitude and probability learning? To answer these questions, we measured the subjects' learning rate dynamics directly through real-time continuous reports during magnitude and probability learning. We found that subjects dynamically adapt their learning rate in both types of learning. After a change point, they increase their learning rate suddenly for magnitudes and prolongedly for probabilities. Their dynamics are driven differentially by two determinants: change-point probability, the main determinant for magnitudes, and prior uncertainty, the main determinant for probabilities. These results are fully in line with normative theory, both qualitatively and quantitatively. Overall, our findings demonstrate a remarkable human ability for dynamic adaptive learning under uncertainty, and guide studies of the neural mechanisms of learning, highlighting different determinants for magnitudes and probabilities.

Significance statement

In a dynamic world, we must constantly update our knowledge, but not all data we observe is equally important. How do humans adjust the weight given to each observation? Here, we've demonstrated two principles that underlie humans' adjustments and their dynamic adaptive learning capabilities. Firstly, when observing a highly surprising event indicating a likely change in the environment, humans reset their knowledge and learn in one shot. Secondly, when their knowledge is more uncertain, humans update it more quickly. These two forces each dominate in one of two key learning contexts, magnitude and probability learning. Our findings advance understanding of the mechanisms of human learning, with implications for the brain and the development of adaptive machines.

Introduction

Humans live in a dynamic world that requires them to constantly update their knowledge as events unfold. For example, new construction works that cause significant delays in transportation should prompt users to revise their estimate of their typical commute time. This learning process is a fundamental part of our adaptive behavior, and a field of study that spans the sciences of behavior, brain, and machines (Rescorla & Wagner, 1972; Rosenblatt, 1961; Schultz et al., 1997; Sutton & Barto, 2018).

Learning is challenging because the environment is often not only dynamic, but also stochastic. When we take public transportation, for example, our commute time varies from day to day due to stochastic fluctuations. However, a delay can also be caused by an abrupt change, such as construction work, that may persist for several weeks. In this case, we face an ambiguity when a delay occurs: is it due to stochastic fluctuations (in which case we should not significantly change our estimated average commute time), or a genuine change point (in which case we should revise our estimate more drastically)? The challenge lies in determining the appropriate weight to give to the current observation (the observed delay) in our learning process.

A descriptive tool to quantify the weight given by the learner to an observed outcome is the *apparent learning rate*, which we will simply call *learning rate* hereafter (Heilbron & Meyniel, 2019; Nassar et al., 2010). The learning rate measures the amount of update of the learned value (from v_{t-1} to v_t) induced by the observed outcome x_t in proportion to its deviation from the previously learned value:

$$\alpha_t = (v_t - v_{t-1}) / (x_t - v_{t-1}) \quad [1]$$

This tool is useful for characterizing the learning process and distinguishing several levels of adaptive learning (Foucault & Meyniel, 2021; Soltani & Izquierdo, 2019). Level zero (non-adaptive) corresponds to a constant learning rate, as in the delta rule, a widely used learning model in psychology and neuroscience. This model accounts for the basic fact that the amount of update increases with the discrepancy between the expectation and the observed outcome (also known as prediction error). It has led to numerous successes across various forms of learning, including associative learning and reinforcement learning in humans, animals, and artificial agents (O'Doherty et al., 2003; Rescorla & Wagner, 1972; Rosenblatt, 1961; Schultz et al., 1997; Sutton & Barto, 2018). Beyond level zero, adaptive learning processes involve an adaptation of the learning rate at some level. We distinguish two levels of adaptive learning: one is at the average level over a block of trials, and the other is more fine-grained, at the level of individual trials.

A first level of adaptive learning corresponds to adjusting the average learning rate to the global statistics of the environment. Two relevant statistics for adjusting the average learning rate are the average frequency of change points (also known as volatility) and the level of stochasticity in the environment (Soltani & Izquierdo, 2019). To illustrate, change points are more frequent in transportation systems where construction campaigns are more frequent, and commute times are less stochastic in transportation systems that are better organized. When change points are more frequent, the average learning rate should be higher to update the estimated value more quickly. Accordingly, previous studies have shown that humans use a higher average learning rate in a block of trials containing many change points compared to a block with no change points (Behrens et al., 2007; Browning et al., 2015; Cook et al., 2019). When the environment is more stochastic, the average learning rate should be lower to stabilize the estimated value. Such an effect has also been observed in humans (Lee et al., 2020).

A second level of adaptive learning corresponds to dynamically adjusting the learning rate from one outcome to the next depending on what is observed—we refer to it as *dynamic adaptive learning*. Such dynamic adjustments are particularly critical to learn effectively in a dynamic and stochastic environment, so as to increase the learning rate locally when a single change point is detected (Foucault & Meyniel, 2021; Nassar et al., 2010). Compared to the first level (different average learning rates between blocks of trials), less is known in humans about this second level (dynamic adjustments of the learning rate at the trial level).

Dynamic adaptive learning has been demonstrated in humans in the case of magnitude learning. Empirical evidence is available for several kinds of magnitudes including symbolic numbers (Nassar et al., 2012), positions (McGuire et al., 2014), and angular directions (Vaghi et al., 2017). However, these studies leave unexplored another common type of learning: probability learning.

The literature on probability learning is vast, from the learning of reward probabilities to stimulus occurrence to the validity of attentional cues, but it remains unknown whether humans dynamically adapt their learning of probabilities. Previous studies are not suitable to examine trial-level adjustments of the learning rate. Many of them are based on choices, from which only an average learning rate can be estimated, as a model parameter that is fitted across many choices and compared between blocks of trials (Behrens et al., 2007; Browning et al., 2015; Cook et al., 2019; Cools et al., 2002).

From a learning perspective, magnitudes and probabilities present a fundamental difference: the amount of information provided by a single outcome about the quantity to be learned is typically much lower for probabilities than for magnitudes (see Fig. S1 for common examples). This amount of information should in principle regulate the learning rate: the more information the outcome provides about the quantity to be learned, the higher the learning rate should be. In particular, a single outcome may be informative enough to detect a change point and thus immediately increase the learning rate. Conversely, when outcomes are less informative, change points are more ambiguous (difficult to distinguish from stochastic fluctuations), making dynamic adaptive learning all the more challenging.

In light of these differences, what has been demonstrated about human adaptive learning in the context of magnitudes may not generalize to probabilities. A comprehensive understanding of dynamic adaptive learning requires to study probability learning and to compare it directly to magnitude learning, in order to uncover the extent and determinants of dynamic adjustments in each case. Here, we address two main questions:

1) Do humans dynamically adapt their learning rate in probability learning? What are their adjustment dynamics, and how do they compare to those of magnitude learning?

2) What are the computational determinants of the dynamic adjustments of learning rates?

To answer these questions, we conducted a study in which human subjects performed two learning tasks, a magnitude learning task and a probability learning task. We developed a new experimental paradigm to measure the subject's learning rate directly from their report at each outcome. This measure of the learning rate serves to characterize the human learning process.

We used a normative Bayesian approach to find how the learning rate should be adjusted. We identified two computational determinants and we show that they play a quantitatively different role in the two types of learning. Overall, our results show a remarkable alignment of human behavior with normative behavior in dynamic adaptive learning, even in the difficult case of probability learning.

Results

Measuring the learning rate at each outcome during magnitude or probability learning. To measure subjects' learning rate at each outcome as they are learning, we developed a new experimental paradigm where subjects update their estimate in real time as they observe the outcomes occurring one at a time at regular intervals of 1.5s (Fig. 1). We obtained real time reports by continuous motion tracking: the estimation is done with a slider that follows the subject's motion, captured on a touchpad or a mouse (Fig. 1 A and B). The subject's estimate (v_t, obtained from the slider position) is thus collected at each outcome (x_t, t indexes the outcome). This allows us to calculate the learning rate for each outcome using equation [1].

Note that the learning rate calculated in this way is directly obtained from the subject's estimates (Nassar et al., 2010), rather than derived from model fitting, as is often done (Behrens et al., 2007; Browning et al., 2015; Cook et al., 2019). This learning rate is a behavioral measure that makes no assumptions about the computational process used by subjects for learning. The learning rate can thus be measured for all kinds of learning models and compared to subjects.

