

Supplemental Methods and Materials

Participants

The included bipolar patients met the criteria for euthymia for at least 2.5 months prior to recruitment, as indicated by the Beck Depression Inventory (BDI, [105]), Altman Self-Rating Mania Scale [106], Hospital Anxiety and Depression Scale (HADS, [107]), and State-Trait Anxiety Inventory (STAI, [108]). The healthy individuals did not have neurological or mental health conditions and were not currently taking any medication for anxiety or depression. Participants in both groups were right-handed, had normal or corrected-to-normal vision, and demonstrated the ability to perform controlled finger movements.

The BDI ranges from 0 to 63, with values above 14 denoting light to severe depression. The Altman Self-Rating Mania Scale ranges 0–20, with values up to 5 considered normal. The HADS, which assesses anxiety and depression separately, ranges from 0 to 21, with normal scores in the range of 0 to 7. We additionally assessed state anxiety on the day of the experimental session to determine potential group differences, which could affect decision-making [59]. We used the state subscale of the Spielberger State-Trait Anxiety Inventory [108], ranging from 0 to 80, with values up to 30 considered low to medium.

In addition, in both cohorts, we assessed cognitive performance including executive functioning using the Trail Making Test (TMT, part 1 and 2, [109,110]), and reversal learning using the Wisconsin Card Sorting Test (WCST, [111]). Cognitive impairment was assessed with the Mini-Mental State Examination (MMSE, [112]) scale. See the test scores in **Table 1**. In the BD sample, the average duration of BD was 7.4 years (SEM 1), and the average duration of euthymia was 12.9 months (SEM 2.6). All BD participants had depression as their last episode, in line with one of the inclusion criteria of the study.

Blinding of data collection and analysis

The experimenter conducting the study was not blinded to the group allocation. Preprocessing of MEG, MRI, and behavioural data was not blinded.

Sample size

Power analysis was informed by a recent study assessing between-group differences in mean volatility estimates in a similar paradigm in healthy individuals with high and low trait anxiety levels ([45]; non-parametric probability of superiority, $\Delta \sim 0.74$, equivalent to a Cohen's d of 0.91, resulting in a minimum of 20 participants per group for 80% power). Differences in frequency-domain amplitude changes during belief updating in that study were associated with a non-parametric effect size of $\Delta \sim 0.735$, equivalent to a Cohen's d of 0.89, resulting in a minimum of 21 participants per group for 80% power. Although larger

effects in volatility and neural differences between bipolar patients and their healthy counterparts were anticipated, we recruited 22 BD patients to account for potential variations in effects across computational and neural variables.

Reward-based motor sequence learning task

Before commencing the main task, participants performed 20 simple sequences of eight finger movements: 1-2-3-4 rightwards and 4-3-2-1 downwards, corresponding to the index, middle, ring, and little fingers on a four-button response box (**Figure 1a**). Each button press produced a specific sound with pitch values E5, F5, G5, A5 (**Figure 1a**). This served as a baseline measure of their fine motor skills.

The sequences used in the primary task were selected from our previous study [57] and were chosen in that study, based on pilot work, to be matched in RT, tempo, and error rate. In the current study, performance in both tasks was also similar. See below (“**Task validation: Similar performance of sequences 1 and 2**”).

The primary task had both a familiarisation and test phase. In the *familiarisation phase*, participants learnt two sequences of four finger presses, each linked to a specific fractal image (**Figure 1a**). First, they practiced the sequences for 15 trials each, without a concurrent probabilistic task, followed by practice of the probabilistic task: 10 trials at a fixed contingency mapping of 70/30, as in our previous work [45,59]. This practice aimed at helping participants understand the timeline of stimuli and the type of feedback they would receive. Regarding the memorisation of the motor sequences, we validated that participants performed without errors in the last 5 of the 15 practice trials, which was achieved in all cases as the sequences were short (4 key presses). There were no accuracy criteria set for the practice version of the probabilistic task, in line with previous work [45,59,61].

The *test phase* comprised two blocks, totalling 320 trials. In each trial, participants selected between two fractal images, playing the corresponding motor sequence (sequence1, sequence2) for a reward. These sequences, and the corresponding fractals, were implicitly associated with reward values of 5 (rewarded) or 0 points (non-rewarded). The probability of a sequence yielding a reward was reciprocal ($p(\text{sequence1}|\text{reward}) = 1 - p(\text{sequence2}|\text{reward})$) and shifted pseudorandomly every 26-38 trials. The probability mappings could adopt values of 0.9/0.1, 0.7/0.3, 0.1/0.9, 0.3/0.7, or 0.5/0.5 in each block (**Figure 1b**). Participants received feedback post-trial, indicating 'You earned 5 points' or 'You earned 0 points'. They had a 3300 [\pm 125] ms window to start and finish the sequence. Feedback was presented 1000 [\pm 200] ms after sequence completion and remained visible for 1900 [\pm 100] ms. If participants took over 3300 ms (time out) or made a sequencing error, they were notified of either the delay or mistake, respectively, and garnered no points. The inter-trial-interval (ITI) was 1750 [\pm 250] ms. Control analyses on the pseudo randomised

probabilistic mappings demonstrated that both participant samples experienced the same true volatility. See details in next section.

Validation of paradigm: Probabilistic structure

The contingencies were generated pseudorandomly in each participant, which could result in both groups experiencing differences in the true volatility by chance. We examined this possibility by comparing the true rate of contingency change in each group: the average number of trials where a change takes place across both blocks. A greater number here would denote smaller true volatility (i.e. changes occur less frequently). On average, contingencies changed every 32 trials in both groups (32.3 [SEM 0.25] in HC, 32.9 [SEM 0.42] in BD), and Bayes factor analysis supported they were equivalent ($BF_{10} = 0.61$, anecdotal evidence; $P = 0.1957$, non-significant difference).

