Neural mechanisms behind identification of leptokurtic noise and adaptive behavioral response

Supplementary Information

Mathieu d'Acremont and Peter Bossaerts

Contents

1 Behavioral Results

1.1 Prolonged Effect of Outliers

To estimate the duration of the outlier effect, we conducted additional analyses on the behavioral data. Data for outlier trials as well as for the 5 trials following an outlier were averaged across participants (trials labeled “0” to “5” in Fig. S1). Data in all other trials were averaged to form the baseline (trial labeled “-1”). If a second outlier occurred before 5 trials elapsed, the trial counter was reset to 0. Thus only the trial labeled “0” contains outliers. The Bayesian solution is also shown for the learning rate.

Participants followed the Bayesian prescription by increasing the learning rate in response to outliers in the fundamental treatment and by reducing it in the transitory treatment (Fig. S1a). However, the positive and negative adjustments were less pronounced when compared to the Bayesian solution. Thus participants were “conservative” when reacting to outliers. Setting aside the outlier trials, one notices that the learning rate was always lower than the Bayesian solution. Thus participants were also “conservative” when reacting to the target movements. Concerning the duration of the change following outliers, the effect became relatively small two trials after it occurred. For the prediction error, the effect progressively diminished until the fifth trial (Fig. S1b). The deliberation time returned more abruptly to the baseline (and the outlier effect was delayed in the transitory condition, Fig. S1c).

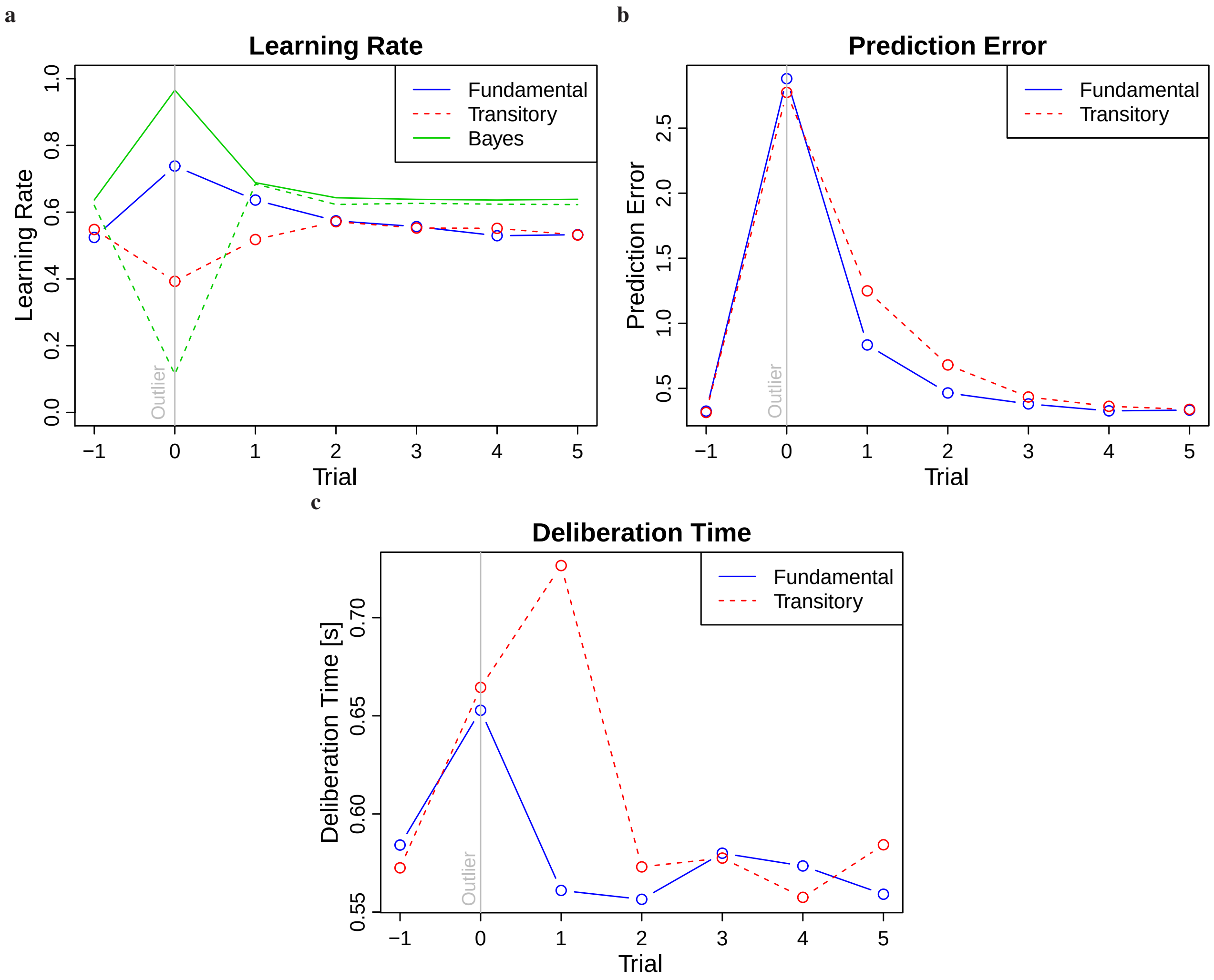

Figure S1: Learning rate, prediction error and deliberation time in the fundamental (red) and transitory (blue) treatment. The Bayesian solution is shown for the learning rate (green). The vertical gray line indicates the occurrence of an outlier.

1.2 Individual Differences and Training Effect

Mixed linear models were fitted to the behavioral data to predict the learning rate as a function of the trial type (Regular, Outlier, or Post-Outlier/Reversal, see manuscript for a definition). The contrast “Outlier minus Regular” measured the impact of outliers on the learning rate (Table S1 and S3). Adding the random effects to the fixed effects for this contrast allowed us to assess individual differences (Fig. S2a). After observing an outlier, almost all participants increased the learning rate in the fundamental treatment (28 blue points above the gray line) and decreased it in the transitory treatment (28 red points bellow the gray line). Thus most of the 31 participants learned to adjust their learning rate in the correct direction.

A factor with a level for each block of 40 trials was added to the mixed linear model to assess how fast participants learned to adjust the learning rate. The regression coefficients for the contrast Outlier-Regular are displayed for each block of 40 trials and each treatment in Figure S2b. In both treatments the impact of outlier on the learning rate was positive in the first 40 trials (however the effect was not significant, $p = .23$ and $p = .14$ for the fundamental and transitory treatments respectively). This trend supports the hypothesis that – a priori – people anticipate fundamental changes. In the transitory treatment, the positive impact of outliers increased rapidly after the first 40 trials. In the transitory treatment, the impact became negative with training, but the change was slower compared to the fundamental treatment. All contrasts after the first 40 trial block were significant ($p < .01$).

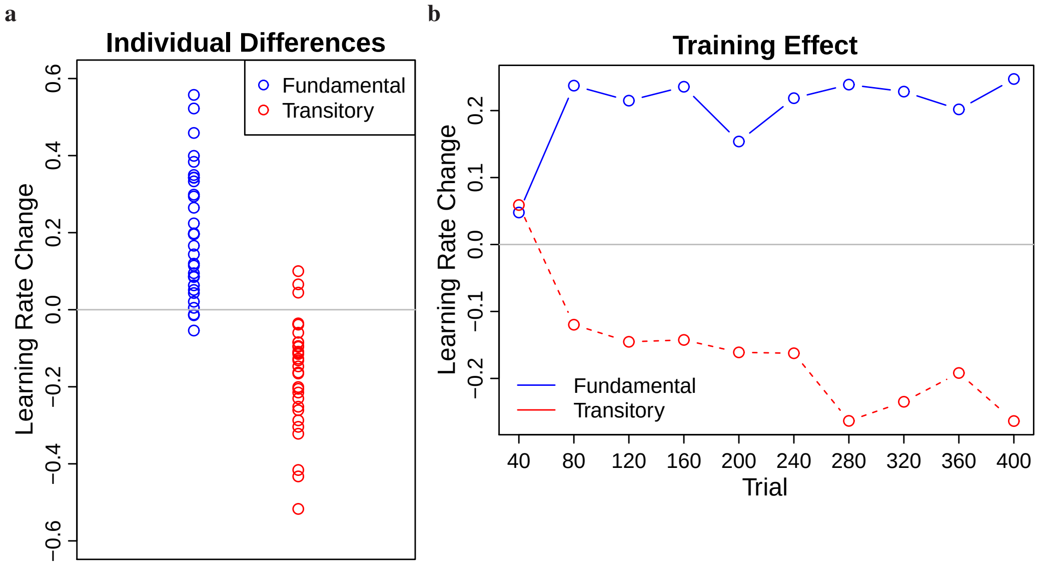

Figure S2: (a) Individual differences for the impact of outliers on the learning rate. (b) Progression of the learning rate across blocks of 40 trials.