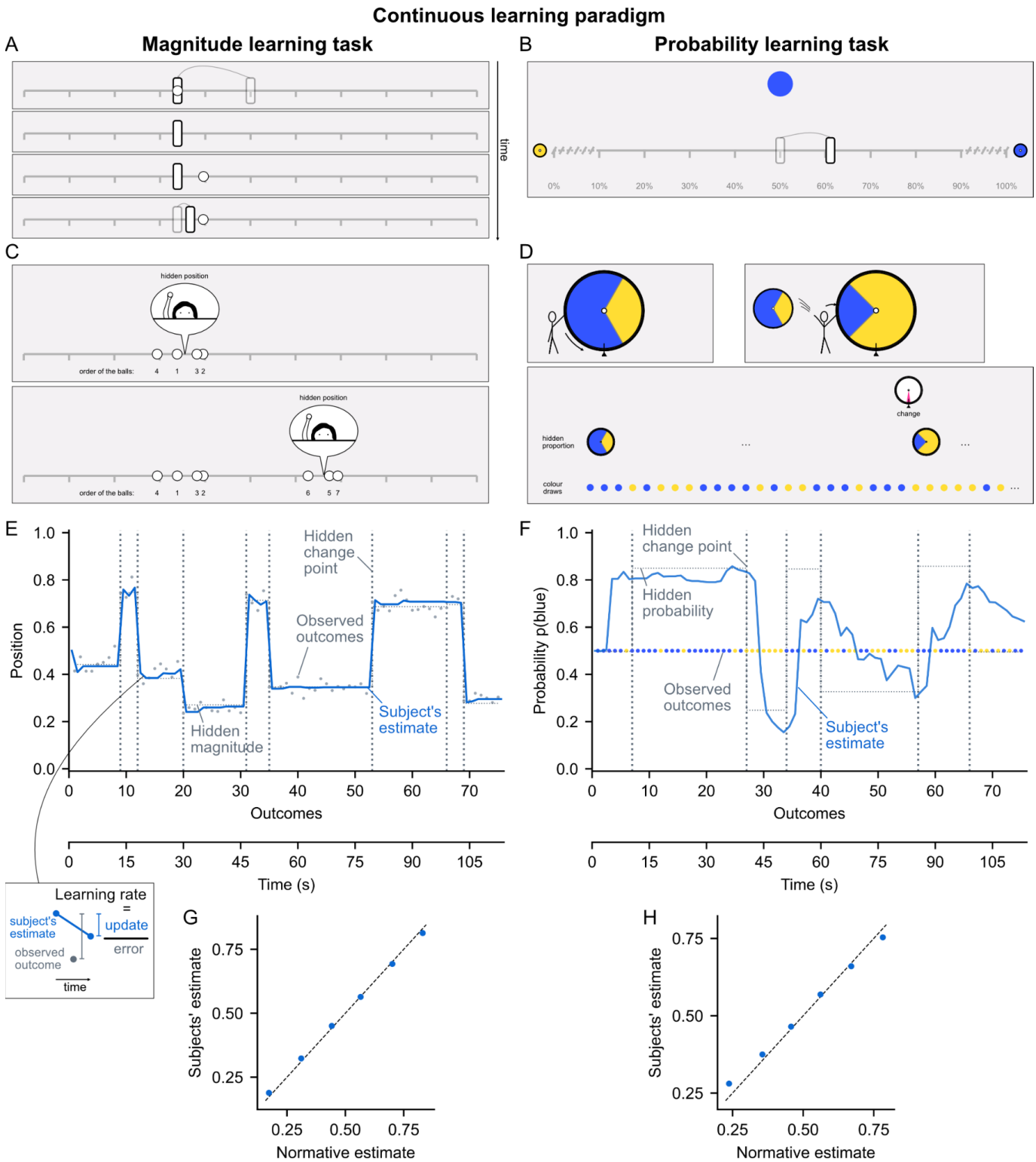

Fig. 1. Magnitude and probability learning tasks with estimates reported in real-time. (A and B) Screenshots of the two tasks. The subject must estimate a hidden quantity (magnitude or probability) based on the stimuli which appear one by one at regular intervals (1.5s). This quantity changes at hidden, unpredictable times called change points. The subject reports their current estimate by moving a slider in real time via motion tracking. For the magnitude task (A, frames show successive times), the hidden quantity is the mean horizontal position generating the stimuli (white circles). For the probability task (B), it is the probability that the circle appearing in the center is blue vs. yellow. Semi-transparent elements indicating slider movements were added here for explanation. (C and D) Screenshots of the task instructions. For the magnitude task (C), subjects were told that they should estimate the position of a hidden person throwing snowballs based on where the balls are landing (the white circles), and that the person could change position at random times without their knowledge. For the probability task (D), subjects were told they should

estimate the proportion of blue and yellow on a hidden wheel used to draw the observed colors, and that the wheel could be changed at random times without their knowledge. (E and F) Example of one session of the magnitude (E) and probability (F) task. Vertical dotted lines indicate the first outcome after a change point. Thin dotted lines show the hidden quantity values. Dots represent outcomes (positions in (E), colors in (F)). The blue line shows the subject's estimate after each outcome (0: left edge of the slider, 1: right edge). (G and H) Accuracy of subjects' estimates for each outcome compared to normative (i.e. optimal) estimates in the magnitude (G) and probability (H) task . The data were binned in 6 equal quantiles of normative estimate and averaged within-subjects. Points and error bars (too small to be seen) show mean ± s.e.m. across subjects, dashed line is the identity line.

We developed two learning tasks using this paradigm: a magnitude learning task and a probability learning task. In order to compare their results, we adopted a similar design and task structure. In both tasks, the subject's goal is to estimate a hidden quantity (magnitude or probability) from the outcomes they observe, and to produce estimates as accurate as possible at each outcome. During the outcome sequence, the hidden quantity undergoes discrete changes at unpredictable random times called change points, which are also hidden (subjects are not told when a change point occurs). In the magnitude learning task, the hidden magnitude is the mean horizontal position generating the observed outcomes. Subjects were instructed that a hidden person was throwing snowballs and that their goal was to estimate the position of this hidden person based on where the snowballs are landing, knowing that the hidden person could change position from time to time without their knowledge (Fig. 1C). In the probability learning task, the hidden probability is the probability of the centrally-presented outcome being blue (vs. yellow). Subjects were instructed that the colors were drawn by spinning a wheel filled with a certain proportion of blue and yellow and that their goal was to estimate this proportion, knowing that the wheel could be changed from time to time without being told when it changes (Fig. 1D). The outcome sequences were generated, in the magnitude task, using the same parameters as (Nassar et al., 2012) for comparison (same or very similar parameters were also used in other magnitude learning studies (McGuire et al., 2014; Vaghi et al., 2017)). In the probability task, we used a probability of change point occurrence that made the task engaging and difficult but still tractable for subjects. (See Methods for more details.)

The human subjects (n=96) performed the two task one after the other, in a counterbalanced order across subjects (see an example session for each task Fig. 1 E and F). We evaluated subjects' estimate accuracy at the level of each outcome by comparing their estimates with the normative estimates, i.e. the optimal estimates given the observed outcomes. Although change points were frequent, subjects were able to produce accurate estimates (close to normative) of magnitudes and probabilities (Pearson r=0.96±0.01 and 0.80±0.01 mean±s.e.m., t_{95}>55.7, $p<10^{-73}$ in both cases, see Fig 1 G and H).

Subjects dynamically adapt their learning rate after a change point for probabilities, and differently than for magnitudes. As per question 1), we investigated subjects' ability to dynamically adapt their learning rate in response to change points, by calculating their learning rate with equation [1] and analyzing its dynamics around the change points. An optimal learner should increase their learning rate when they believe that a change point has occurred recently, to quickly update their estimate, and then decrease their learning rate to stabilize their estimate (see illustration of this behavior and how it differs from that of a learner using a fixed learning rate in Fig. 2C). Note that change points are more difficult to detect in the probability learning task because the amount of information provided by a single outcome is much lower than for magnitude learning (Fig. S1). Despite this difficulty, subjects dynamically adapted their learning rate remarkably well, not only in the magnitude learning task, but crucially also in the probability learning task (Fig. 2 A and B).

In the magnitude learning task, the learning rate dynamics we found (Fig. 2A) replicate those found in previous studies of magnitude learning (McGuire et al., 2014; Nassar et al., 2012; Vaghi et al., 2017). The increase in learning rate occurs immediately and mainly at the first outcome after a change point, with a learning rate suddenly close to 1. In the probability learning task, the learning rate dynamics show an increase that is smoother and more prolonged, remaining significantly higher at the eighth outcome after a change point compared to before the change point (Fig. 2B). Importantly, these different dynamics are very similar to those prescribed by normative theory (see insets in Fig. 2; the Pearson correlation between subject's mean dynamics and the normative one is 0.99 in both tasks; the amplitudes of their dynamics are also similar to the normative ones, shown in further detail in Fig. S2). We also controlled for the frequency at which subjects updated: our results are replicated even when excluding from analysis all outcomes where subjects did not report an update (Fig. S4). (Note that subjects frequently updated their estimate in our study, see Fig. S3).

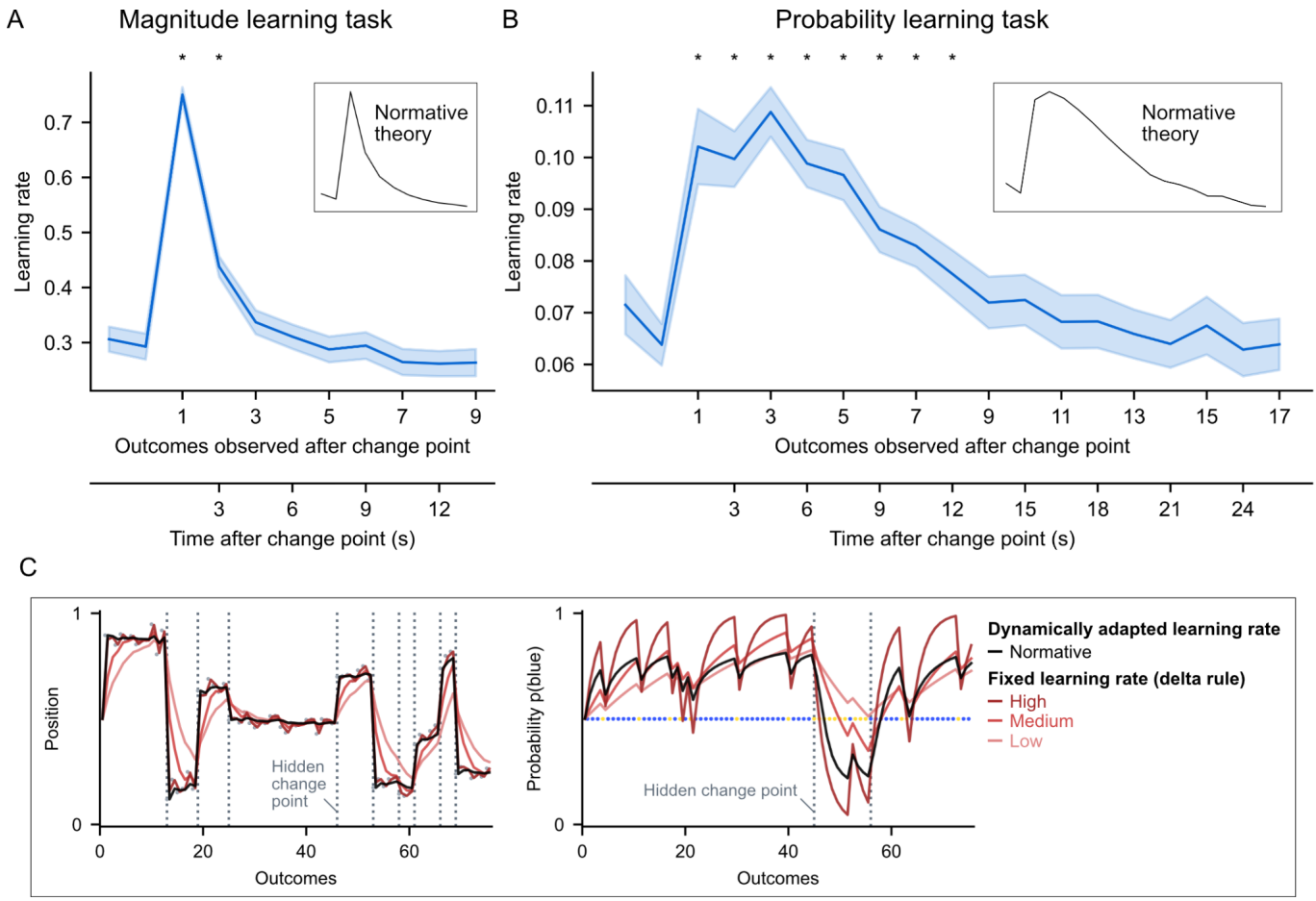

Fig. 2. Dynamics of subjects' learning rate in response to a change point in the two tasks. The subjects' learning rates were measured at each outcome (equation [1]), then aligned to change points and averaged within-subject. Lines and error bands show mean ± s.e.m. across subjects. Subjects dynamically adapt their learning rate after a change point, increasing it transiently and mainly at the first outcome in the magnitude learning task (A), and more prolongedly in the probability learning task (B). Stars show statistically significant differences compared to the baseline measured at the two outcomes before the change point ($p<0.05$, two-tailed, FWE cluster-corrected for multiple comparisons across time). The subjects' learning rate dynamics are very similar to those prescribed by normative theory, shown in the insets. (C) Examples illustrating the difference in behavior between a fixed learning rate (delta-rule, red curves, for different values of the learning rate) and a dynamic one (normative learner, black curve). When the learning rate is fixed, regardless of its value, the behavior shows systematic deviations making learning less efficient: after a change point, the estimates are too slowly updated, and in the absence of change points, they fluctuate too much following the random variability in outcomes. By dynamically increasing and then decreasing the learning rate after a change point, the normative learner quickly updates and then stabilizes its estimate.