Validation of paradigm: Similar performance of sequences 1 and 2

Difficulty in both sequences was similar (as expected based on [57]). First, we assessed reaction time (RT, time from stimulus to first key press) and performance tempo (mean inter-key press interval in a sequence, mIKI; used interchangeably). RT and performance tempo did not differ between sequences 1 and 2, as indicated by substantial evidence for the null hypothesis from Bayes Factor analysis (BF in range 1/10 to 1/3). This was assessed in each group separately and in the full cohort. On average, RT was 0.523 seconds (SEM 0.02) for sequence 1 and 0.529 seconds (SEM 0.02) for sequence 2 (joint sample, $N = 49$, $P = 0.3154$, $BF_{10} = 0.2568$). The performance tempo was 0.348 seconds (SEM 0.01) and 0.347 seconds (SEM 0.01) for sequence 1 and 2, respectively ($P = 0.7382$, $BF_{10} = 0.1650$). The error rate was also similar between the two sequences (error rate < 1% in both cases, $P = 0.3036$, $BF_{10} = 0.2589$). Similar statistical results were obtained in each group separately.

General task performance

Following [45,59], win-stay rates were estimated as the ratio of the number of trials in which participants repeated the same sequence following a win, divided by the total number of win (rewarded) trials. The remaining types of trials after a win were either win-shift or, in very few instances, win-error trials (denoting an error following a win trial). Accordingly, $p(\text{win-stay}) = 1 - p(\text{win-shift}) - p(\text{win-error})$. Similarly, lose-shift rates were calculated as the ratio of trials in which a response was changed after losing (no reward), divided by the total number of loss trials. And $p(\text{lose-shift}) = 1 - p(\text{lose-stay}) - p(\text{lose-error})$. Due to the near-perfect complementarity between win-stay and win-shift rates (except for a very small error rate,

see **Supplementary Results**), and similarly between lose-shift and lose-stay rates, we did not analyse between-group differences in win-shift and lose-stay rates to avoid redundancy and additional multiple comparisons.

Control analyses assessed whether any differences between groups in our primary dependent variables (DV) could be confounded by differences in error rates. Two types of errors were considered, as in [57]: performance errors, resulting from participants making a sequence performance error (e.g., playing at least one incorrect note in the sequence), and timeouts, when participants did not complete the sequence within the allocated 3300 ms.

Modelling decision-making behaviour using hierarchical Gaussian filters

We employed a validated Bayesian framework with a Hierarchical Gaussian Filter (HGF) to model how input about probabilistic reward outcomes and their change over time is integrated with prior beliefs during learning, resulting in posterior beliefs about the hidden states causing the observed outcomes [26,27].

Drawing inspiration from Behrens et al. [22], in the HGF a sequence of hidden states $\{x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}\}$ —with k representing a trial or unit of time—is conceptualised within a generative model consisting of hierarchically coupled Gaussian random walks. These walks are coupled through their variance (inverse precision) and evolve over time.

The HGF generates dynamic and adaptive learning rates, capturing the process of learning under uncertainty in a volatile environment. In this hierarchy, higher levels represent the dynamic structure of the world, and the step size of each random walk is influenced by the state at the level above. Inverting this generative model using variational approximation provides update rules for the temporal trajectories of beliefs held by the agent.

The modelling framework consists of two components: a perceptual model and an observational model. The perceptual model delineates how sensory input—here, reward outcomes—is mapped to the hidden states of the world that generate these inputs. The observational model describes the mapping from the agent's probabilistic representations or beliefs to the produced responses. In other words, the response model accounts for how the decisions of the agent we are observing are derived based on their perception, encompassing both their observations and the inferences drawn from them.

To estimate each participant's individual learning characteristics and belief trajectories during our binary reward-based learning task in a volatile setting, we implemented the three-level perceptual HGF for binary categorical inputs. At the lowest level, the hidden state x_1 corresponds to the binary categorical variable of the experimental stimuli: whether sequence 1 is rewarded in trial k ($x_1^{(k)} = 1$) or not ($x_1^{(k)} = 0$). Beliefs are represented on the second and third levels and modelled as Gaussian distributions. The second level

represents the trajectory of participants' beliefs about the contingency between actions (sequence 1 or 2) and their outcomes (rewarded or not), and the third level represents the rate of change in that tendency (volatility). Gaussian belief distributions are represented by their posterior mean ($\mu_2^{(k)}, \mu_3^{(k)}$ for levels 2 and 3 respectively) and posterior variance (uncertainty: σ_2, σ_3). Precision is the inverse variance or uncertainty, π_i ($i = 2$ and 3). As described in the main text, the HGF includes several forms of uncertainty used in different modelling frameworks [26,60,61,113]: irreducible, estimation and volatility uncertainty, which are represented by $\sigma_1, \sigma_2, \sigma_3$. While estimation (also known as informational) and volatility uncertainty can be reduced through learning, irreducible uncertainty is not: it emerges from the probabilistic relationships between responses and their outcomes, which is an inherent property of most real-world interactions. Another form of uncertainty that can be assessed in the HGF, aligning with previous models [22,60], is environmental uncertainty, which is induced by changes in the environment (volatility), and is defined as $\exp(\kappa\mu_3^{(k-1)} + \omega_2)$.