2 Ideal-Observer Model

2.1 What Does The Ideal Observer Know?

A Bayesian model was developed to calculate the learning rate an agent should adopt to optimally respond to outliers in the fundamental and transitory treatments. We opted to approximate the infinite-range and continuous-state nature of the task with a finite-range, discrete-state, Markov state transition model. This simplified computations of the Bayesian updating: as will become clear later, Bayesian posterior and probabilities could be obtained from straightforward matrix multiplications.

In deriving optimal learning rates (i.e., how much the decision maker is to catch up with the target after a target move), and consistent with the literature on Kalman filtering, we assumed that the decision maker knew the parameters of the processes, and whether he/she was in the fundamental or transitory treatment. Later, this last assumption will be relaxed for the Bayesian model. A reinforcement learning model will also be presented for which no knowledge about the process parameters or the treatment is necessary.

2.2 Finding the Position of the Hidden Target

First, because our state-space model potentially (though rarely) generates target moves of more than one circle (more than 2π radians), we started from a state space that covered 4 circles. That is, the state space ran from 0 to 8π radians. Second, we discretized the state space in 2880 segments of $8\pi/2880$ radians each. Moves could be clockwise (increase in radians) or counter-clockwise (decrease in radians).

We then determined the transition probability matrix $\mathbf{P}$ of the hidden target X. $\mathbf{P}[i,j]$ gives the probability that the target will move to segment i at time $t+1$ knowing that it was at segment j at time t. Figure S3a shows a subset of the transition probability matrix for each treatment. $\mathbf{P}$ is calculated with the Algorithm 1. k serves to formalize the possibility that the target can reach the new segment directly or after several rotations on the hyper-circle (up to 3 rotations). In the fundamental treatment, $\sigma_1 = 0.25$, $\sigma_2 = 2$ and $p_2 = 0.15$. Thus the hidden target is subject to rare but large movements. In the transitory treatment, $\sigma_1 = \sigma_2 = 0.25$ and $p_2 = 0$. Thus the target movements are Gaussian.

Algorithm 1 Transition Probabilities

```
for $j = 1 : 2880$ do // Go through all departure segments
    for $i = 1 : 2880$ do // Go through all arrival segments
        Initialize the transition probabilities: $\mathbf{P}[i,j] = 0$
        for $k = -3, -2, -1, 0, 1, 2, 3$ do // The distance can be larger than 1 hyper-circle
            Calculate the movement size: $d = \frac{8\pi}{2880}(i-j) + 8\pi k$
            Density of the $1^{st}$ Gaussian at $d$: $f_1 = e^{\frac{-d^2}{2\sigma_1^2}} / \sqrt{2\pi\sigma_1^2}$
            Density of the $2^{nd}$ Gaussian at $d$: $f_2 = e^{\frac{-d^2}{2\sigma_2^2}} / \sqrt{2\pi\sigma_2^2}$
            Transition probability (mixture): $\mathbf{P}[i,j] = \mathbf{P}[i,j] + (1-p_2)f_1 + p_2 f_2$
        end for
    end for
    Normalize: $\mathbf{P}[.,j] = \mathbf{P}[.,j] / \sum_i \mathbf{P}[i,j])$
end for
```

The likelihood matrix $\mathbf{L}$ defines the relation between the hidden target X and the observed target Y. $\mathbf{L}[i,j]$ gives the probability that X is at segment j knowing that Y is at segment i. Figure S3b shows a subset of the likelihood matrix for each treatment. $\mathbf{L}$ is calculated with Algorithm 2. In the fundamental treatment, $\sigma_1 = \sigma_2 = 0.25$ and $p_2 = 0$. This implies that the distance between the observed and the hidden target follows a Gaussian distribution. In the transitory treatment, $\sigma_1 = 0.25$, $\sigma_2 = 2$ and $p_2 = 0.15$. Thus the observed target is sometimes very far away from the hidden target.

Algorithm 3 calculates the posterior belief after observing the target movement. The vector $\mathbf{b}$ quantifies the belief on the hidden target position. $\mathbf{b}_t[i]$ gives the probability that X is at segment i at time t. The posterior belief is stored in $\mathbf{b}'_t$. y_t is the segment at which Y was observed at time t. $\mathbf{l}[i]$ with $\mathbf{l} = \mathbf{L}[y_t, .]$ gives the probability of observing Y at the segment y_t

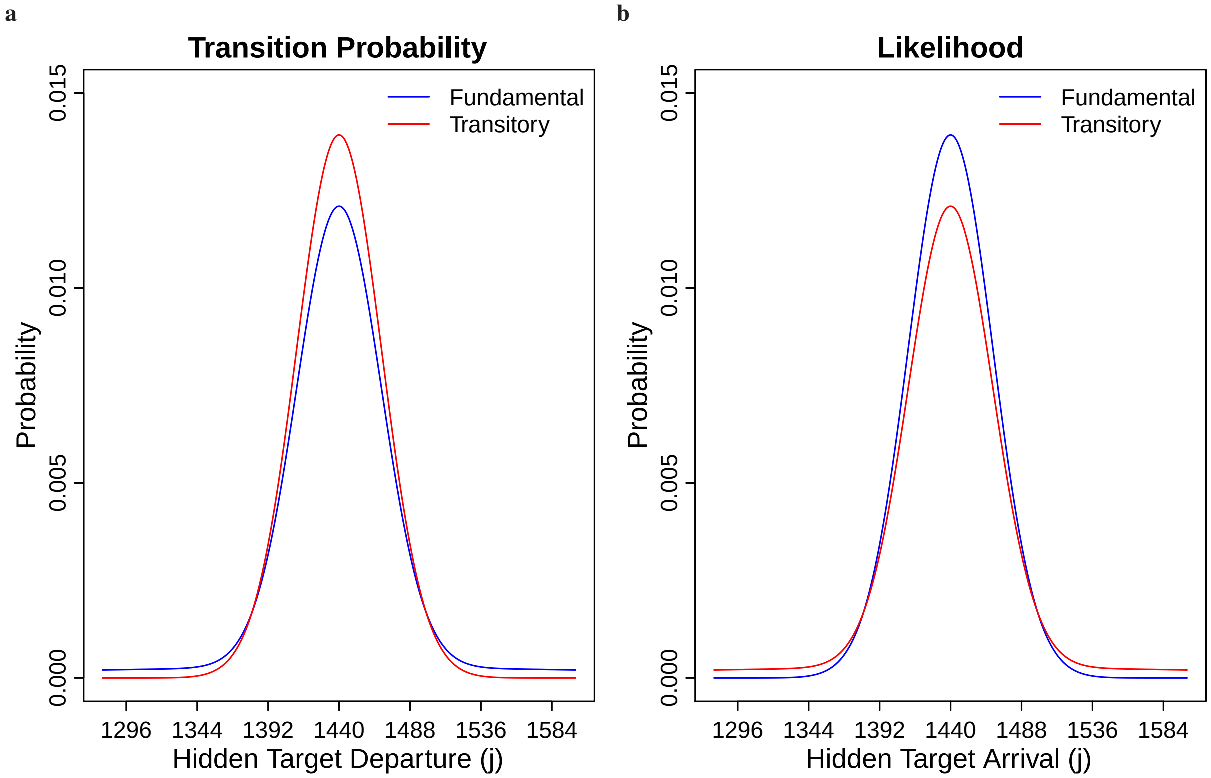

Figure S3: (a) Transition probabilities. Given a hidden target positioned at segment $i = 1440$, the plot shows the probability that the hidden target was at segment j in the previous time step. Large movement of the hidden target are plausible in the fundamental treatment but almost impossible in the transitory condition. (b) Likelihood. Given an observed target at segment $i = 1440$, the plot shows the probability that the hidden target is currently at segment j. Large distance between the hidden and observed target are plausible in the transitory treatment, but almost impossible in the fundamental treatment.

knowing that the hidden target X is at segment i. c_t is the probability of observing Y at the segment y_t regardless of where X is hidden.