Two normative determinants regulate subjects' learning rate dynamics, and differentially in magnitude vs. probability learning. Having demonstrated the dynamic nature of subjects' learning rate, we sought to identify the determinants that drive these dynamics (question 2). To this end, we formalized the two tasks within the same probabilistic framework and resorted to the general principles that govern learning according to normative theory. Below, we summarize the essential elements for understanding the normative determinants identified.

During learning, the normative model computes a distribution for the value of the hidden quantity (h_t) given the outcomes observed so far ($x_{1:t}$), i.e. $p(h_t|x_{1:t})$. The model's (point) estimate is the mean of this distribution (which is optimal in the sense of minimizing the mean squared error with the true hidden quantity). Each update depends on the observed outcome x_t and the distribution prior to that outcome, $p(h_{t-1}|x_{1:t-1})$. We identified two main computational factors that modulate the (apparent) learning rate of this normative updating process (illustrated in Fig. 3A).

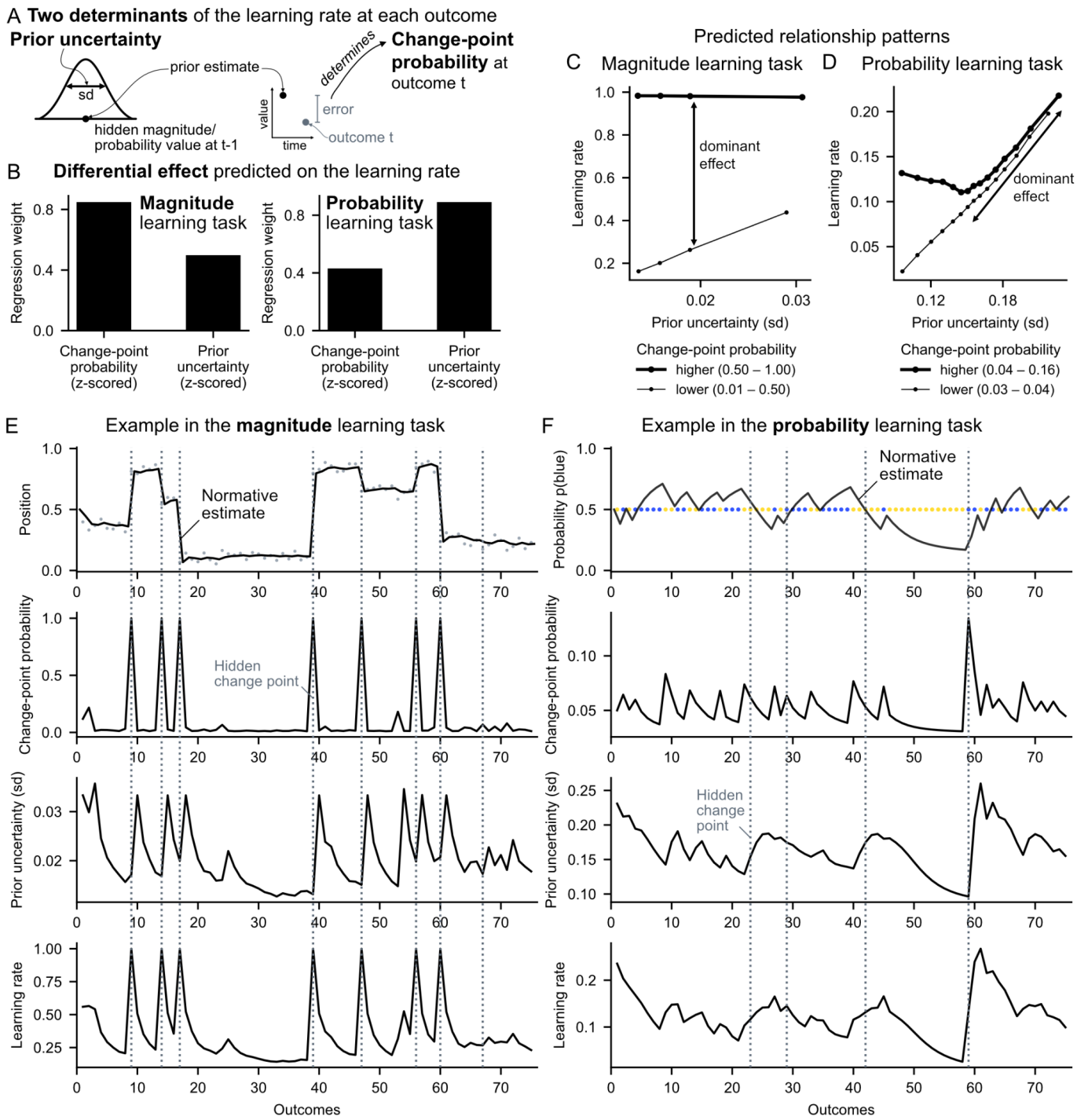

Fig. 3. Normative theory: Two determinants of the dynamic adjustments of the learning rate. (A) The prior uncertainty is the standard deviation of the normative distribution associated with the estimate prior to the last-observed outcome t (equation [2]). The change-point probability is the inferred probability that a change point occurred at that last outcome (equation [3]). (B) Differential effect predicted by the normative model in the two tasks: In the magnitude task, the effect of change-point probability should dominate over that of prior uncertainty, whereas in the probability task, the effect of prior uncertainty should dominate. Effects are quantified by the weights of a multiple regression of the two factors on the normative learning rate (to make the weights commensurable, all variables were z-scored). (C and D) Relationship patterns predicted between the learning rate, prior uncertainty and change-point probability in the magnitude (C) and probability (D) task. In B–D we carried out the analyses on the model in the same way as for subjects in Fig. 4. (E and F) Examples illustrating the dynamics of the different normative variables as a function of the sequence in the magnitude (E) and probability (F) task.

(i) The *prior uncertainty* (u_t), i.e., the uncertainty associated with the estimate *prior to the last outcome*, quantified by the standard deviation of the model distribution:

$$u_t = SD[h_{t-1}|x_{1:t-1}] \quad [2]$$

(ii) The *change-point probability* (Ω_t), i.e., the inferred probability that a change point occurred *at the very last outcome*:

$$\Omega_t = p(h_t \neq h_{t-1}|x_{1:t}) \quad [3]$$

Note that the change-point probability reflects the magnitude of the error ($\delta_t = x_t - v_{t-1}$): the larger the error magnitude, the more probable it is that a change point has just occurred (Spearman correlation between change-point probability and error magnitude across outcome sequences: ρ=0.97 and 1.00 in the magnitude and probability learning task, respectively).

The normative model predicts a differential effect of prior uncertainty and change-point probability on the learning rate: both should influence the learning rate, but the effect of change-point probability should dominate over prior uncertainty in the magnitude learning task, and the effect of prior uncertainty should dominate over change-point probability in the probability learning task (Fig. 3B). Note that the ranges of values of change-point probability and prior uncertainty are different, not only within a task, but also between tasks, due to the differences in the informativeness of outcomes in magnitude vs. probability learning (Fig. S1). Thus, we z-scored these quantities to make them commensurable when comparing their effects on the learning rate in Fig. 3B. We further characterized the effects of change-point probability and prior uncertainty by relating them jointly to the learning rate and plotting the qualitative patterns that should be found in the two tasks (Fig. 3 C and D). In the magnitude learning task (Fig. 3C), change-point probability, when it exceeds 0.5, should produce a large increase in learning rate with a reset close to 1 regardless of the prior uncertainty; otherwise prior uncertainty should also produce a smaller increase in learning rate. In the probability learning task (Fig. 3D), prior uncertainty (which is larger than in the magnitude task and varies over a wider range) should cause the largest increase in learning rate; additionally when prior uncertainty is low, change-point probability (which is smaller than in the magnitude task) should produce a small increase in learning rate. These effects are best understood when viewed over time, by relating the variables to each other and to the observed outcomes, see Fig. 3 E and F.

Do human subjects regulate their learning rate according to the two determinant factors as predicted by normative theory? To find out, we performed the same analyses as we had done with the normative model but using the subjects' learning rates instead. All the normative signatures we had identified were found in the subjects. First, the weights of change-probability and prior uncertainty on subjects' learning rates are significant in both tasks (Fig. 4 A and B) (in the magnitude task, mean±s.e.m weight and two-tailed t-test against zero for change-point probability: 0.43±0.02, t_{95}=23.4, $p=10^{-41}$, and for prior uncertainty: 0.14±0.01, t_{95}=13.8, $p=10^{-24}$; in the probability task, for change-point-probability: 0.07±0.02, t_{95}=4.3, $p=10^{-5}$, and for prior uncertainty: 0.15±0.01, t_{95}=17.0, $p=10^{-30}$). Second, in the magnitude task, the weight of change-point probability is significantly higher than that of prior uncertainty (Fig. 4A; two-tailed paired samples t-test: t_{95}=17.1, $p=10^{-30}$; see also Supplementary Results for comparison with previous studies). Conversely, in the probability task, the weight of prior uncertainty is significantly higher than that of change-point probability (Fig. 4B; t_{95}=-4.4, $p=10^{-5}$). Third, a graded co-variation of the learning rate with prior uncertainty can be observed in the magnitude task (Fig. 4C) and even more so in the probability task (Fig. 4D). Fourth, the interaction effects between prior uncertainty and change-point probability on the learning rate in subjects are similar to the normative ones (Fig. 4 E and F).