Belief updating on each level i ($i = 2$ and 3) and trial k is driven by PEs, modulated by precision ratios, weighting the influence of precision or uncertainty in the current level and the level below:

$$\Delta \mu_i^k = \mu_i^{(k)} - \mu_i^{(k-1)} \propto \frac{\hat{\pi}_{i-1}^{(k)}}{\pi_i^{(k)}} \delta_{i-1}^{(k)} \quad (4)$$

Following equation (4), the expectation of the posterior mean on level i , $\mu_i^{(k-1)}$, is updated to its current level $\mu_i^{(k)}$ proportionally to the prediction error of the level below, $\delta_{i-1}^{(k)}$. The influence of PEs is weighted by the ratio of precision values, with the *prediction* (denoted by “^”) of the precision of the level below in the numerator and the precision of the current level (inverse uncertainty, σ_i) in the denominator. Here we assume the prediction from trial $k-1$ remains constant until the beginning of the current trial, without drifting to a new value before the agent observes the outcome. Therefore, $\hat{\mu}_i^{(k)} = \mu_i^{(k-1)}$. See Weber and colleagues [114] for a generalised version of the HGF that explicitly models drifts in the prediction.

In the HGF for binary outcomes, the precision ratio updating beliefs on level 2 in equation (4) is reduced to $\sigma_2^{(k)} (1/\pi_2^{(k)})$, as shown in equation (1) in the main text. Accordingly, the posterior mean of the belief about the action-outcome contingencies is updated via low-level PE about action (stimulus) outcomes, scaled by the degree of informational uncertainty. For level 3, the precision ratio is proportional to the uncertainty about volatility, $\sigma_3^{(k)}$ (inverse precision on level 3: $1/\pi_3^{(k)}$).

The HGF perceptual model was coupled with a unit-square sigmoid response model where choice probability is shaped by a free fixed (time-invariant) parameter ζ , interpreted as inverse decision noise: the sigmoid approaches a step function as ζ tends to infinity. The mapping between the predictive probability $m^{(k)}$ for an outcome on trial k onto the probability that the individual will choose response 1 or 0, $p(y^{(k)} = 1)$ and $p(y^{(k)} = 0)$ respectively, takes this form (omitting trial index k again for simplicity):

$$p(y|m, \zeta) = \left(\frac{m^\zeta}{m^\zeta + (1-m)^\zeta} \right)^y \cdot \left(\frac{(1-m)^\zeta}{m^\zeta + (1-m)^\zeta} \right)^{(1-y)} \quad (5)$$

See further detail in Eq. 18 in Mathys et al. [27]. This combination constituted our first perceptual-response model (M1). Next, we used a perceptual 2-level HGF model with volatility fixed to a constant level and coupled it with this unit-square sigmoid response model (M2). Our third model combined the 3-level HGF with a response model where the sigmoid function depends on the trial-wise prediction of log-volatility, $\zeta = e^{-\mu_3^{(k-1)}}$, (M3; [25]). In this observational model, higher estimates of volatility lead to a 'noisier' relationship between beliefs and decision making. As a result, there is an increased likelihood of choosing responses that deviate from predictions. Last, model M4 was constructed similarly to M3 but replaced the free parameter ω_2 with κ [32].

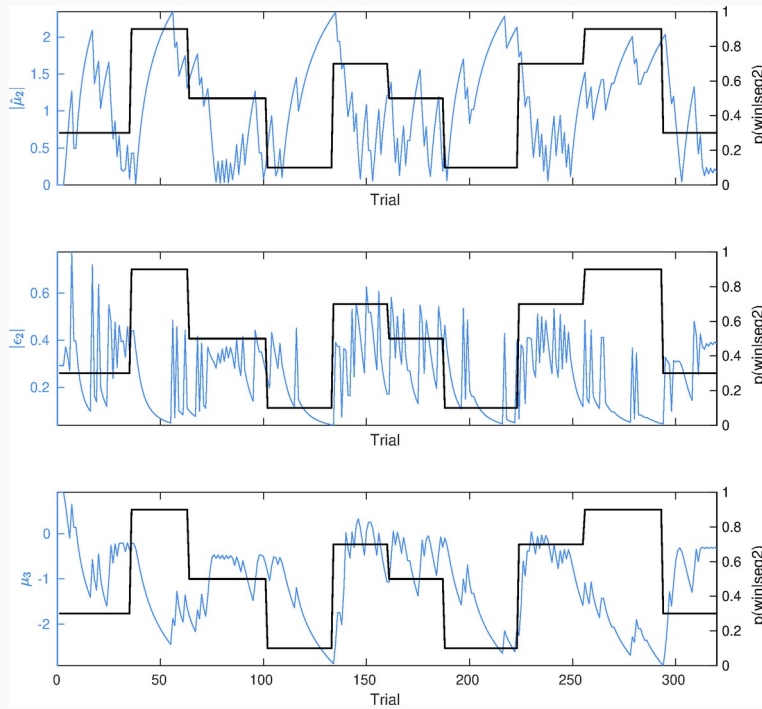


Figure S1. Same as **Figure 2b** in the main manuscript, but with an overlay of the probability of reward for sequence 2 ($p(\text{reward}|\text{seq2}) = 1 - p(\text{reward}|\text{seq1})$). The chosen variables, $|\hat{\mu}_2^{(k)}|$, $|\varepsilon_2|$, and μ_3 represent the regressors used in the Bayesian multilevel modelling (top), one regressor used in the GLM analysis (middle), and the trajectory of log-

volatility, respectively. Top: The *strength* of predictions about response-outcome contingencies is higher (closer to 1) for probabilistic blocks 90/10 and 10/90, as expected, while volatility is lower in these phases (bottom).