The next algorithm uses the posterior belief to estimate the uncertainty of the belief (entropy) and the expected position of the hidden target (Alg. 4). This position is an unsigned number ranging from 1 to 2880 and points to a unique segment on the hyper-circle. This number needs to be converted into a signed value with no boundaries. Indeed, the process controlling the position of the hidden target is a random walk and produces values that are unbounded. The algorithm takes care of converting the unsigned position x_t $(\in 1,\ldots,2880)$ into a signed position x^s_t $(\in \mathbb{Z})$.

When the change in position from $t-1$ to t is very high ($>$ 1203 segments), it indicates that the hidden target has moved counter-clockwise through the junction of the hyper-circle. This junction directly connects segment 2880 to segment 1 (closing of the hyper-circle). This would happen for instance if X moved from segment 1 to segment 2880 and thus produced a displacement d of 2879. Simulation shows that positive differences > 1203 are impossible and d is thus converted to an innovation term v equals to -1 segment (counter-clockwise movement). An opposite example is if X moved from segment 2880 to segment 1 and thus produced a displacement d of -2879. Such a low negative difference is impossible and is thus converted to an innovation v equals to 1 segment (clockwise movement). The innovation v is added to the previous signed position of the hidden target x^s_{t-1} to obtain the new signed position x^s_t (signed index of a segment). This strategy allows us to recover the position of the hidden target on an unbounded space. It requires a discrete sate space that covers enough radians, hence the use of a hyper-cycle of 8π. Otherwise it is impossible to decide if X moved through the junction or not.

In the decision making task program, the current position of the observed target was given in radians and is called y^r_t in the algorithm. It is transformed into y^s_t, the signed index of a segment. This value can be compared to the signed index of the previous hidden target position x^s_{t-1} in order to calculate the prediction error δ. The optimal (Bayesian) learning rate r is calculated by dividing the innovation v by the prediction error. This division leads to infinite or undefined values

Algorithm 2 Likelihood

```
for $j = 1 : 2880$ do // Go through all hidden target segments
    for $i = 1 : 2880$ do // Go through all observed target segments
        Initialize the likelihood: $\mathbf{L}[i,j] = 0$
        for $k = -3,-2,-1,0,1,2,3$ do // The distance can be larger than 1 hyper-circle
            Calculate the target distance: $d = \frac{8\pi}{2880}(i-j) + 8\pi k$
            Density of the $1^{st}$ Gaussian at $d$: $f_1 = e^{\frac{-d^2}{2\sigma_1^2}} / \sqrt{2\pi\sigma_1^2}$
            Density of the $2^{nd}$ Gaussian at $d$: $f_2 = e^{\frac{-d^2}{2\sigma_2^2}} / \sqrt{2\pi\sigma_2^2}$
            Likelihood (mixture): $\mathbf{L}[i,j] = \mathbf{L}[i,j] + (1-p_2)f_1 + p_2 f_2$
        end for
    end for
    Normalize: $\mathbf{L}[.,j] = \mathbf{L}[.,j] / \sum_i \mathbf{L}[i,j])$
end for
```

Algorithm 3 Bayesian Inference

```
Set the initial belief: $\mathbf{b}'_0[i] = 1/2880; \quad i = 1,\ldots,2880$
for $t = 1 : 200$ do // Go through all time steps in the run
    Compute the prior belief using the transition probabilities: $\mathbf{b}_t = \mathbf{P}\mathbf{b}'_{t-1}$
    Normalize the prior belief: $\mathbf{b}_t[i] = \mathbf{b}_t[i] / \sum_j \mathbf{b}_t[j]; \quad i = 1,\ldots,2880$
    Likelihood of seeing Y: $\mathbf{l} = \mathbf{L}[y_t, .]$
    Compute the posterior belief: $\mathbf{b}'_t[i] = \mathbf{l}[i]\mathbf{b}[i]; \quad i = 1,\ldots,2880$
    Probability of seeing Y: $c_t = \sum_j \mathbf{b}'_t[j]$
    Normalize the posterior belief: $\mathbf{b}'_t[i] = \mathbf{b}'_t[i]/c_t; \quad i = 1,\ldots,2880$
end for
```

Algorithm 4 Prediction Error and Learning Rate

```
Set the initial position of the hidden target: $x_0 = 1$
Set the signed position of the hidden target: $x_0^s = 1$
for $t = 1 : 200$ do // Go through all time steps in the run
    Expected position on the hyper-circle: $x_t = \arg\max_i \mathbf{b}'_t[i]$
    Calculate the entropy: $h_t = h(\mathbf{b}'_t)$
    Change in position: $d = x_t - x_{t-1}$
    if $d > 1203$ then // The change in position is too high, it must be counter-clockwise
        Calculate the signed innovation: $v = -(2880 - d)$
    else if $d < -1203$ then // The change in position is too low, it must be clockwise
        Calculate the signed innovation: $v = 2880 + d$
    else // The change in position is possible
        Calculate the signed innovation: $v = d$
    end if
    New signed position: $x_t^s = x_{t-1}^s + v$
    Signed position of the observed target (integer division): $y_t^s = 1 + y_t^r \setminus (8\pi/2880)$
    Prediction error: $\delta_t = y_t^s - x_{t-1}^s$
    Learning rate: $r_t = v/\delta_t$
end for
```

when the prediction error equals 0 segment. This limit is inherent to the finite state space. Infinite and undefined values were replaced by missing values (NA). The Bayesian learning rate was missing for less than 1% of the trials. The optimal learning rate for each trial type was calculated by averaging the learning rate r obtained for all the Regular, Outlier, and Post-Outlier/Reversal movements of the observed target y.

It might be worth stressing here that the prediction errors and the learning rates do not drive the learning process in the Bayesian model. These variables are byproducts of the probabilistic inference and are calculated to compare the Bayesian model with the behavioral data and the reinforcement learning model developed later in the SI.

The accuracy of the Bayesian inference is illustrated in Figure S4. In the fundamental treatment (Fig. S4a), large movements of the observed target (red) are associated with large movements of the hidden target (green). In the transitory treatment (Fig. S4b), large movements of the observed target (red) are due to errors and are not associated to underlying movements of the hidden target (green). It can be seen that the inferred position (blue) tracks the position of the hidden target. It follows the large movements of the observed target in the fundamental treatment and successfully ignores them in the transitory treatment.