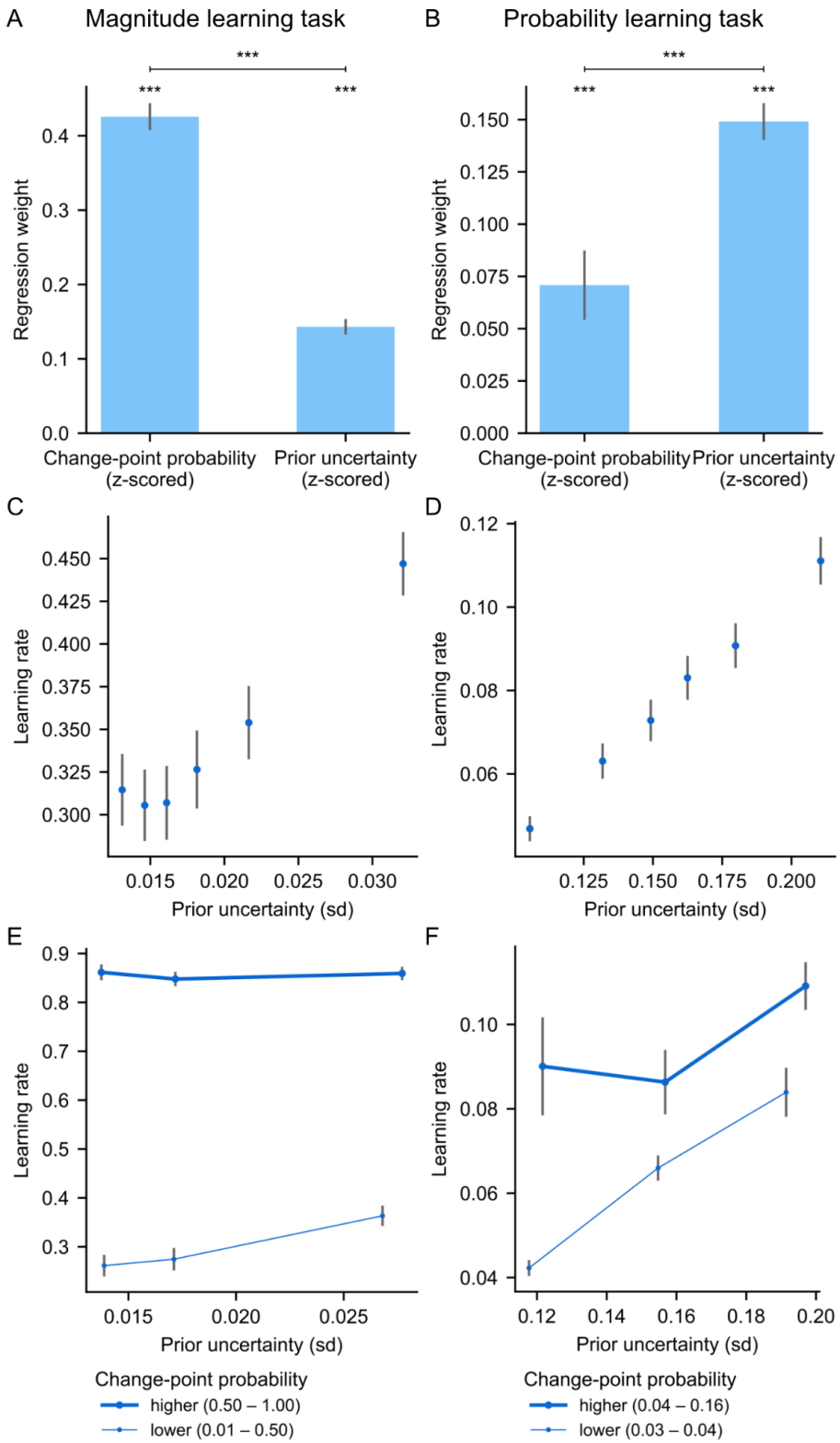

Fig. 4. Subjects' learning rates are differentially affected by the two determinants, as predicted by normative theory. (A and B) Weights of the change-point probability and prior uncertainty on subjects' learning rates in the magnitude (A) and probability (B) task, calculated per subject by multiple regression on the subject's learning rate. As predicted in Fig. 3B, the effect of change-point probability dominates in the magnitude task; that of uncertainty dominates in the probability task. ***: $p<0.001$, two-tailed t-tests. (C and D) Subjects' learning rate increases with prior uncertainty, especially in the probability task. (E and F) The interaction effects between prior uncertainty and change-point probability on subjects' learning rate are similar to the normative model (Fig. 3 C and D). In all plots, subjects' learning rates were measured at each outcome (equation [1]), and prior uncertainty and change-point probability were calculated using the normative model for the same outcomes. Bars/dots and error bars show mean ± s.e.m. across subjects.

Average human learning is close to normative. So far, we have investigated the many ways in which human learning behaves normatively, but it might also have systematic deviations. Here, we studied systematicity at the group level by quantifying a systematic deviation as the degree to which the average estimate of human participants for a given outcome deviates from normative behavior. Systematic deviations should be distinguished from non-systematic deviations, arising from mere variability. What proportion of the deviations can be attributed to the systematic vs. non-systematic component at the group level?

To disentangle the two, we presented identical outcome sequences across subjects to have multiple estimation samples for the same sequence (see Fig. 5, examples at the bottom). This allows us to empirically calculate, for each outcome in the sequence, the systematic deviation formally referred to as bias (the squared difference between the subjects' mean estimate and the normative estimate), and the non-systematic deviations attributed to estimate variance. Note that the variance measured in this way at the group level encompasses both inter-subject variability proper (which accounts for individual biases non-reproducible across individuals) and intra-subject variability (which could be observed through multiple presentations of the same sequence to the same subject; in our case, with one presentation per subject, this variability is reflected across subjects) (Drugowitsch et al., 2016). According to the bias-variance decomposition, the total deviation is equal to the sum of the bias term and the variance term (Hastie et al., 2009). Thus, we can calculate the proportion of deviations attributed to the systematic component (bias) or the non-systematic component (variance).

Using this method, we found that only a small proportion of the deviations were systematic: on average across sequences, the bias/variance ratio was 26.8/73.2% (±1.0% s.e.m.) in the magnitude learning task and 17.2/82.8% (±0.7% s.e.m.) in the probability learning task (Fig. 5). This indicates that there are relatively few systematic deviations in average human learning.

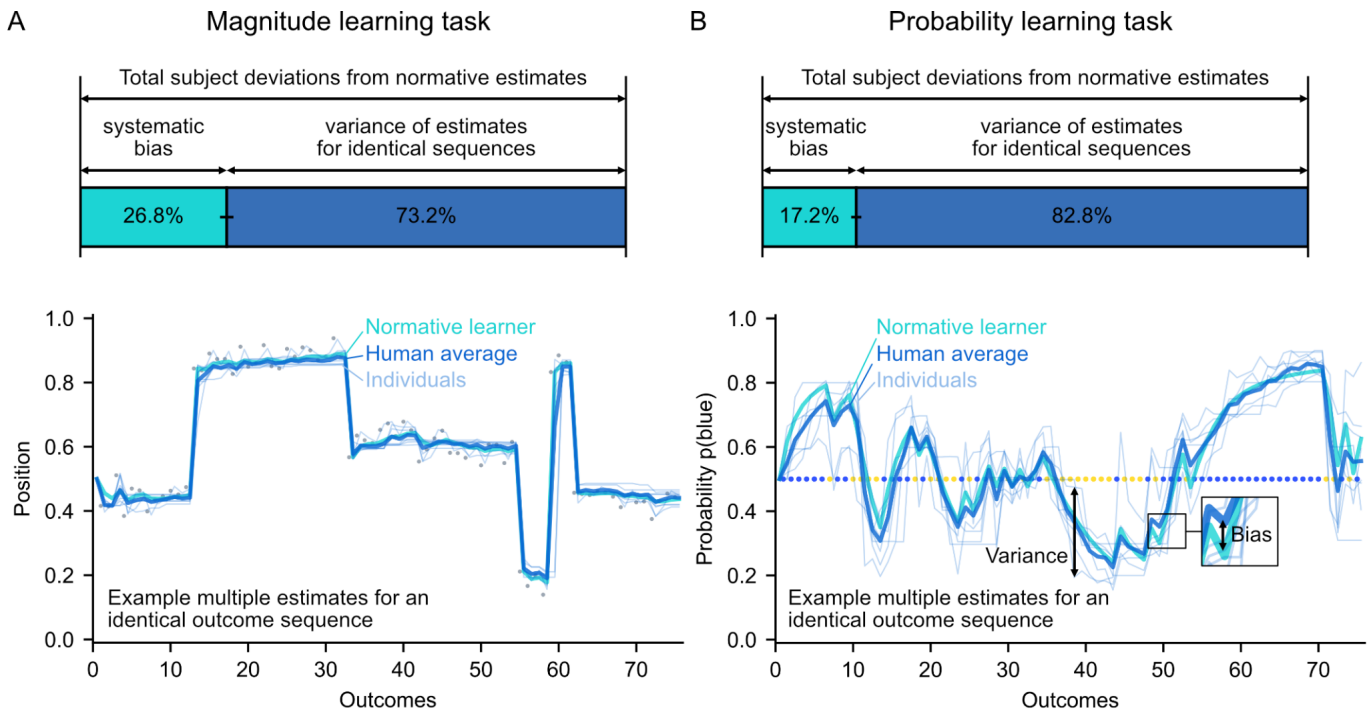

Fig. 5. Group-level decomposition of deviations from normative learning reveals minor systematic deviations compared with the overall variance. Identical outcome sequences were presented across subjects. The bottom plots show, for one of the sequences in each task and each outcome in the sequence, the estimates of different subjects, their average and the normative estimate. The total deviation between the subjects' estimates for a given outcome and the normative estimate is equal to the sum of two terms: the bias and the variance, which respectively measure systematic and non-systematic deviations across subjects. We calculated the proportion of bias/variance for each sequence. The bar graphs at the top show the mean ± s.e.m proportion across sequences for the magnitude (A) and probability (B) task.