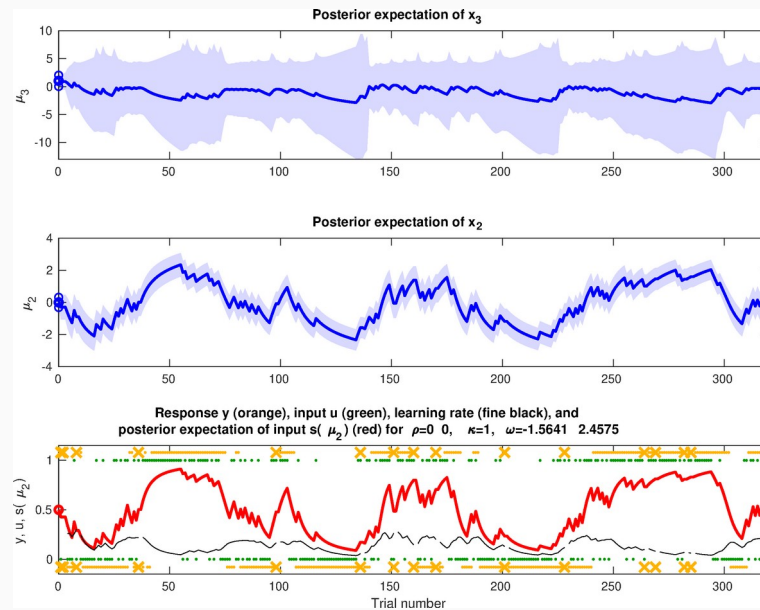


Figure S2. Representation of the three levels of the HGF for binary outcomes and the associated belief trajectories across the total 320 trials in a representative participant (same as in **Figure S1**). At the lowest level, the inputs u correspond to the rewarded outcome of each trial (1 = sequence2, 0 = sequence1; shown as green dots). The participant's responses y are shown in orange dots tracking those trial outcomes (errors as "x"). The learning rate about stimulus outcomes at the lowest level is also given in black. The posterior expectation of inputs (in red) is a sigmoid transformation of beliefs on level 2, $s(\mu_2)$. Middle: The belief on the second level, μ_2 (σ_2), represents the participant's estimate of the stimulus tendency x_2 and the step size or variance of the Gaussian random walk for x_2 depends on parameters κ and ω_2 , in addition to the estimates of the level above, x_3 . The belief on the third level, μ_3 (σ_3), represents estimates of volatility x_3 , whose step size is governed by parameter ω_3 . Figure created using the HGF toolbox function `tapas_hgf_binary_plotTraj()`.

Selection of priors for HGF models

We selected prior values for our HGF models based on estimates obtained from our data through an ideal observer model. An ideal observer is defined as a model that adopts a range of parameter values aimed at minimising the overall surprise an agent experiences upon receiving a series of inputs (refer to Weber et al. [115] and [80] for different applications of an ideal observer model). This approach was preferred over the use of previously reported prior values for the binary categorical HGF model, primarily because the code in the TAPAS toolbox has undergone modifications and optimisations, leading to improved parameter recovery compared to earlier versions. Consequently, prior values used in previous studies may not be directly applicable to newer model implementations. We used the HGF release v7.1 in Matlab R2020b, and functions 'tapas_ehgf_binary'.

Using ideal observer models, the group median of the prior values estimated separately for HC and BD were:

$$\omega_2, \omega_3 = [-2.5 -0.5] \text{ in HC}$$

$$\omega_2, \omega_3 = [-2.6 -0.1] \text{ in BD}$$

These values were comparable between groups ($BF_{10} = 0.4219, 0.7014$, support for H_0 , based on anecdotal and substantial evidence; $P = 0.3447, 0.1458$, no significant differences). Accordingly, in our main analyses, we selected the median values for $[\omega_2, \omega_3]$ from the full sample ($N = 49$), which were $[-2.6 -0.3]$. For complete details on our prior parameters, see **Table S1**.

In a control analysis employing group-specific priors, we evaluated the consistency of our main computational findings. This analysis, however, should be interpreted with caution due to the comparison of computational variables derived from distinct models for each group. We observed significant group differences in ω_2 ($P_{FDR} = 0.0049$), as well as in volatility at onset and average volatility ($P_{FDR} = 0.0049, 0.0056$ respectively). Interestingly, the evidence for differences between groups was stronger than that obtained using common prior settings, with Bayes Factors exceeding 9. This indicates substantial evidence supporting the alternative hypothesis (H_1).

Table S1. Priors (means and variances) on perceptual parameters and starting values of the belief distributions for the winning HGF model M3.

Prior	Mean	Variance
κ	$\log(1)$	0
ω_2	-2.6	4
ω_3	-0.3	4
$\mu_2^{(0)}$	0	0
$\sigma_2^{(0)}$	$\log(0.1)$	0
$\mu_3^{(0)}$	$\log(1)$	1
$\sigma_3^{(0)}$	$\log(1)$	1

Quantities are estimated in their native space when they are unbounded, such as ω_2, ω_3 . Conversely, quantities with a natural lower bound at zero, like $\kappa, \mu_3^{(0)}$, and $\sigma_3^{(0)}$, are estimated in log-space. In the winning HGF model M3, $\omega_2, \omega_3, \mu_3^{(0)}, \sigma_3^{(0)}$ were free parameters ($\kappa, \sigma_2^{(0)}, \mu_2^{(0)}$ were fixed). The prior variances are in the space in which the corresponding parameter is estimated.

Assessing motor invigoration

The nigrostriatal dopamine pathway, central to the 'dopamine hypothesis' of BD [54,55], is crucial for invigorating future movements [53]. This hypothesis posits that aberrant dopamine processing is a cornerstone of the disorder, with dopamine dysregulation hypothesised to underlie both manic episodes and the broader episodic features of BD [54,55]. Previous research has shown that dopaminergic pathways, essential for motivating movement, play a key role in modulating motor vigour in response to reward [53,116]. Moreover, recent animal studies provide robust mechanistic evidence of how dopamine modulates vigour, partly by synchronising with cholinergic activity in the basal ganglia and integrating inputs from the cerebellum [117,118].