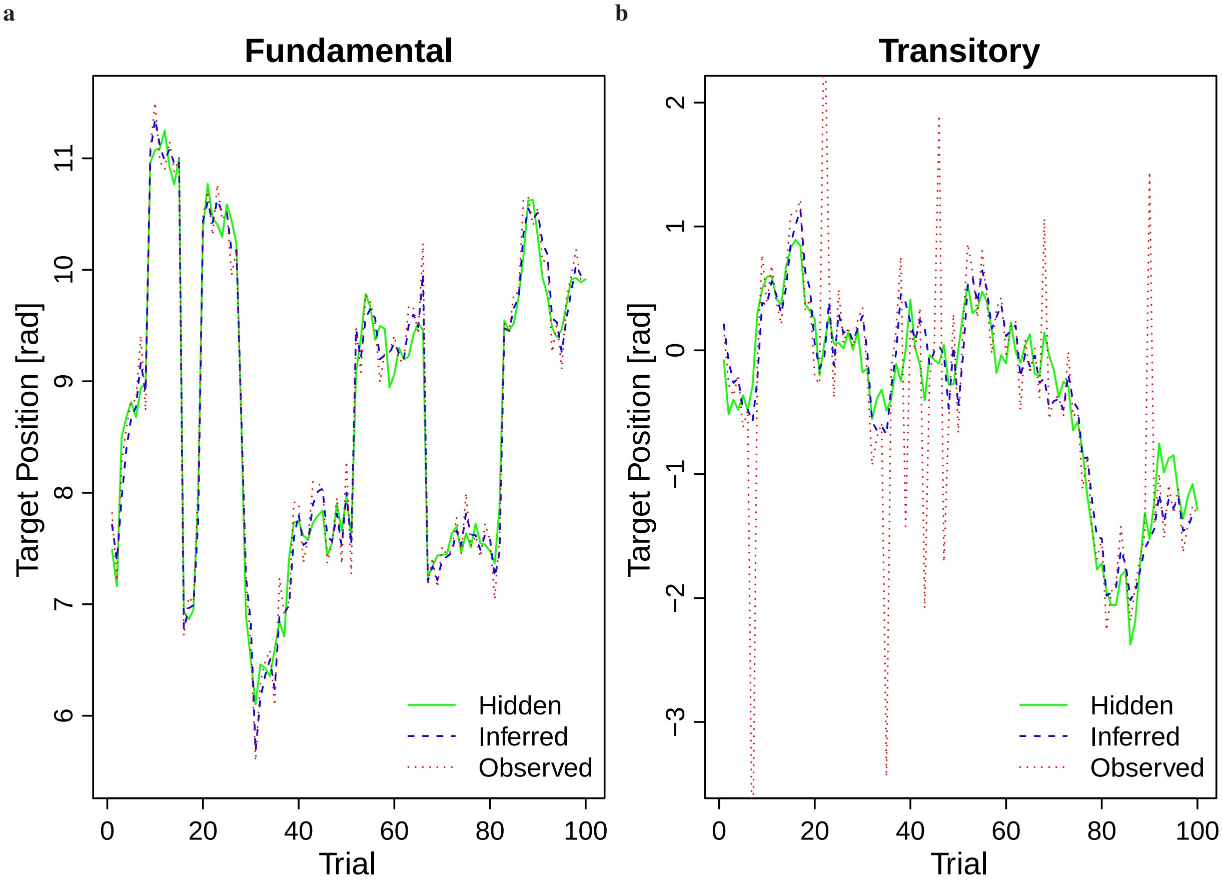

Figure S4: Ability of the Bayesian model to track the position of the hidden target. (a) Fundamental treatment. (b) Transitory treatment.

2.3 Relation between Prediction Error, Learning Rate, and Uncertainty

Figure S5 shows how the learning rate increases with the prediction error, consistent with the idea that large target moves (outliers) are most likely to be caused by the leptokurtic driving process of the state variable, and hence, demand substantial belief adjustment. The uncertainty in the origin of the target moves (noise; state changes) translates into corresponding entropy of the posterior belief. Entropy is lower for tiny and large target moves. This also implies that learning is most intense after medium-sized target moves. Results of the Bayesian inference were smoothed to obtain the figure.

In contrast Figure S6 shows how the learning rate decreases with the prediction error in the transitory treatment, consistent

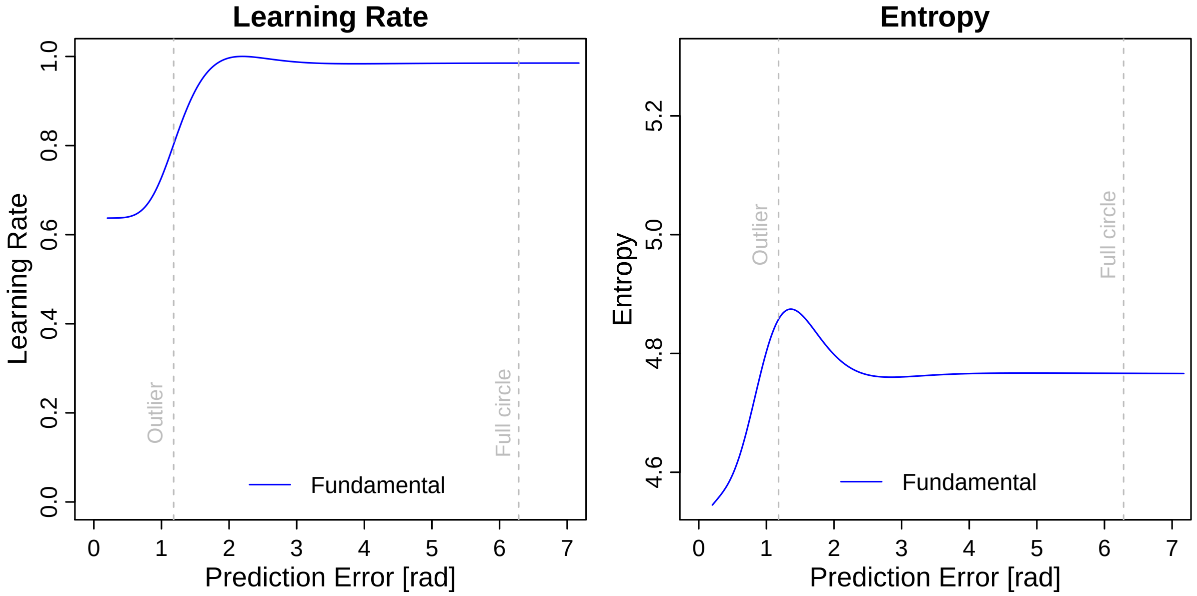

Figure S5: Fundamental Treatment. (a) Learning rate as a function of the prediction error (difference between the observed target location and the prior inferred position). (b) Entropy (inverse of precision) of posterior beliefs of the hidden target location as a function of prediction error.

with the idea that large target moves (outliers) are most likely to be caused by leptokurtic noise, and hence, require little belief adjustment. There is a slight non-monotonicity: tiny target moves are more likely to be generated by leptokurtic noise as well (the noise distribution exhibits both fat tails and more peakedness than the Gaussian distribution). Consequently, the learning rate actually decreases slightly as the prediction error becomes tiny. The uncertainty in the origin of the target moves (noise; state changes) translates into corresponding posterior belief entropy. Entropy is lower for tiny and large target moves. This also implies that learning is most intense after medium-sized target moves (near the outlier frontier).

2.4 Knowing Which Regime One Is In

In the ideal-observe model presented in the previous section, we made the assumption that the Bayesian agent knew which treatment he/she was in. This simplification is acceptable only if the agent can discover rapidly the treatment currently in place. To test this assumption, we looked at the evolution of beliefs at the beginning of a new block of trials. In this model, the Bayesian agent knows the distributions that generate the observed target in each treatment (like in the ideal-observer model), but starts with a neutral prior regarding the treatment at play. At any given trial, the likelihood to see y_t is given by c_t (see Alg. 3). This likelihood can be computed under the hypothesis of being in the fundamental (c_t^F) or transitory treatment (c_t^T). The posterior probability of being in the transitory treatment p_t^T is given by the Bayes law:

$$p_t^T = \frac{c_t^T p_{t-1}^T}{c_t^T p_{t-1}^T + c_t^F (1 - p_{t-1}^T)} \tag{1}$$

Before the first observation, the initial probability of being in the transitory treatment is set at $p_0^T = 0.5$. Figure S7 shows the evolution of beliefs in each treatment. The belief changes abruptly after the observation of an outlier (vertical gray line) and gets close to 0 in the fundamental treatment and close to 1 in the transitory treatment. It can be concluded that the Bayesian agent discovers in which treatment he/she is immediately after the occurrence of an outlier.