Discussion

Many studies in psychology and neuroscience use a fixed learning rate to model the subject's behavior (with models known by various names such as delta rule, Rescorla-Wagner, temporal difference (TD) learning, Q-learning, SARSA...) (O'Doherty et al., 2003; Rescorla & Wagner, 1972; Rosenblatt, 1961; Schultz et al., 1997; Sutton & Barto, 2018). Our results actually show that in humans, the learning rate varies dynamically from one outcome to the next based on what is observed, and is normatively adapted in response to inferred changes in the environment (Fig. 2). We found evidence for these normative dynamic adjustments of the learning rate in two common types of learning, namely magnitude learning and probability learning, despite fundamental differences that exist between these two types of learning, demonstrating the generality of dynamic adaptive learning in humans.

The dynamic adaptation we have shown is distinct from an average adaptation to global environmental statistics such as volatility. In the latter, the average learning rate was found to be higher in a block of trials containing many change points (volatile block) than in a block containing no change points (stable block) (Behrens et al., 2007; Browning et al., 2015; Cook et al., 2019). Here, the adjustment of the learning rate we report is not on average across trials but from trial to trial, in response to single change points. This dynamic adaptation implies an adaptation to volatility (because more volatile environments have more change points, inducing more frequent increases in the learning rate and thus a higher average learning rate), but the reverse is not necessarily true. For instance, adaptation to volatility can arise from a switch between strategies that are tailored to different volatility conditions, with each strategy having its own static learning rate; for example, in a choice task, a strategy closer to choice repetition can be used in the stable condition and one closer to win-stay, lose-shift in the volatile condition (Cook et al., 2019).

The generality of our finding is further strengthened by several advantages of our paradigm compared to the numerous probability learning paradigms based on choices used previously, such as probabilistic reversal learning paradigms (Behrens et al., 2007; Browning et al., 2015; Cook et al., 2019; Cools et al., 2002). First, our measure of the subjects' learning rate does not require any model. It is obtained directly from the subjects' behavior (equation [1]). It is therefore not sensitive to the researcher's modeling assumptions, as is the case when the learning rate is estimated by fitting a model parameter across trials. Second, change points in our tasks are not limited to simple reversals and they occur at unpredictable random times, rather than in a predictable way as in other tasks (e.g. at regular intervals, or within a predetermined window, or according to a deterministic rule based on the subject's choices). This eliminates any possibility that subjects may have taken advantage of a predictable task structure (as do some models of choices in reversal tasks such as (Costa et al., 2015; Wilson et al., 2014)).

Given that dynamic adaptive learning has been demonstrated in our study, research in neuroscience should now uncover its neural mechanisms. To study this question, our paradigm could be reused in neuroimaging experiments, and as a starting point, neural measurements could be related to the computational determinants we presented. Catecholaminergic neuromodulators (noradrenaline and dopamine) are good candidates for investigation as several theoretical and empirical works indicate that they may play a role in regulating learning based on surprise and uncertainty (Aston-Jones & Cohen, 2005; Yu & Dayan, 2005). These neuromodulators provide a biological implementation of gating, also known as gain modulation, a computational mechanism that we have previously shown to play a crucial role for artificial neural networks to achieve the very same dynamic adaptations that we have found here in humans (Foucault & Meyniel, 2021).

Differences between magnitude and probability learning. In light of the differential effects we found in the magnitude and probability learning tasks (Fig. 4), it will be important in future studies of learning to distinguish the contribution of, on the one hand, the effects related to the immediately observed outcome and the quantities associated with it, such as error magnitude, surprise and change-point probability, and on the other hand, the effects of prior uncertainty arising from past outcomes. Moreover, it will be important to distinguish the role of these two types of effects in magnitude learning vs. probability learning, as they do not necessarily involve the same mechanisms in these two types of learning. In the present study, we found that there was overall little correlation between the magnitude and probability task across subjects in terms of how much they weighted change-point probability (no correlation, r=-0.01) and prior uncertainty (significant but weak correlation, r=0.26), compared to e.g. how frequently they updated their reports (r=0.63) (see Supplementary Results for more details). This invites us to be particularly careful and to not

consider a priori that the results discovered in one type of learning will generalize to all types of learning under uncertainty, because it runs the risk of confusing what are potentially very different processes.

Comparing previous fMRI studies conducted separately on magnitude learning (McGuire et al., 2014) and probability learning (Bounmy et al., 2023; Meyniel & Dehaene, 2017), we noted differences between the two in the neural correlates that were specific to each type of effect. Specifically, when examining the selective effects of the error (quantified by change-point probability or surprise), we noted that frontal regions were implicated mainly in probability learning (right superior frontal gyrus, left and right supplementary and cingulate eye fields). In magnitude learning, they involved primarily medial regions of the visual cortex (calcarine sulci), whereas in probability learning the effects in visual cortex were more lateral (left/right V3/V4). When examining the selective effects of prior uncertainty (quantified by relative uncertainty or confidence), we noted that ventromedial and anterior prefrontal cortex regions were involved primarily in magnitude learning. Interestingly, both types of learning showed neural correlates in the parietal cortex, but in magnitude learning these correlates were specific to prior uncertainty, whereas in probability learning they were common to both prior uncertainty and surprise.

In general, several factors may give rise to differences between the two types of learning. First, there may be distinct neural mechanisms for processing different types of quantities (magnitudes vs. probabilities). Second, the ranges and dynamics of change-point probability and prior uncertainty are different in magnitude and probability learning (Fig. 3), and different neural substrates may be recruited in these different regimes. Third, these two factors may be processed differently in the two types of learning because, in magnitude learning, change-point probability is more diagnostic than prior uncertainty for detecting change points, while in probability learning, prior uncertainty is a more reliable signal for inferring a change point in the recent past (Fig. 3 E and F). In addition, there may be a trade-off in the degree of processing one can allocate to compute and use these two factors that favors the most relevant factor in a given learning context, namely the change-point probability when learning magnitudes and the prior uncertainty when learning probabilities (Fig. 3B, 4 A and B).

Frequent updating during the learning process. A previous study proposed that the brain does not update its probability estimate at each outcome, but only on rare occasions (Gallistel et al., 2014). This conclusion was based on an empirical observation: in Gallistel et al.'s study, subjects overtly updated their reported probability estimate only once every 18.1 outcomes on average, with only 7% of updates occurring after a single outcome. However, a lack of updating in the subject's report may not necessarily reflect a lack of updating in the subject's internal learning process. In contrast to Gallistel et al.'s study, our study shows frequent updating during the learning process. Subjects updated their estimate at each outcome in the vast majority of cases: 84% of updates occurred after a single outcome (vs. 7% in Gallistel et al.), and on average subjects updated their estimate every 1.4 outcome (vs. 18.1) (Fig. S3). This makes our data on human probability learning rather novel.

This difference in the frequency of report updates between Gallistel et al.'s study and our study can be explained by task differences: the cost for the subject to overtly update their estimate greatly varies depending on how the task is designed, and when this cost is high, subjects may not manifest their update. In Gallistel et al.'s task, the action required to update the report induces a time cost which, by reproducing the task interface, we estimated to be 4 to 5 s per outcome. If subjects had performed this action for all 1,000 outcomes of a session, it would have taken them more than an hour (estimated 67 to 83 min) per session, instead of the 25.6 min they took on average (Gallistel et al., 2014). In our task, there is no time cost since the outcomes occur at fixed time intervals. This design choice seems to have made subjects much more willing to express their update. Subjects in our study also received a performance bonus encouraging them to report their updates.

The frequency of subjects' updating is critical to the results we have shown on the dynamics of the learning rate at the scale of an outcome. Note that we replicated all our results after stringently excluding from analysis all trials in which the report was not updated (Fig. S4).

Computational models of human learning. The results shown here impose strong computational constraints on models of human learning: they must be able to exhibit the same dynamic adaptations as humans do in response to change points and depending on change-point probability and prior uncertainty (Fig. 2 to 4). They should also be able to reproduce the variability of human learning, and the fact that it is on average close to normative (Fig. 5).

Various models have been proposed in the literature, including the Kalman filter (Kalman, 1960), the Hierarchical Gaussian Filter (HGF) (Mathys et al., 2011), a model estimating the volatility and stochasticity of the environment (assumed to follow a random walk as in the Kalman filter) (Piray & Daw, 2021), the proportional-integral-derivative (PID) controller (Ritz et al., 2018), an adaptive mixture of delta-rules (Wilson et al., 2013), a model with metaplastic synapses guided by a change detection system (Iigaya, 2016), the normative model we used here (which computes the optimal solution via Bayesian inference) (Adams & MacKay, 2007; Behrens et al., 2007; Heilbron & Meyniel, 2019), particle filters (which uses sampling to approximate the optimal solution) (Brown & Steyvers, 2009; Prat-Carrabin et al., 2021), and several architectures of recurrent neural networks (RNNs) (Foucault & Meyniel, 2021; Wang et al., 2018). The Kalman filter is a particular case because it qualifies for level 1 but not level 2 of adaptive learning as we presented in the introduction: its learning rate depends on the global level of stochasticity and volatility in the environment, but does not depend on what is observed (the changes in learning rate over time are simply dictated by the number of observed outcomes). The other models qualify for level 2 (they dynamically adapt their learning rate), but it remains to be evaluated to what extent they can, in both types of learning, conform to the specific normative properties demonstrated here in humans.

In a previous article (Foucault & Meyniel, 2021), we have shown that a small recurrent neural network (as small as three recurrent units) could solve the probability learning task quasi-optimally and reproduce all the above-mentioned normative properties (Fig. 2 to 4). This model thus provides a low-cost, neurally feasible solution. In addition, it has the advantage of not incorporating a generative model of the environment a priori: through mere exposure to the environment, the network spontaneously develops its dynamic adaptive learning capabilities (which is sometimes referred to as ‘meta-learning’) (Foucault & Meyniel, 2021; Wang et al., 2018). One mechanistic insight from our study was that small networks demonstrated optimal adaptive learning capabilities only when equipped with a gating mechanism. This is interesting to put in perspective with the other proposed models as they often involve multiplicative interactions (i.e. multiplication between two state variables or a state variable and an input variable) like gating.