Accordingly, an open question in BD research is whether the hypothesised dopamine dysregulation results in motor vigour alterations. Empirical studies have shown that individuals with BD exhibit heightened energy and effort following success, indicating enhanced motor vigour effects [4,56]. Moreover, a recent longitudinal study monitoring mood, physical activity, energy, and sleep found that changes in physical activity and energy precede mood changes, rather than the reverse [119]. This underscores the importance of understanding motor vigour alterations in BD.

Here, we aimed to determine whether trial-wise expectations about the tendency of the action-outcome contingencies would be associated with faster performance tempo in sequences of finger movements, an invigoration effect that we described previously in a similar task [57]. Additionally, we assessed whether decision-making differences between bipolar and healthy participants led to distinct motor invigoration effects. This was motivated by findings that BD individuals, including those in euthymic phases, mobilise more effort, confidence, and energy while expecting and receiving rewards [56]. We investigated whether trial-by-trial expectations about the probabilistic action-outcome mapping differentially influenced motor performance in our groups, focusing on performance tempo and, separately, RT. Notably, faster sequence completion would expedite the display of the outcome (at 1000 [\pm 250] ms post-performance).

This analysis was implemented using Bayesian multilevel regression modelling (BML) with the logarithm of trial-wise timing variables as the DVs. BML models were separately implemented using timing variables: (i) RT and (ii) performance tempo (mKI), both measured in milliseconds and transformed to logarithmic milliseconds (log-ms) for BML.

BML estimates approximate posterior probability distributions for model parameters and incorporates accurate uncertainty estimates, even with unbalanced data and small sample sizes [120]. We conducted BML using the R package brms [121,122], employing informative priors and estimating models by Markov-

Chain Monte Carlo (MCMC) sampling. This process involved drawing 20,000 samples across four chains, discarding the first 1000 iterations of each chain as warm-up. We validated chain convergence using the Gelman–Rubin statistics ($R\text{-hat} \leq 1.01$; [123,124]). As an additional convergence indicator, we used the effective sample size (ESS). As recommended by [124], ESS should exceed 100 for each chain (> 400 in our analysis with four chains; criterion applied to the bulk-ESS—the effective sample size based on the rank normalized draws—and the tail effective sample size, tail-ESS). This recommendation is valid also for larger numbers of draws (16000 post-warmup in our case).

Following Tecilla et al. [57], given the sign of $\hat{\mu}_2^{(k)}$ in the HGF is arbitrary, we used as predictor for BML the absolute value, $|\hat{\mu}_2^{(k)}|$, representing the *strength* of those predictions, and assessed its association with performance variables (log-tempo, logRT) as DVs. Our hypothesis was that the relationship between $|\hat{\mu}_2^{(k)}|$ and performance tempo would exhibit a non-zero (negative) slope, indicating that stronger beliefs about probabilistic contingencies correlate with faster sequence performance. Furthermore, we posited that performance tempo in euthymic bipolar patients might exhibit increased sensitivity to the strength of predictions [4,56], reflected in a comparatively steeper slope than that observed in healthy controls. Similar analyses were conducted for RT, to determine whether the invigoration effects were specific to performance tempo, as shown previously in this task [57].

To evaluate our motor vigour hypotheses, we first constructed a series of BML models for tempo as DV [57]. The most complex model (**Table S2**) incorporated an interaction between group (a categorical variable distinguishing between HC and BD participants) and the centred continuous variable $|\hat{\mu}_2^{(k)}|$ at trial k . This centring step of the continuous predictor is recommended to improve model stability and facilitate the interpretability of the intercept. The interaction term allowed us to determine group-specific associations between predictions and our DV. Additionally, the model included a random effects structure to account for subject-specific variations in intercepts and slopes for the centred predictor, as well as trial-to-trial variations.

Subsequent models of decreasing complexity were constructed by successively omitting random effects or interaction terms. HC served as the reference group for between-group comparisons, providing posterior distributions of differences in slopes and intercepts (see **Table S2** for a full list of models).

Model comparison was conducted using leave-one-out cross-validation (LOO-CV) with Pareto-smoothed importance sampling [125]. The best-fitting model was identified based on the highest expected log point-wise predictive density (ELPD). We also ensured that the absolute mean difference in ELPD between the

two top models (elpd_diff in brms) exceeded twice the standard error of these differences (2se_diff). If elpd_diff was smaller than 2se_diff, we opted for the more parsimonious model.

Similar models were constructed to assess the association between the strength of predictions and logRT, which we expected to be less robust than for performance tempo, as shown in [57] for this task. However, the aim of the RT models was to assess whether the invigoration effects were specific to performance tempo or extended to RT as well.

In all models, trial-wise outliers in performance tempo or RT, defined as values exceeding three standard deviations from the mean, were excluded. On average, we removed 4.37 (SEM 0.4) trials from the mIKI dataset and 3.67 (SEM 0.3) trials from the RT dataset.

Table S2. Bayesian Multilevel Models with a Gaussian distribution assessing the effect of strength of predictions on tempo or RT. Models of decreasing complexity were defined to assess whether a timing variable (tempo [logmIKI] or logRT, represented by y) was modulated by the strength of predictions about the tendency of the action-outcome contingency, $|\hat{\mu}_2^{(k)}|$. This predictor variable was centred and is denoted by *prediction.c* in the table. The most complex model included the random effect of trials, and the random effect of participants on the slope of the timing by prediction association and the intercept.