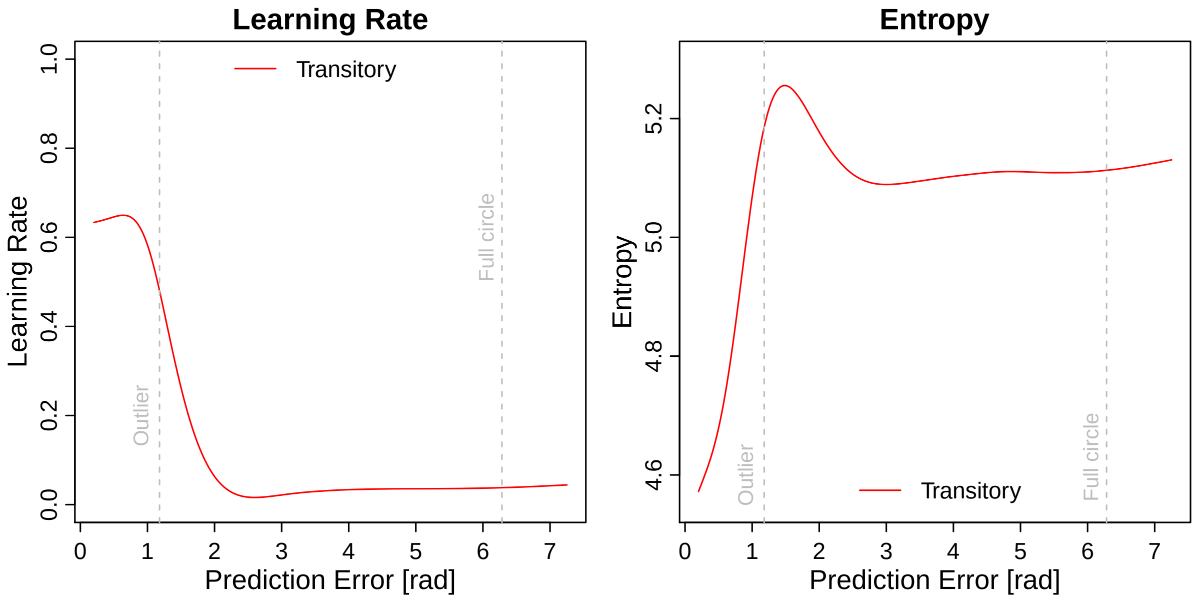

Figure S6: Transitory Treatment. (a) Learning rate as a function of the prediction error (difference between the observed target location and the prior inferred position). (b) Entropy (inverse of precision) of posterior beliefs of the hidden target location as a function of prediction error.

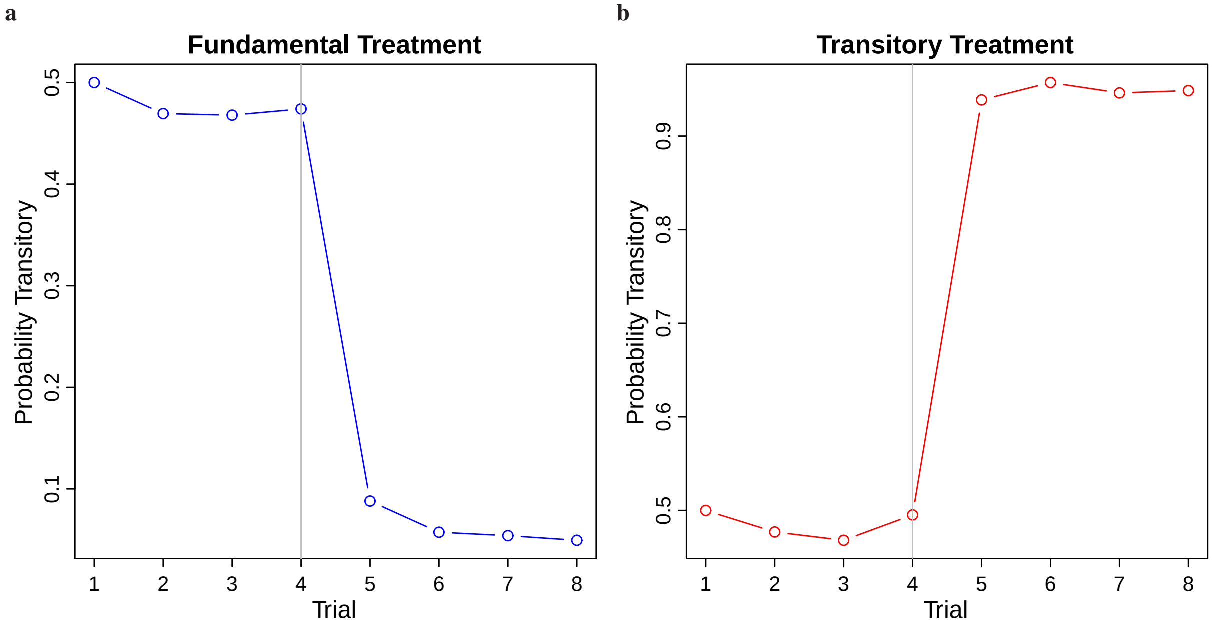

Figure S7: Evolution of the belief of being in the transitory treatment across the first 8 trials of a block. The vertical gray line corresponds to the occurrence of an outlier. (a) The Bayesian agent is in the fundamental treatment. (b) The Bayesian agent is in the transitory treatment.

3 Reinforcement Learning

3.1 Model Description

The Bayesian model was necessary to discover the optimal reaction to outliers in each treatment. However, this model requires full knowledge of the distributions used to generate the target moves. This is rarely the case in a natural environment. The Bayesian model is also demanding in terms of computation. Given the limits of human cognition, participants might rely on a simpler model to learn the correct learning rate (meta learning). We thus developed a reinforcement learning model that is cognitively more realistic. We call it "Contrarian Reinforcement Learning" (Contrarian RL) because the learning rate can decrease when the prediction error increases.

The central idea of this new model is to take advantage of the prediction error autocorrelation to control the reaction to outliers. Adjustment of the learning rate as a function of autocorrelation is an attempt to make beliefs form a martingale because Bayesian beliefs do so (Doob, 1948). Hence, the contrarian RL model is an attempt to emulate Bayesian learning. The principle of using the autocorrelation has been presented before to accelerate learning (Sutton, 1992). However, in our model, the autocorrelation does not directly influence the learning rate, it influences the relation between the learning rate and the prediction error. This gives our new algorithm the unique ability to momentarily inhibit a response when the change in the environment is too drastic (e.g a sudden sell-off in the stock market).

The autocorrelation carries diagnostic information on how to react to large changes. Indeed, if the autocorrelation is positive, this means the agent under-reacts to outliers. In this case, the autocorrelation is used to set a positive relation between the prediction error and the learning rate (like in Fig. S5). If the prediction error is large, the learning rate will increase in that particular trial and the agent will follow the outlier. This is the correct strategy in the fundamental treatment. If the autocorrelation is negative, this means the agent over-reacts to outliers. In this case, the autocorrelation is used to set a negative relation between the prediction error and the learning rate (like in Fig. S6). If the prediction error is large, the learning rate will decrease in that particular trial and the agent will resist the change (contrarian attitude). This is the correct strategy if large moves reflect observational errors like in the transitory condition. The detailed Contrarian RL algorithm is presented first (Alg. 5) and its various components are then reviewed.

Algorithm 5 Contrarian Reinforcement Learning

```
Set the free parameters $\theta, a, f, s$
Initialize the inferred position of the target: $z = 0$
Initialize the prediction error autocovariance: $c = 0$
Initialize the past prediction error: $\delta_0 = 0$
for $t = 1 : 2000$ do // Go through all time steps
    Calculate the new prediction error: $\delta = y_t - z$
    if $t = 1$ then // First time step
        Initialize the prediction error variance: $v = \delta^2$
    end if
    Update the prediction error variance: $v = v + \theta(\delta^2 - v)$
    Update the prediction error autocovariance: $c = c + \theta\delta\delta_0$
    Get the scaled-unsigned prediction error: $\delta^+ = |\delta|/\sqrt{v}$
    Get the prediction error autocorrelation: $r = c/v$
    if $\delta^+ < f$ then // The prediction error is within the outlier frontier
        The learning rate is constant: $\alpha = a$
    else // The prediction error is large
        Adjust the learning rate as a function of the autocorrelation and error: $\alpha = a + sr(\delta^+ - f)$
    end if
    Bound the learning rate in $[0, 1]$: $\alpha = \min(\max(0, \alpha), 1)$
    Update the inferred position of the target: $z = z + \alpha\delta$
    Record the current prediction error: $\delta_0 = \delta$
end for
```

z is the inferred position of the hidden target and y is the observed target position (in radians). The key variable in the algorithm is the learning rate α that determines the reaction to the target movement on a trial to trial basis. To adjust this variable, the algorithm needs to estimate two quantities: The prediction error variance v and its covariance c. The meta learning rate θ is a free parameter that controls the estimation of the prediction error variance and covariance. Dividing the covariance and the prediction error by the variance makes the model immune to an arbitrary change of units. Dividing by the variance, one obtains the autocorrelation r and the unsigned-scaled prediction error δ^+. A second free parameter f defines the outlier frontier. If the prediction error is smaller than this frontier, the learning rate equals the value of a third free parameter a. That is, for small target movements, the learning rate remains constant. If the prediction error is larger than the frontier, the learning rate becomes a function of the (scale-unsigned) prediction error: It increases with the prediction error if the autocorrelation is positive (activation) and decreases if the autocorrelation is negative (inhibition). The strength of this relationship (not its direction) is controlled by a fourth free parameter s.