Using our behavioral dataset, future studies will be able to evaluate and compare models to probe the algorithms underlying human learning. New data can also be collected easily and in large amounts using the experimental paradigm that we make available to the community (1 report per outcome every 1.5 s).

Relevance for artificial intelligence. Using a dynamic adaptive learning rate is relevant for machine learning and artificial intelligence. The two determinants we identified could guide the search for new efficient, adaptive learning methods. One advantage of change-point probability and prior uncertainty, compared to state of the art methods such as Adam, AdaGrad and RMSprop (Goodfellow et al., 2016), is that they do not require to evaluate the cost function of the task, which in many cases is not accessible at the level of one input sample (in our task for example, the cost function is the error between the subject’s estimate and the true quantity, which is hidden). This is especially relevant for learning from few input samples, as in one-shot or few-shot learning. According to our findings, change-point probability might be particularly relevant for one-shot learning in contexts where individual samples are highly informative, and prior uncertainty for few-shot learning in contexts where individual samples are less informative. This could be linked to current research in machine learning aimed at improving systems’ ability to assess their uncertainty (Gal & Ghahramani, 2016; Hüllermeier & Waegeman, 2021; Kendall & Gal, 2017; Kompa et al., 2021).

Materials and Methods

Participants.

96 human subjects participated in our study (38 female, median age 30 years, interquartile range 25–37 years, from 19 different nationalities). Participants were recruited from Prolific, an online platform for recruiting participants focusing on academic research. They were required to use a computer with a touchpad or mouse and a screen large enough to perform the task. They were rewarded £8.40 for their participation (£9/h at an a priori estimated completion time of 56 min, which also turned out to be the median completion time of the participants), plus a performance-based bonus of up to £4.20 (50% of the base pay). Informed consent was collected before the start of the experiment. No participants were excluded from the analyses.

Experiment steps.

The experiment consisted of the following steps: Task 1 instructions, Task 1 sessions, Task 2 instructions, Task 2 sessions. Each task session consisted of a sequence of 75 outcomes, taking under 2 min to perform (see examples Fig. 1 E and F). In total, the subjects took a median time of 56 min to complete the whole experiment, including instructions (interquartile range 52–65 min). They performed twenty-one task sessions (six of the magnitude task and fifteen of the probability task; this split was determined a priori with an independent pilot study of eight subjects in order to obtain a sufficiently reliable measure of a subject's learning rate adjustments, with approximately equal relative error in the measurement in both tasks). At the end of each task session, subjects received feedback whose purpose was to maximize subjects' engagement with the task. The feedback display showed the subject's score for the session and the corresponding monetary gain (proportional to the score), and also showed the true values of magnitude/probability along with the subject's estimates. The score was calculated based on the mean absolute error between subjects' estimates and the true values of magnitude/probability. The function mapping the error to the score was designed to: normalize the score in 0–100%; keep the score of the normative learner at a constant level (close to 100%); prevent too low scores (equal or close to 0%) that would demotivate subjects, using a softplus nonlinearity.

Task design.

The graphic elements of the tasks were designed with several goals in mind: a) to allow subjects to easily monitor the stimuli as they are appearing and adjust their estimate at the same time, b) to strike a good balance between speed and accuracy of adjustments, c) to share as much as possible between the two tasks (see Fig. 1 A and B and links below to run the tasks). The total length of the slider track was 640 CSS pixels, such that the leftmost and rightmost reachable position of the slider was at about 6–7º of visual angle from the center (1 CSS px corresponds to 0.0213º visual angle at the typical viewing distance of the user's display, https://www.w3.org/TR/css-values-3/#reference-pixel). This allowed the subject to see the slider and the stimuli at the same time. Tick marks were displayed at every 10% length of the slider to make location easier for subjects. These elements were shared between the two tasks. Additionally, there were a few design elements specific to the probability learning task: labels were put below the tick marks showing the percentage corresponding to the estimated probability (in the magnitude task, this is irrelevant as the position itself is the estimate); the portions of the slider between 0–10% and 90–100% were hatched to indicate that the hidden probability never lies within these intervals; a small yellow wheel and blue wheel were shown at the left and right edge of the slider track, respectively, to indicate the direction of estimation in relation to the two colors; we chose the blue and yellow colors to be easily recognizable by humans even with color-blindness.

Within the 1.5 s interval separating the onset of each outcome, the stimulus offset occurred at 1.3s, and in the 1.3–1.5 s inter-stimulus interval, the slider thumb was highlighted (lighter stroke color) to indicate to subjects when their estimate was recorded for calculating their performance.

Throughout the session, subjects could at any time move the slider via motion tracking to update their estimate. Motion tracking was implemented by continuously tracking movements of the subject's pointer (corresponding to finger movements on a touchpad, or to mouse movements). The pointer itself was hidden during the task, to make it clear to the subject that they were controlling the slider.

In the task instructions, for the magnitude learning task, the choice of the cover story (a hidden person throwing snowballs) was inspired by (Prat-Carrabin et al., 2021).

The tasks can be freely tested online, separately at
https://run.pavlovia.org/cedricfoucault/ada-learn-task/ada-pos-study.html?skipConsent (magnitude learning task) and
https://run.pavlovia.org/cedricfoucault/ada-learn-task/ada-prob-study.html?skipConsent (probability learning task), and and the combined experiment can be tested at
https://run.pavlovia.org/cedricfoucault/ada-learn-task/ada-pos-prob-study.html (magnitude task first) and
https://run.pavlovia.org/cedricfoucault/ada-learn-task/ada-prob-pos-study.html (probability task first).

Outcome sequences and sequence-generating processes of the tasks.

We generated a set of 100 sequences for the magnitude learning task and 150 sequences for the probability learning task. Each time a subject performed the tasks, they were shown sequences randomly sampled without replacement from these sets.

For the sake of clarity and to be able to compare the two tasks, we use here as the reference unit the horizontal position within the slider interval normalized between 0 and 1 (left and right edge of the slider track, respectively). This unit applies to the estimates (slider positions) and the hidden magnitudes/probabilities these estimates are about, and to the outcomes (for the probability task, outcome 'yellow' and 'blue' correspond to '0' and '1', respectively, since when the estimate is equal to 1 this corresponds to a 100% probability of the outcome being blue). For the magnitude learning task, the outcome-generating process was the same as in (Nassar et al., 2012): the outcomes were drawn from a Gaussian distribution with a standard deviation of 10/300 (in normalized unit), and whose mean (the hidden magnitude) changed with a probability of zero for the first three outcomes following a change point and 1/10 for all trials thereafter. For the probability learning task, the outcomes were drawn from a Bernoulli distribution whose parameter (the hidden probability) changed with a probability of zero for the first six outcomes following a change point and 1/20 for all trials thereafter. The probability was sampled uniformly between 0.1 and 0.9 initially, and resampled uniformly in the same interval at each change point subject to the constraint that the resulting change in the odds p/(1-p) be no less than fourfold.

Note that, for the magnitude learning task, we used the same parameter values as previous studies (Nassar et al., 2012; Vaghi et al., 2017). Equalizing the two tasks in terms of the informative value of the outcomes would not be practically feasible (see also Fig. S1).

Behavioral analyses.

The subject's estimate v_t, for outcome x_t, was taken as the normalized position of the slider at the end of the 1.5s interval just before observing the next outcome (the estimate prior to the first outcome, v_0, is 0.5, as the slider is initially positioned in the middle). The value x_t is equal to, for the magnitude task, the (normalized) position of the stimulus, and for the probability task, 1 for a blue draw and 0 for a yellow draw. v_{t-1} and v_t are the subject's estimate prior to and after observing that outcome, and α_t is the subject's learning rate for that outcome.

In the results relating to the subjects' estimate accuracy, the significance was calculated by computing the Pearson correlation at the subject-level between the subject's and the normative estimates, and then performing at the group-level a two-tailed, one-sample t-test of those correlations against zero.

When analyzing learning rates in the magnitude learning task, similar to (Nassar et al., 2012; Vaghi et al., 2017), we ignored from the analyses outliers in the learning rates (which might occasionally occur when the error is very close to 0), corresponding to learning rates above 1.3 or below -0.6. These values were determined a priori using the distribution of the subjects' learning rates in an independent pilot study of eight subjects by taking the values above and below which the empirical density was less than 0.05.

Normative models.

These models compute (using probability calculus and in particular, Bayes' rule) the optimal probability distribution of the latent quantity generating the outcomes, given the outcomes observed by subjects and knowing the actual generative process of outcomes implemented in each task (Ma et al., 2023). Below we highlight the main properties of these models presented in previous studies.

Magnitude learning task. We implemented the reduced Bayesian model described in (Nassar et al., 2012), which has been used in other studies of similar tasks (McGuire et al., 2014; Vaghi et al., 2017). This model computes at each outcome a probability distribution for the value of the hidden magnitude (i.e. the mean of the Gaussian distribution generating the outcomes) given the outcomes observed so far. We computed the change-point probability as described in (Nassar et al., 2012). We computed the prior uncertainty, u_t, according to the definition we provided, as the standard deviation of the distribution for the hidden magnitude (equation [2]). This can be calculated from the variables described in (Nassar et al., 2012)—the relative uncertainty (τ_t) and the total variance on the predictive distribution (σ_t^2)—with the formula: $u_t = \sigma_t \tau_t^{1/2}$.

Note that we deliberately used a quantification of uncertainty that is commensurable—i.e. has the same unit—in the magnitude and the probability learning tasks, to be able to compare the uncertainty between the two tasks, by taking the standard deviation of the model's distribution whose support is the [0–1] interval of the estimation slider, rather than using the "relative" uncertainty as in other studies, which would break this commensurability (McGuire et al., 2014; Nassar et al., 2012; Vaghi et al., 2017).

Probability learning task. We reused the Bayesian model described in (Meyniel, 2020) (code available online at https://github.com/florentmeyniel/TransitionProbModel). We made two modifications to the model to match the subject's actual task:

- As the subjects were told that the hidden probabilities are drawn in [0.1, 0.9] (with the restriction of the slider to that interval), we also incorporated this knowledge into the model, by setting the corresponding distribution equal to the uniform distribution on [0.1, 0.9], rather than [0, 1] as in the original version.
- Since we asked (and rewarded) subjects to estimate the value of the hidden probability at the last-observed outcome, we used the corresponding distribution, for the value of the hidden probability at the last outcome (rather than the next outcome), to compute the model estimate (the mean of the distribution) and uncertainty (the standard deviation, equation [2]).