Model #	Model
1	$y \sim 1 + \text{prediction.c}$
2	$y \sim 1 + \text{group} * \text{prediction.c}$
3	$y \sim 1 + \text{prediction.c} + (1 \text{subject})$
4	$y \sim 1 + \text{group} * \text{prediction.c} + (1 \text{subject})$
5	$y \sim 1 + \text{prediction.c} + (1 + \text{prediction.c} \text{subject})$
6	$y \sim 1 + \text{group} * \text{prediction.c} + (1 + \text{prediction.c} \text{subject})$
7	$y \sim 1 + \text{group} * \text{prediction.c} + (1 + \text{prediction.c} \text{subject}) + (1 \text{trial})$
	The brms family was Gaussian

We applied a default prior distribution for the intercept in all models, and a normal distribution for each fixed effect (assuming a constant effect of group and prediction.c: normal prior with mean 0 and standard deviation 2; interaction term group x trial: normal prior with mean 0 and standard deviation 1). For the priors on the standard deviation (sd) parameters associated with the random effects (individual subjects,

trials), we also employed the default half Student-t prior with 3 degrees of freedom. Models 5-7 include a prior on the LKJ correlation coefficient set to 2.

Control analyses: Assessing practice effects in timing performance

We controlled for between-group differences in the timing of sequence performance and practice effects over trials using a series of Bayesian multilevel models with a Gaussian distribution for performance tempo and, separately, for RT. These timing variables were log-transformed (log-ms) to serve as DVs. The models included fixed effects, such as Group (BD, HC; with HC as the reference) and Trial (1:320), and random effects, capturing individual variation in subject effects on the intercept and/or slope. Fixed effects refer to effects that are assumed not to vary across subjects, while random effects are considered to vary among individuals (e.g., in their intercept or slope).

The most complex model (Model 8 in **Table S3**) included Group, Trial, and their interaction as fixed effects and permitted individual variation in both intercepts and slopes. This approach aimed to reveal how tempo/RT evolved across trials in each participant. For comparative purposes, additional models with varying degrees of complexity were constructed (refer to **Table S3**). Trial-wise outliers in mIKI or RT, defined as values exceeding three standard deviations from the mean, were also excluded from the analysis.

For model comparison, we applied the same LOO-CV approach with Pareto-smoothed importance sampling as used in the motor vigour analysis. The selection of models was based on the highest ELPD as noted in previous section.

Table S3. Bayesian Multilevel Models with a Gaussian distribution assessing practice effects on tempo or RT.

Decreasingly complex models were constructed to assess practice effects on tempo (logmIKI, measured in log-ms) or, separately, on RT (logRT, in log-ms). The same models were assessed for each dependent variable, separately, denoted by “y” in the table. Models 1, 2, and 3 explain y through the fixed effects of group (BD, HC; with HC as the reference group; fixed here denotes an effect assumed not to be changing across subjects) and trial (1:320), or their additive combination, while Model 4 includes their interaction effect. Models 5-7 incorporate the random effect of the intercept by subject, thus modelling how the intercept changes across subjects. The most complex model, Model 8, allows for individual variation in both intercepts and slopes, potentially revealing how individual subjects' tempo/RT changes over trials. Models 4, 7, and 8 enable the assessment of slope differences between BD and HC groups in the practice effect.

Model #	Model
1	$y \sim 1 + \text{group}$
2	$y \sim 1 + \text{trial}$

3	$y \sim 1 + \text{group} + \text{trial}$
4	$y \sim 1 + \text{group} * \text{trial}$
5	$y \sim 1 + \text{trial} + (1 \text{subject})$
6	$y \sim 1 + \text{group} + \text{trial} + (1 \text{subject})$
7	$y \sim 1 + \text{group} * \text{trial} + (1 \text{subject})$
8	$y \sim 1 + \text{group} * \text{trial} + (1 + \text{trial} \text{subject})$
	The brms family was Gaussian

As for the models listed in **Table S2**, we used a default prior distribution for the intercept, and a normal distribution for each population-level effect (effect of group and trial: normal prior with mean 0 and standard deviation 2; interaction term group x trial: normal prior with mean 0 and standard deviation 1). For the priors on the sd parameters associated with the group-level effects (individual subjects, trials), we used the default half Student-t prior with 3 degrees of freedom, as recommended by [121]. In Model 8, the prior on the LKJ-Correlation was set to 2.

MEG recording and preprocessing

Initial preprocessing of the MEG signals consisted of correcting for head movements, reducing noise, and eliminating bad channels using the temporally extended signal-space separation (tSSS) method [126], integrated into the Elekta Neuromag software (Maxfilter, Elekta Neuromag). The settings were adjusted to a sliding window of 10 seconds and a subspace correlation threshold of 0.9.

Source reconstruction of MEG signals

To source reconstruct MEG signals, integrating planar gradiometers and magnetometers, we used Linearly Constrained Minimum Variance beamforming (LCMV [65]) in the MNE-Python toolbox. We used individual T1-weighted magnetic resonance imaging (MRI) images to establish surface-based cortical divisions for each hemisphere with Freesurfer 6.0 software ([66,67]; <http://surfer.nmr.mgh.harvard.edu/>). MRIs were unavailable for two BD participants; consequently, in those cases, we utilised the fsaverage template brain files provided by Freesurfer.

We selected the Desikan-Killiany atlas (DK) atlas, which parcellates the cerebral cortex into 68 distinct anatomical regions [68]. The alignment of MRI and MEG coordinate systems was achieved with an

automated procedure within the MNE-Python toolbox (mne.gui.coregistration). This process relied on head position indicator coils and the registered points on the head's surface. Additionally, we ensured that the alignment of three fiducial anatomical points (both preauricular points and the nasion) was accurate across the coordinate systems.

Next, forward modelling was implemented with boundary element conductivity models (BEM) for each participant, with the inner skull surface serving as the chosen volume conductor layout. We then generated a surface-based source space at an "oct6" resolution, which provided 4098 positions (vertices) for every hemisphere with an average adjacent distance of 4.9 mm.

Solving the inverse problem with LCMV beamforming involved computing adaptive spatial filters using a data-covariance matrix for a target interval, selected based on expectations of task effects on neural activity. Informed by prior research indicating modulation of oscillatory activity by pwPE in similar decision-making tasks [45,46], we defined the target interval as 0–1.8 s post-outcome. The noise-covariance matrix for this analysis was estimated from -1 to 0 seconds pre-outcome.