In contrast with the Bayesian model, the learning rate and prediction error in the Contrarian RL are not byproducts because they drive the learning process. Another characteristic of the algorithm is that it learns to follow or resist outliers using the same set of free parameters. Thus it does not need to know in which treatment it is operating. It also learns without knowing the full distributions generating the stimuli. Instead it uses the prediction error autocorrelation. It is important to note that it is not the number of free parameters that gives the Contrarian RL the flexibility to momentarily activate or inhibit a response, but the way the autocorrelation is factored in the algorithm.

3.2 Model Performance

The performance of the Contrarian RL was tested against three models: a) the benchmark Bayesian model (Ideal-Observer Model, see Section 2), b) the reinforcement learning model written by Sutton (1992), and c) a simple update rule. The Bayesian model has no free parameter but it requires full knowledge of the generating distributions (and in which regime it is). The model by Sutton (1992) has one meta learning rate θ and the learning rate depends on the correlation between present and past adjustments. Finally, the simple update rule has one free parameter, a constant learning rate α:

$$z \leftarrow z + \alpha\delta \tag{2}$$

To estimate the free parameters, the error was measured by the distance between the inferred (z) and the true hidden target position (z^*). We chose the RMSE for the loss function. The hidden target position was known because we used stimulated target movements. The loss function was minimized with constrained optimization in R. For the new contrarian model, the constrains were $\theta > 0$, $f > 0$, $0 < a < 1$, and $s > 0$. For Sutton's model, the constrain was $\theta > 0$. For the simple update rule, it was $\alpha > 0$. All models were fitted on 2000 trials per treatment.

The estimated value for the constant learning rate a in the Contrarian RL model was 0.57. The outlier frontier f was 0.91, a value close to one standard deviation. Beyond this value, the learning rate α depends on the product between the (unsigned-scaled) prediction error and the autocorrelation. The value of s was 0.57, meaning that this product needs to be scaled. The non-zero value also suggests that this product improves the model. The meta learning rate for the variance and covariance was $\theta = 0.012$. Results also show that the learning rate $alpha$ increases in response to outliers in the fundamental treatment and decreases in the transitory treatment (purple lines, Fig. S8a). This pattern follows the optimal Bayesian solution (green line).

The meta learning rate in Sutton's algorithm was $\theta = 0.023$. Results show that the learning rate α is high in the fundamental treatment and low in the transitory treatment (orange lines, Fig. S8a). Unlike the Contrarian and the Bayesian models, the learning rate remains unchanged when an outlier occurs. The learning rate for the simple update rule was estimated at 0.61 and remained constant across trials (black dashed line, Fig. S8a). The researcher might have no access to the position of the hidden target. However, similar estimates were found after minimizing the prediction error. The prediction error can always be computed because it is based on the inferred position and not the true position of the hidden target.

For the four models we found that the RMSE was larger in the transitory compared to the fundamental treatment (Fig. S8b). This result is in line with the larger errors made by participant in the former condition (see behavioral data in Fig. S1b). Joining the two treatments, the RMSE was smallest for the Bayesian model (0.23), followed by the Contrarian RL (0.28), Sutton's model (0.36), and the simple update rule (0.47). To take into account the number of free parameters in each model, the BIC was calculated based on the MSE (a lower BIC is better). The BIC was -11591 for the optimal Bayesian model, -10256 for the Contrarian RL, -8143 for Sutton's model, and -6017 for the simple update rule. These results show that the novel RL model performs better in the fundamental and transitory treatment when compared to the

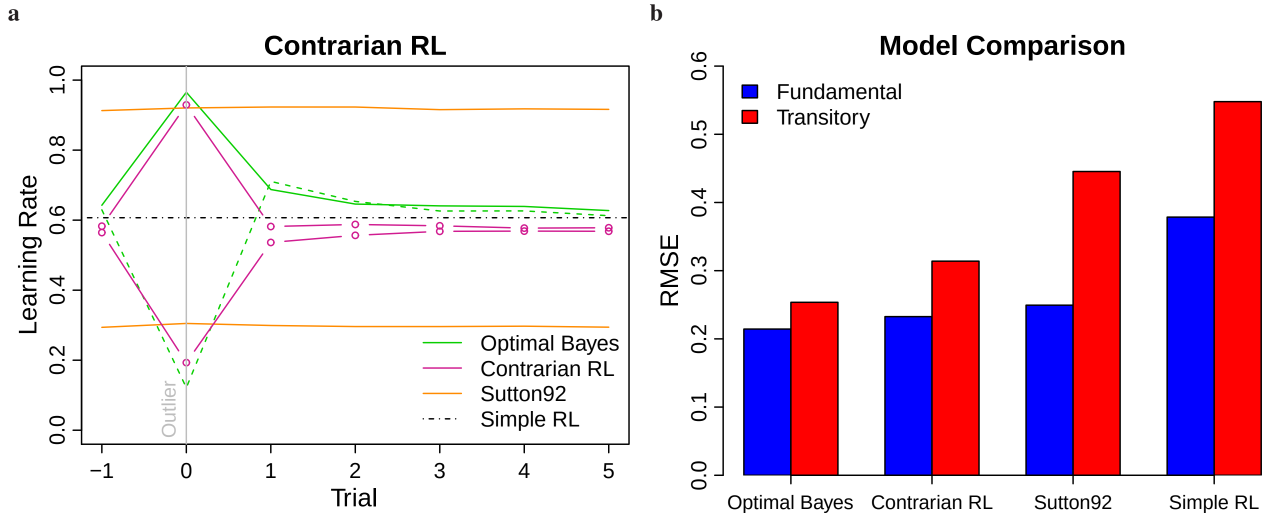

Figure S8: The RL models parameters were estimated by minimizing the distance between the inferred and true target position. The Bayesian model has no free parameter to estimate. (a) Adjustment of the learning rate in response to outliers in the fundamental (upper colored lines) and transitory treatments (lower colored lines). The Simple RL model has the same learning rate in the two treatments (black dashed line). (b) RMSE for each model in the fundamental and transitory treatments. Lower is better.

other RL algorithms.

4 Regressions

Mixed linear models were defined with the learning rate (Tables S1, S3), the unsigned prediction error (Tables S8, S9), or the deliberation time (Tables S16, S17) as the dependent variable. Regressions were estimated separately for the fundamental and transitory treatment. On the right-hand-side of the regression, a single factor was entered to label the type of trials. In the fundamental treatment, the labels were: Regular, Outlier, and Post-Outlier. In the transitory treatment, the labels were Regular, Outlier, and Reversal. To estimate the difference between the types of trials, the contrasts **Outlier-Regular** (i.e., Outlier minus Regular) and **PostOutlier-Outlier** were defined for the fundamental treatment. The contrasts **Outlier-Regular** and **Reversal-Outlier** were defined for the transitory treatment. Thus the intercept corresponds to the **Outlier** trials. Random effects capturing subject variability were introduced for both the intercept and the factor.

To compare the participant learning rate to the optimal Bayesian solution and assess the relative impact of outliers, additional mixed linear models were defined. A regression was estimated separately for the fundamental and transitory treatment (Tables S2, S4). Post-Outlier and Reversal trials were removed from the analysis. Two measurements were taken on each participant and at each trial: its own learning rate and the learning rate calculated with the Bayesian model (Alg. 4). The dependent variable is thus the learning rate. The dummy variable **Participant** was entered on the right-hand-side to indicate if the measurement was the Bayesian learning rate or the participant one. The dummy variable **Outlier** indicated if the type of trial was Regular or Outlier. The interaction **Outlier x Participant** will show if the impact of outliers on the learning rate is smaller or larger for participants in comparison to the Bayesian solution. Random effects capturing participant variability were introduced for both the intercept and the two dummy variables.

To compare the prediction error to the optimal Bayesian solution, data from the fundamental and transitory treatment were joined and than split by the type of trial (Regular, Outlier, Post-Outlier or Reversal). A separate regression was estimated for each type of trial (Tables S10, S11, S12). The dependent variable was the difference *AbsPeObs* minus *AbsPeBayes*. *AbsPeObs* is the unsigned participant prediction error. *AbsPeBayes* is the unsigned Bayesian prediction error. On the right-hand-side of the regression, a dummy variable was entered to label the **Treatment**. A significant and positive effect of the dummy variable will indicate that the prediction error (relative to the Bayesian prediction error) is larger in the transitory condition. Random effects due to subject variability were introduced for the intercept and the dummy variable.