We computed the uncertainty as the standard deviation of the distribution for the hidden probability computed by the model (equation [2]). To compute the change-point probability, we used the definition given in equation [3].

Data and code availability

All data, and code to reproduce the results will be made freely available at the point of publication on OSF, and on our GitHub repository.

The tasks can also be readily tested online at the links given above.

Acknowledgements

We gratefully acknowledge the support received for this work. C.F. was supported by a PhD fellowship from ENS Paris-Saclay (France). F.M. was supported by the European Research Council (ERC grant #947105) and the Inserm. We thank Maëva L'Hôtellier and Alexander Paunov for useful feedback on the project. We also thank them as well as Steven Geysen and Tiffany Bounmy for piloting the experiment.

References

Adams, R. P., & MacKay, D. J. C. (2007). *Bayesian Online Changepoint Detection*. https://arxiv.org/abs/0710.3742

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*(1), Article 1. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221.

Bounmy, T., Eger, E., & Meyniel, F. (2023). A characterization of the neural representation of confidence during probabilistic learning. *NeuroImage*, 119849.

Brown, S. D., & Steyvers, M. (2009). Detecting and predicting changes. *Cognitive Psychology*, *58*(1), 49–67.

Browning, M., Behrens, T. E., Jocham, G., O'reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, *18*(4), 590–596.

Cook, J. L., Swart, J. C., Froböse, M. I., Diaconescu, A. O., Geurts, D. E., Den Ouden, H. E., & Cools, R. (2019). Catecholaminergic modulation of meta-learning. *Elife*, *8*, e51439.

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *Journal of Neuroscience*, *22*(11), 4563–4567.

Costa, V. D., Tran, V. L., Turchi, J., & Averbeck, B. B. (2015). Reversal learning and dopamine: A bayesian perspective. *Journal of Neuroscience*, *35*(6), 2407–2416.

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.

Foucault, C., & Meyniel, F. (2021). Gated recurrence enables simple and accurate sequence prediction in stochastic, changing, and structured environments. *ELife*, *10*, e71801. https://doi.org/10.7554/eLife.71801

Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *International Conference on Machine Learning*, 1050–1059. http://proceedings.mlr.press/v48/gal16.html

Gallistel, C. R., Krishan, M., Liu, Y., Miller, R., & Latham, P. E. (2014). The perception of probability. *Psychological Review*, *121*(1), 96.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning book. *MIT Press*, *521*(7553), 800.
Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Vol. 2). Springer.
Heilbron, M., & Meyniel, F. (2019). Confidence resets reveal hierarchical adaptive learning in humans. *PLoS Computational Biology*, *15*(4), e1006972.
Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, *110*, 457–506.
Iigaya, K. (2016). Adaptive learning and decision-making under uncertainty by metaplastic synapses guided by a surprise detection system. *ELife*, *5*. https://doi.org/10.7554/eLife.18073
Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, *82*(1), Article 1. https://doi.org/10.1115/1.3662552
Kendall, A., & Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision? *Advances in Neural Information Processing Systems*, *30*.
Kompa, B., Snoek, J., & Beam, A. L. (2021). Second opinion needed: Communicating uncertainty in medical machine learning. *NPJ Digital Medicine*, *4*(1), 4.
Lee, S., Gold, J. I., & Kable, J. W. (2020). The human as delta-rule learner. *Decision*, *7*(1), 55.
Ma, W. J., Kording, K. P., & Goldreich, D. (2023). *Bayesian Models of Perception and Action: An Introduction*. MIT press.
Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, *5*, 39. https://doi.org/10.3389/fnhum.2011.00039
McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally dissociable influences on learning rate in a dynamic environment. *Neuron*, *84*(4), 870–881.
Meyniel, F. (2020). Brain dynamics for confidence-weighted learning. *PLOS Computational Biology*, *16*(6), Article 6. https://doi.org/10.1371/journal.pcbi.1007935
Meyniel, F., & Dehaene, S. (2017). Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *Proceedings of the National Academy of Sciences*, *114*(19), E3859–E3868.
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.
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.
O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*(2), 329–337.
Piray, P., & Daw, N. D. (2021). A model for learning based on the joint estimation of stochasticity and volatility. *Nature Communications*, *12*(1), 6587.
Prat-Carrabin, A., Wilson, R. C., Cohen, J. D., & da Silveira, R. A. (2021). Human Inference in Changing Environments With Temporal Structure. *Psychological Review*, *128*(5), 879–912. https://doi.org/10.1037/rev0000276
Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current Research and Theory*, *2*, 64–99.
Ritz, H., Nassar, M. R., Frank, M. J., & Shenhav, A. (2018). A Control Theoretic Model of Adaptive Learning in Dynamic Environments. *Journal of Cognitive Neuroscience*, *30*(10), Article 10. https://doi.org/10.1162/jocn_a_01289
Rosenblatt, F. (1961). *Principles of neurodynamics. Perceptrons and the theory of brain mechanisms*. Cornell Aeronautical Lab Inc Buffalo NY.
Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.
Soltani, A., & Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nature Reviews Neuroscience*, 1.
Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
Vaghi, M. M., Luyckx, F., Sule, A., Fineberg, N. A., Robbins, T. W., & De Martino, B. (2017). Compulsivity reveals a novel dissociation between action and confidence. *Neuron*, *96*(2), 348–354.
Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., Hassabis, D., & Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience*, *21*(6), 860.
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), Article 7. https://doi.org/10.1371/journal.pcbi.1003150
Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron*, *81*(2), Article 2. https://doi.org/10.1016/j.neuron.2013.11.005
Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), Article 4. https://doi.org/10.1016/j.neuron.2005.04.026

Supplementary information

Supplementary Results

Regression on subject's learning rate in the magnitude learning task performed as in a previous study. For comparison with previous studies on magnitude learning, we additionally performed a regression analysis on the subject's learning rate as it was done in (McGuire et al., 2014), that is, without z-scoring regressors as we did in the regression reported in the main text, and replacing our prior uncertainty regressor by the RU*(1-CPP) regressor. We obtained regression weights similar to but slightly higher than those reported in (McGuire et al., 2014): The median and interquartile range of the regression weights in this analysis are 0.83 [0.56–0.92] and 0.51 [0.24–0.73] in our data for change-point probability and RU*(1-CPP) respectively (all two-tailed signed-rank $p<0.001$), vs. 0.53 [0.40–0.76] and 0.32 [0.11–0.44] in (McGuire et al., 2014).

Correlations across subjects between the magnitude learning task and the probability learning task. We correlated the regression weights obtained in the magnitude learning task with those obtained in the probability learning task across subjects (the weights were obtained using the same regression as in Fig. 4). The weight of change-point probability was not significantly correlated between the two tasks (r=-0.01, p=0.92), and that of prior uncertainty was weakly though significantly correlated (r=0.26, p=0.012) (partial Pearson correlation controlling for the individual's average update frequency, two-tailed p values). This is indeed due to differences between the two tasks: within each task, when performing the same correlation analysis on two halves of the data (even and odd sessions), we obtained strong correlations (these were, in the magnitude and probability task respectively, r=0.64 and 0.95 for change-point-probability, r=0.38 and 0.74 for prior uncertainty, all $p<0.001$, two-tailed). For comparison, another behavioral measure, the average update frequency of the subject, was more strongly correlated across subjects between the two tasks: r=0.63, $p<0.001$ (Pearson correlation, two-tailed).

Supplementary Figures

Continued on next page

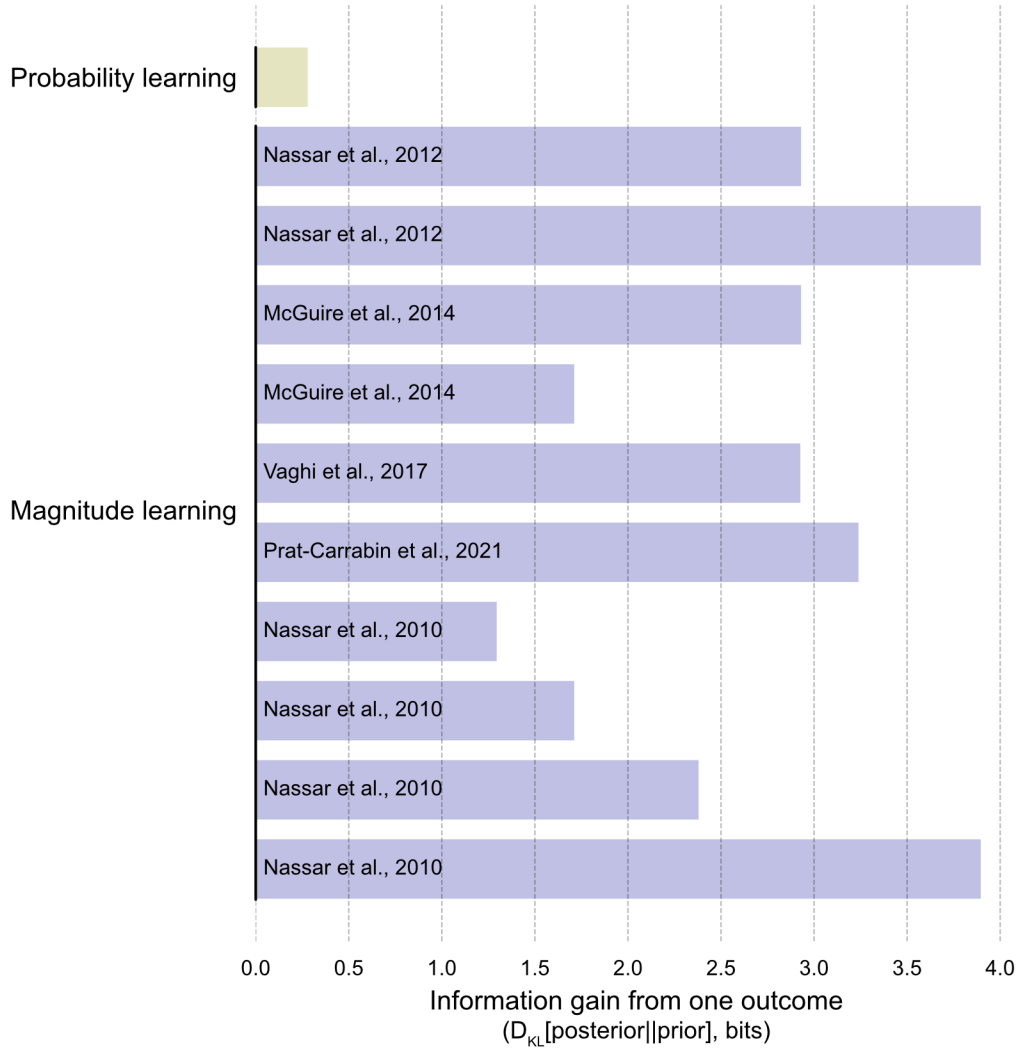

Fig. S1. Information gain provided by a single outcome about the quantity to be learned in magnitude learning and probability learning. We computed the information gain as the KL divergence between the prior (uniform) distribution before having observed any outcome, and the posterior distribution about the underlying quantity after having observed the outcome, using the prior as reference distribution (i.e. D_{KL}[posterior||prior]), on average over the possible outcomes. The posterior is obtained from the prior and the likelihood function relating the outcome to the underlying quantity using Bayes rule. In probability learning, the information gain is minimal. This is due to the binary nature of the outcome. In magnitude learning, the information gain is larger because the outcome is quantitative and typically fairly representative of the underlying magnitude. Although the latter depends on the experimenter's choice of standard deviation with which outcomes are generated, we computed the information gain for numerous experiments previously conducted and each condition of these experiments, and as shown above, in all cases it was substantially higher than that obtained in probability learning (McGuire et al., 2014; Nassar et al., 2010, 2012; Prat-Carrabin et al., 2021; Vaghi et al., 2017).