To analyse modulation of activity in the theta (4–7 Hz), alpha (8–12 Hz) and beta (13–30 Hz) frequency bands, the MEG signals were first filtered with a 1–40 Hz band-pass filter, prior to LCMV beamforming. Separately, to assess activity in the gamma frequency range (32–100 Hz), the data were band-pass filtered within 30–124 Hz (as in [45]).

Subsequently, we extracted time courses for individual vertices within cortical labels corresponding to our regions of interest (ROIs), focusing on brain areas consistently associated with decision making, reward-based learning and belief updating under uncertainty. Our selected ROIs included: (1) the ACC, (2) the OFC along with the ventromedial Prefrontal Cortex (vmPFC), (3) the dorsomedial Prefrontal Cortex (dmPFC), (4) the dorsolateral PFC (dlPFC). The ACC and medial PFC are central to emotional regulation, reward processing, and decision making [43,70,71]. Within the medial PFC, the vmPFC is associated with representing reward probability, magnitude, and outcome expectations [72]. Gamma activity in the dmPFC has been correlated with unsigned reward prediction errors during exploration-exploitation behaviour [43]. On the other hand, the dlPFC has been shown to encode belief uncertainty prior to making a decision [73], though some reports link this region to belief updating under uncertainty [74]. Concerning the OFC, this region is crucial for emotional processing and reward/punishment decision making [72], yet different OFC circuits make unique contributions to flexible decision making [127]. Research has associated the medial OFC (mOFC) with encoding reward value, and the lateral OFC (lOFC) with processing nonreward and punishment [128]. In rats, specific OFC circuits have been causally linked to different computations underlying reversal learning [127]. This evidence is particularly pertinent in the context of bipolar disorder

as individuals with this condition exhibit impairments in fronto-striatal reward circuitry, including the anterior insula and ventral striatum, but also cortical areas such as the vmPFC and dlPFC [4,9,76,77].

In addition, we included (5) the primary motor cortex (M1) and (6) premotor cortex (PMC) to assess modulation of motor activity during belief updating, and the effect of decision making on motor vigour [78].

Overall, our study's ROIs correspond to **sixteen bilateral labels** in the DK atlas, encompassing **eight** distinct areas: (1) rostral and caudal ACC (rACC, cACC); (2) lateral and medial OFC (lOFC, mOFC), which include the vmPFC according to some MEG studies [129,130]; but see [131] for a debate on the vmPFC delineation); (3) superior frontal gyrus (SFG), representing the dmPFC and the supplementary motor area (SMA); (4) rostral middle frontal gyrus (rMFG); (5) precentral gyrus (M1), and (6) caudal MFG (cMFC).

The representative time series for each label were derived using the 'PCA flip' method in MNE-Python. This method employs singular value decomposition (SVD) on the time courses associated with each vertex within a specific brain region or label. Its primary function is to extract the first right singular vector, representing the direction of maximal power in each source. Following this extraction, the singular vector undergoes scaling and sign-flipping, which then yields the time course for the region of interest (ROI). Although the 'flip' operator was not relevant for our time-frequency (TF) analysis, it was essential for preparing the source-reconstructed time series for subsequent connectivity analysis.

Convolution modelling of time-frequency responses during outcome processing

We employed a general linear model (GLM) to investigate the frequency-domain, trial-by-trial amplitude changes associated with belief updating and processing uncertainty following the presentation of the outcome. In the winning HGF model, the trajectories of beliefs at levels 2 and 3 are updated proportionally to the magnitude of pwPEs: low-level pwPEs about the action-outcome contingencies, termed ϵ_2 , and high-level pwPEs about the environmental volatility, ϵ_3 . As the sign of pwPE changes updating level 2 is arbitrary in the binary categorical HGF, previous studies have opted for using the unsigned $|\epsilon_2|$ as a regressor in GLM analyses of neural activity [45,46]. Variable $|\epsilon_2|$ represents precision-weighted Bayesian surprise about reward outcomes, independent of whether they are related to sequence 1 or 2. Given the collinearity of ϵ_3 and $|\epsilon_2|$ (in each participant, the trajectories were highly correlated, Pearson R in range 0.52–0.94, $P < 10^{-10}$ in all cases), we selected $|\epsilon_2|$ as our primary parametric regressor, modulating induced oscillatory activity following the outcome presentation. Additional parametric regressors included informational uncertainty, σ_2 , and level 3 uncertainty, σ_3 , as the relevant (inverse) precision terms for the GLM. Furthermore, our GLM incorporated discrete regressors coding for win and lose outcomes, and error trials (performance errors or timeouts).

We implemented the GLM on the time series of induced responses over multiple frequencies using a linear convolution model developed by Litvak and colleagues [79]. This approach, an adaptation of the classical GLM used in fMRI analysis for time-frequency data, has been successfully employed in previous work to identify frequency-resolved neural correlates of belief updating, precision, and predictions during decision-making and perceptual learning [44-46]. It facilitates assessing the modulation of TF responses on a trial-by-trial basis by a specific explanatory regressor while controlling for the effects of other regressors included in the model.

This outcome-locked convolution model was solved in the source space after applying LCMV to the time series of concatenated epochs of MEG data within -0.5 – 1.8 s (see previous section). We tested the hypothesis that bipolar participants exhibit changes in gamma and concomitant alpha/beta activity during the encoding of pwPE signals. Additionally, we hypothesised that euthymic BD individuals would show alterations in alpha/beta oscillatory activity during the representation of uncertainty variables σ_2 (corresponding with precision weights for level 2), and σ_3 .