To test the effect of training on behavioral adjustment, mixed linear models were defined with the learning rate as the dependent variable. Separate models were estimated for the fundamental (Table S5) and transitory treatments (Table S6) and only outlier trials were included. On the right-hand-side of the regressions, a dummy variable **Run** was created to indicate if the (outlier) trial belonged to the first or second run. Random effects due to subject variability were introduced for the intercept and the dummy variable. Data from the fundamental and transitory treatments were then joined. Here again only outlier trials were included in the analysis. Dummy variables coding for the **Treatment** and the **Run** were entered on the right-hand-side of the regression (Table S7). The interaction **Run x Treatment** was added. The interaction will test if the effect of the training is more pronounced in one treatment compared to the other. Random effects due to subject variability were introduced for the intercept, the main effects, and the interaction effect.

To test the effect of training on performance, mixed linear models were defined with the unsigned prediction error as the dependent variable. Separate models were estimated for the fundamental (Table S13) and transitory treatments (Table S14) and only Post-Outlier or Reversal trials were included. On the right-hand-side of the regressions, a dummy variable **Run** was created to indicate if the (outlier) trial belonged to the first or second run. Random effects due to subject variability were introduced for the intercept and the dummy variable. Data from the fundamental and transitory treatments were then joined. Here again only Post-Outlier or Reversal trials were included in the analysis. Dummy variables coding the **Treatment** and **Run** were entered on the right-hand-side of the regression (Table S15). The interaction **Run x Treatment** was added. The interaction will test if the effect of the training is more pronounced in one treatment compared to the other. Random effects due to subject variability were introduced for the intercept, the main effects, and the interaction effect.

To test the effect of training on speed, mixed linear models were defined with the deliberation time as the dependent variable. When participants do not change the learning rate, the value of the dependent variable is missing (NA). Separate models were estimated for the fundamental (Table S18) and transitory treatments (Table S19). All trials were included. On the right-hand-side of the regressions, a dummy variable **Run** was created to indicate if the trial belonged to the first or second run. Random effects due to subject variability were introduced for the intercept and the dummy variable.

5 Tables

Table S1: Learning rate in the fundamental treatment

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
Outlier	0.735∗∗∗	0.684	0.787	0.026	12367	28.01	0.000
Outlier-Regular	0.202∗∗∗	0.141	0.264	0.032	12367	6.42	0.000
PostOutlier-Outlier	−0.099∗∗	−0.175	−0.024	0.038	12367	−2.59	0.010
Random effect (SD)							
Outlier	0.140	–	–	–	–	–	–
Outlier-Regular	0.170	–	–	–	–	–	–
PostOutlier-Outlier	0.205	–	–	–	–	–	–
Error	0.244	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=12400.

Table S2: Learning rate in the fundamental treatment (relative to Bayes)

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
(Intercept)	0.642∗∗∗	0.639	0.646	0.002	24647	370.23	0.000
Outlier	0.324∗∗∗	0.312	0.335	0.006	24647	56.70	0.000
Participant	−0.100∗∗∗	−0.152	−0.048	0.026	24647	−3.77	0.000
OutlierxParticipant	−0.131∗∗∗	−0.199	−0.064	0.034	24647	−3.81	0.000
Random effect (SD)							
(Intercept)	0.000	–	–	–	–	–	–
Outlier	0.000	–	–	–	–	–	–
Participant	0.147	–	–	–	–	–	–
OutlierxParticipant	0.186	–	–	–	–	–	–
Error	0.183	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=24681; Outlier=0 for regular trials; Outlier=1 for outlier trials; Participant=0 for Bayesian prediction; Participant=1 for participants.

Table S3: Learning rate in the transitory treatment

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
Outlier	0.389∗∗∗	0.335	0.444	0.028	12367	14.10	0.000
Outlier-Regular	−0.162∗∗∗	−0.217	−0.108	0.028	12367	−5.83	0.000
Reversal-Outlier	0.111∗∗∗	0.067	0.156	0.023	12367	4.92	0.000
Random effect (SD)							
Outlier	0.147	–	–	–	–	–	–
Outlier-Regular	0.148	–	–	–	–	–	–
Reversal-Outlier	0.108	–	–	–	–	–	–
Error	0.257	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=12400.

Table S4: Learning rate in the transitory treatment (relative to Bayes)

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
(Intercept)	0.627∗∗∗	0.624	0.631	0.002	24660	329.06	0.000
Outlier	−0.513∗∗∗	−0.526	−0.500	0.007	24660	−77.95	0.000
Participant	−0.080∗∗	−0.133	−0.026	0.027	24660	−2.93	0.003
OutlierxParticipant	0.355∗∗∗	0.295	0.416	0.031	24660	11.50	0.000
Random effect (SD)							
(Intercept)	0.000	–	–	–	–	–	–
Outlier	0.000	–	–	–	–	–	–
Participant	0.151	–	–	–	–	–	–
OutlierxParticipant	0.164	–	–	–	–	–	–
Error	0.202	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=24694; Outlier=0 for regular trials; Outlier=1 for outlier trials; Participant=0 for Bayesian prediction; Participant=1 for participants.

Table S5: Learning rate as a function of training in the transitory treatment

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
(Intercept)	0.447***	0.386	0.509	0.031	1003	14.36	0.000
Run	−0.116***	−0.169	−0.064	0.027	1003	−4.37	0.000
Random effect (SD)							
(Intercept)	0.157	–	–	–	–	–	–
Run	0.106	–	–	–	–	–	–
Error	0.294	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=1035.

Table S6: Learning rate as a function of training in the fundamental treatment

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
(Intercept)	0.726***	0.669	0.783	0.029	1103	25.07	0.000
Run	0.017	−0.013	0.048	0.015	1103	1.13	0.260
Random effect (SD)							
(Intercept)	0.152	–	–	–	–	–	–
Run	0.038	–	–	–	–	–	–
Error	0.230	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=1135.

Table S7: Learning rate as a function of training and treatment (interaction)

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
(Intercept)	0.726***	0.671	0.782	0.028	2136	25.68	0.000
Run	0.017	−0.016	0.050	0.017	2136	1.00	0.319
Treatment	−0.279***	−0.357	−0.202	0.040	2136	−7.03	0.000
Run x Treatment	−0.133***	−0.194	−0.072	0.031	2136	−4.27	0.000
Random effect (SD)							
(Intercept)	0.145	–	–	–	–	–	–
Run	0.035	–	–	–	–	–	–
Treatment	0.202	–	–	–	–	–	–
Run x Treatment	0.118	–	–	–	–	–	–
Error	0.262	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=2170; Run=0 for run1; Run=1 for run2; Treatment=0 for fundamental; Treatment=1 for transitory.

Table S8: Prediction error (abs) in the fundamental treatment

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
Outlier	2.890***	2.815	2.966	0.038	12367	75.42	0.000
Outlier-Regular	2.547***	2.471	2.623	0.039	12367	65.36	0.000
PostOutlier-Outlier	−2.041***	−2.183	−1.900	0.072	12367	−28.29	0.000
Random effect (SD)							
Outlier	0.196	–	–	–	–	–	–
Outlier-Regular	0.197	–	–	–	–	–	–
PostOutlier-Outlier	0.382	–	–	–	–	–	–
Error	0.513	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=12400.

Table S9: Prediction error (abs) in the transitory treatment

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
Outlier	2.774∗∗∗	2.738	2.811	0.019	12367	149.18	0.000
Outlier-Regular	2.415∗∗∗	2.377	2.453	0.019	12367	123.99	0.000
Reversal-Outlier	−1.393∗∗∗	−1.552	−1.234	0.081	12367	−17.18	0.000
Random effect (SD)							
Outlier	0.023	–	–	–	–	–	–
Outlier-Regular	0.023	–	–	–	–	–	–
Reversal-Outlier	0.427	–	–	–	–	–	–
Error	0.583	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=12400.