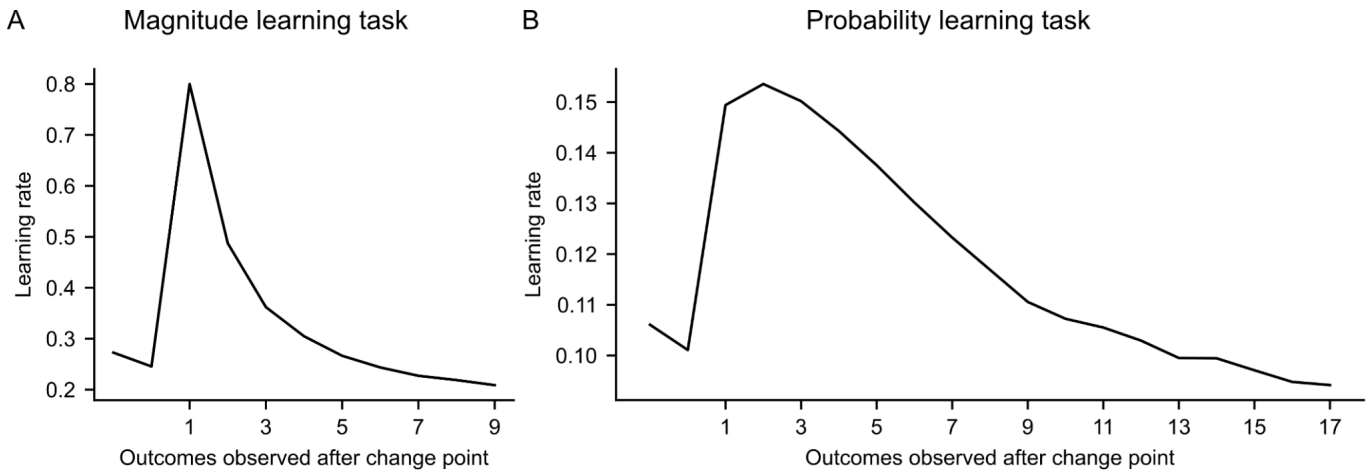

Fig. S2. Dynamics of the normative model's learning rate after a change point, in the magnitude (A) and probability (B) learning tasks. The plots were obtained as in Fig. 2, but rather than using the subjects' learning rate, we used the normative model's learning rate instead, which we obtained by running the normative model on the same sequences as the subject.

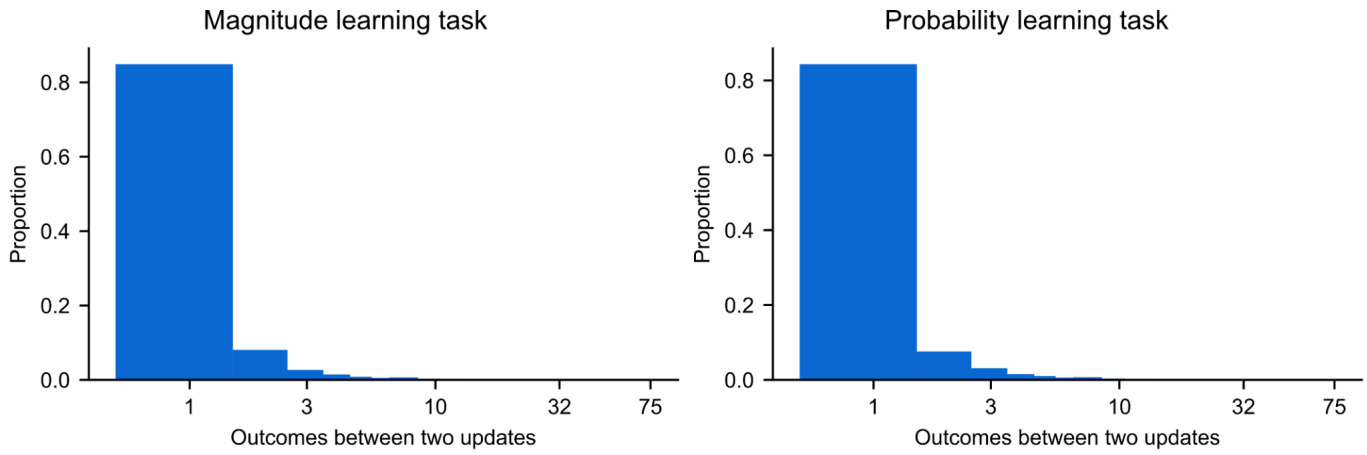

Fig. S3. Distribution of the number of outcomes elapsed between two report updates made by subjects. A log scale was used for the number of outcomes as in (Gallistel et al., 2014) for comparison (the equivalent distribution in Gallistel et al. is shown in their Fig. 11). In contrast to (Gallistel et al., 2014), in our study, updates were made on each outcome most of the time (84% in the above distribution; mean of the distribution: 1.4 outcome).

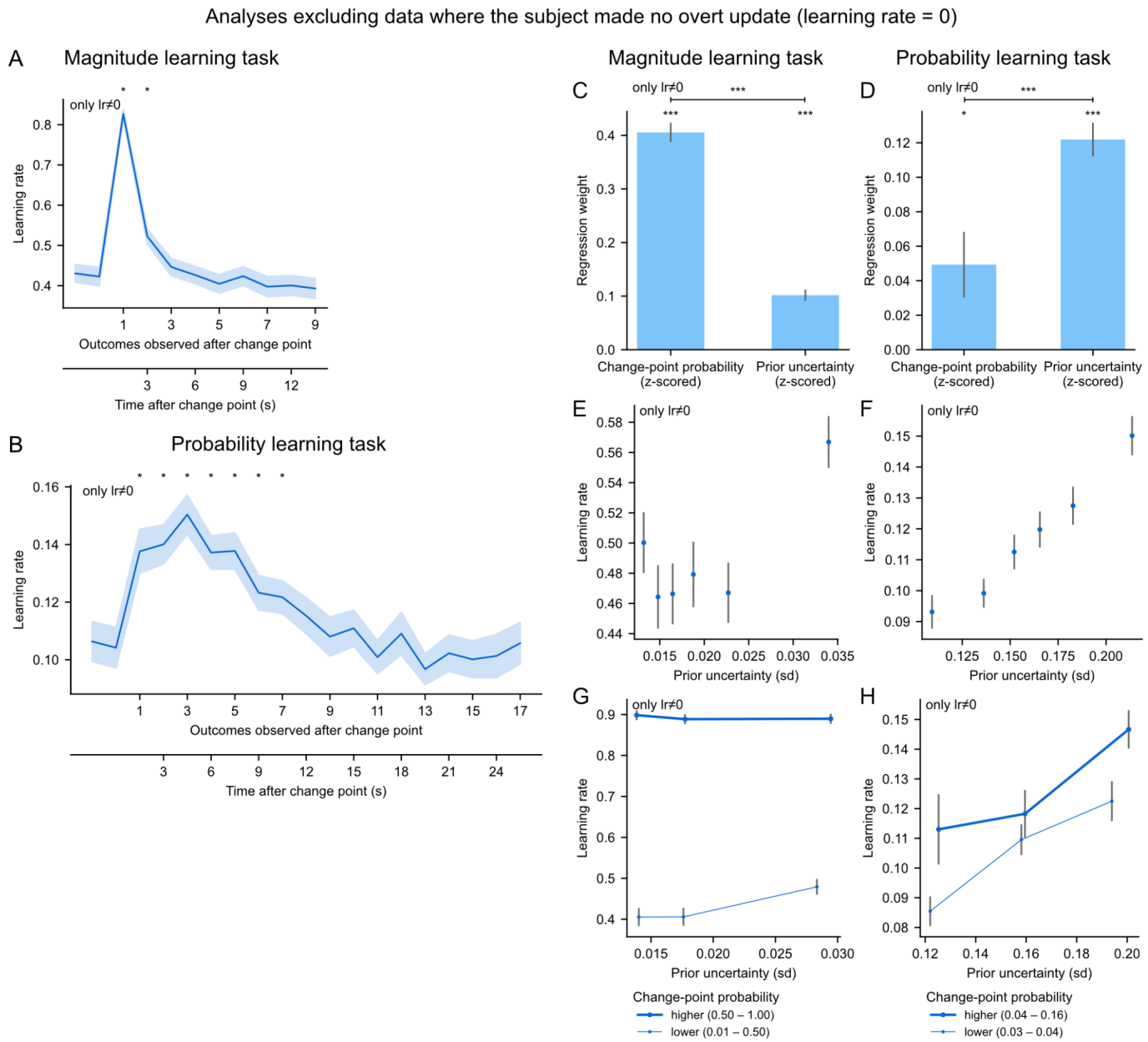

Fig. S4. The results are replicated when excluding all data where the subject did not make an overt update. After excluding all data points where this was the case (i.e. learning rate = 0), we performed the same analyses as in previous figures and obtained the above plots: (A and B) Equivalent to Fig. 2 A and B; (C–H) Equivalent to Fig. 4 (A–F).