Table S10: Prediction error (abs) for regular trial as a function treatment (relative to Bayes)

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
(Intercept)	0.028∗∗∗	0.022	0.035	0.004	20643	8.11	0.000
Treatement	0.025∗∗∗	0.012	0.037	0.006	20643	3.93	0.000
Random effect (SD)							
(Intercept)	0.014	–	–	–	–	–	–
Treatment	0.029	–	–	–	–	–	–
Error	0.243	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=20675.

Table S11: Prediction error (abs) for outlier trials as a function treatment (relative to Bayes)

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
(Intercept)	0.014	−0.010	0.037	0.012	2138	1.14	0.255
Treatement	0.039∗	0.008	0.071	0.016	2138	2.44	0.015
Random effect (SD)							
(Intercept)	0.025	–	–	–	–	–	–
Treatment	0.000	–	–	–	–	–	–
Error	0.374	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=2170.

Table S12: Prediction error (abs) for trials following outliers as a function treatment (relative to Bayes)

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
(Intercept)	0.470∗∗∗	0.343	0.597	0.065	1923	7.27	0.000
Treatement	0.316∗∗	0.116	0.516	0.102	1923	3.09	0.002
Random effect (SD)							
(Intercept)	0.323	–	–	–	–	–	–
Treatment	0.518	–	–	–	–	–	–
Error	0.913	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=1955.

Table S13: Prediction error (abs) as a function of training in the fundamental treatment

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
(Intercept)	0.910***	0.765	1.055	0.074	1006	12.31	0.000
Run	−0.126*	−0.222	−0.029	0.049	1006	−2.56	0.011
Random effect (SD)							
(Intercept)	0.364	–	–	–	–	–	–
Run	0.000	–	–	–	–	–	–
Error	0.785	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=1038.

Table S14: Prediction error (abs) as a function of training in the transitory treatment

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
(Intercept)	1.597***	1.425	1.768	0.087	885	18.31	0.000
Run	−0.435***	−0.614	−0.257	0.091	885	−4.79	0.000
Random effect (SD)							
(Intercept)	0.376	–	–	–	–	–	–
Run	0.255	–	–	–	–	–	–
Error	1.162	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=917.

Table S15: Prediction error (abs) as a function of training and treatment (interaction)

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
(Intercept)	0.909***	0.772	1.046	0.070	1921	13.01	0.000
Run	−0.132*	−0.251	−0.012	0.061	1921	−2.16	0.031
Treatment	0.676***	0.457	0.895	0.112	1921	6.05	0.000
Run x Treatment	−0.297**	−0.515	−0.079	0.111	1921	−2.67	0.008
Random effect (SD)							
(Intercept)	0.307	–	–	–	–	–	–
Run	0.000	–	–	–	–	–	–
Treatment	0.512	–	–	–	–	–	–
Run x Treatment	0.362	–	–	–	–	–	–
Error	0.977	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=1955; Run=0 for run1; Run=1 for run2; Treatment=0 for fundamental; Treatment=1 for transitory.

Table S16: Deliberation time in the fundamental treatment

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
Outlier	0.662***	0.586	0.737	0.038	10453	17.23	0.000
Outlier-Regular	0.071***	0.031	0.110	0.020	10453	3.50	0.000
PostOutlier-Outlier	−0.091***	−0.145	−0.037	0.028	10453	−3.30	0.001
Random effect (SD)							
Outlier	0.201	–	–	–	–	–	–
Outlier-Regular	0.082	–	–	–	–	–	–
PostOutlier-Outlier	0.112	–	–	–	–	–	–
Error	0.413	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=10486.

Table S17: Deliberation time in the transitory treatment

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
Outlier	0.663∗∗∗	0.603	0.722	0.030	10332	21.79	0.000
Outlier-Regular	0.082∗∗∗	0.053	0.112	0.015	10332	5.46	0.000
Reversal-Outlier	0.075∗∗	0.025	0.125	0.026	10332	2.94	0.003
Random effect (SD)							
Outlier	0.149	–	–	–	–	–	–
Outlier-Regular	0.000	–	–	–	–	–	–
Reversal-Outlier	0.081	–	–	–	–	–	–
Error	0.428	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=10365.

Table S18: Deliberation time as a function of training in the fundamental treatment

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
(Intercept)	0.614∗∗∗	0.545	0.682	0.035	10454	17.51	0.000
Run	−0.035∗∗	−0.060	−0.011	0.013	10454	−2.81	0.005
Random effect (SD)							
(Intercept)	0.192	–	–	–	–	–	–
Run	0.053	–	–	–	–	–	–
Error	0.415	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=10486.

Table S19: Deliberation time as a function of training in the transitory treatment

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
(Intercept)	0.600∗∗∗	0.542	0.657	0.029	10333	20.55	0.000
Run	0.001	−0.028	0.030	0.015	10333	0.07	0.948
Random effect (SD)							
(Intercept)	0.159	–	–	–	–	–	–
Run	0.068	–	–	–	–	–	–
Error	0.429	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=10365.

Table S20: Early BOLD response to outliers (attentional network)

Region	*T-value*	*Df*	*x*	*y*	*z*
Left Lingual	9.44	29	−22	−70	−2
Right Lingual	8.30	29	22	−64	2
Left Superior Occipital	6.26	29	−22	−74	34
Right Superior Occipital	6.54	29	26	−70	38
Left Parietal/Post Central	6.01	29	−34	−44	42
Right Parietal/Post Central	7.40	29	36	−44	52
Left Superior Precentral	7.82	29	−34	−6	50
Right Superior Precentral	7.69	29	40	0	44
Medial superior frontal (SMA)	5.50	29	−6	14	48
Left Inferior Precentral	5.84	29	−44	2	30
Right Inferior Precentral	7.39	29	50	8	28
Left Insula	8.04	29	−28	24	0
Right Insula	6.21	29	30	28	−4
Thalamus	6.61	29	−2	−22	−2

Peak activations in clusters surviving qFDR $<$ .05; MNI coordinates.

Table S21: Late BOLD response to outliers (fronto-parietal network)

Region	*T-value*	*Df*	*x*	*y*	*z*
Left Angular Gyrus	6.23	29	−40	−64	44
Right Angular Gyrus	6.43	29	40	−60	48
Left Middle Frontal Gyrus	5.45	29	−34	24	46
Right Middle Frontal Gyrus	5.77	29	36	28	46
Left Ant. Middle Frontal Gyrus	7.16	29	−36	56	0
Right Ant. Middle Frontal Gyrus	6.09	29	38	56	0

Peak activations in clusters surviving qFDR $<$ .05; MNI coordinates.

Table S22: Early BOLD response to outliers (transitory minus fundamental)

Region	*T-value*	*Df*	*x*	*y*	*z*
Left Insula	3.75	29	−28	24	−6
Right Insula	4.76	29	32	26	−6

Peak activations at p $<$ .001, uncorr; MNI coordinates.

Table S23: Insula ROI: Early effect of outliers as a function of the treatement

Variable	*Estimate*	*Lower*	*Upper*	*SE*	*Df*	*t*	*p*
Fixed effect							
(Intercept)	1.420	−1.645	4.484	1.563	49565	0.91	0.364
Side	−1.403∗∗∗	−1.728	−1.079	0.166	49565	−8.47	0.000
Outlier	0.533	−0.435	1.500	0.494	49565	1.08	0.280
Treatement	−0.724	−4.601	3.152	1.978	49565	−0.37	0.714
Outlier x Treatement	2.152∗∗	0.693	3.611	0.744	49565	2.89	0.004
Random effect (SD)							
(Intercept)	8.659	–	–	–	–	–	–
Outlier	2.293	–	–	–	–	–	–
Treatement	10.960	–	–	–	–	–	–
Outlier x Treatement	3.537	–	–	–	–	–	–
Error	18.450	–	–	–	–	–	–

*0 not included in the 95% Confidence Interval; Obs=49600; Side=0 for left, Side=1 for right; Treatment=0 for fundamental, Treatment=1 for transitory.

References

Doob, J. (1948). Le Calcul des Probabilités et ses Applications. In *Colloques Internationales du CNRS Paris* (pp. 22–28).

Sutton, R. S. (1992). Adapting bias by gradient descent: An incremental version of delta-bar-delta. In *Proceedings of the 10th National Conference on Artificial Intelligence. Cambridge (MA). MIT Press* (pp. 171–176).