To conduct the convolution GLM, we estimated standard TF representations of the source-level time series using Morlet wavelets. TF spectral power was extracted between 4 and 100 Hz and transformed to amplitude, following ref. [79]. For the theta (4–6 Hz), alpha (8–12 Hz), and beta (13–30 Hz) frequency ranges, we used 5-cycle wavelets shifted at every sampled point in bins of 2 Hz. For gamma-band activity (32–100 Hz), 7-cycle wavelets sampled in steps of 2 Hz were employed. To represent any physiological response shape, the convolution GLM approach considers a generic basis set (e.g., sines and cosines). Regressors were constructed by convolving a set of basis functions with a set of input functions representing the events of interest. We chose a 20th-order Fourier basis set (comprising 40 basis functions: 20 sines and 20 cosines). Each discrete and parametric regressor was convolved with this set. This setting allowed the GLM to resolve modulations of TF responses up to approximately 8.7 Hz (20 cycles / 2.3 seconds, or ~ 115 ms). See further details in the original methodological publication [79].

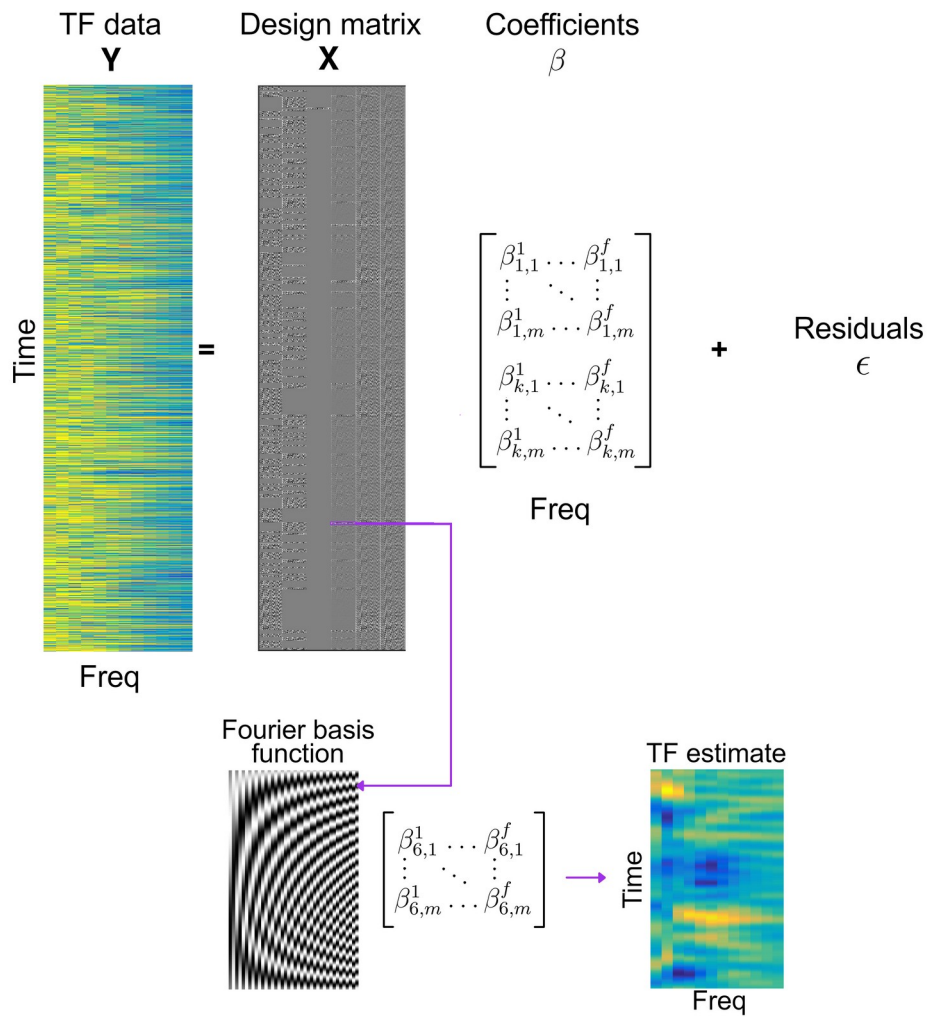


Figure S3. Convolution General Linear Model (GLM). This figure illustrates the standard pseudo-continuous time-frequency (TF) representations of the source-reconstructed MEG signal (Y), estimated using Morlet wavelets. In the GLM framework, the signal Y is modelled as a linear combination of explanatory variables or regressors (contained in matrix X), modulated by regression coefficients (β), with an additional noise term (ϵ). The design matrix X , depicted in the lower left inset, is constructed by convolving Fourier functions ($m = 20$; comprising 20 sine and 20 cosine functions) with input functions that represent the timings and values of relevant discrete or parametric events. For this example, six regressors were used (from left to right in the figure): Outcome Win, Outcome Lose, Outcome Error/No Response, and absolute precision-weighted prediction error (pwPE) on level 2, σ_2 , and σ_3 , each defined over time. The solution to the convolution GLM yields response images (represented by the TF estimate in arbitrary units in the figure) as a combination of the basis functions m and the regression coefficients β_i for each specific regressor type i , across frequencies f and basis functions m . Consequently, the convolution GLM effectively estimates deconvolved TF responses (shown as the TF estimate in the rightmost bottom image) to event types and their associated parametric regressors. See details in Litvak and colleagues [79].

Frequency-resolved connectivity

To assess functional connectivity between frequency-resolved activity in our reconstructed sources, we employed a robust metric of directed information flow between time series: time-reversed Granger causality (TRGC, [82,83]). This metric builds upon Granger causality (GC), a statistical method used to determine directional information flow between signals [132]; frequency domain: [133]). GC operates on the principle that, if information from the past of signal $x(t)$ improves the prediction of the present of signal $y(t)$ relative to what can be predicted by the past of $y(t)$ alone, then $x(t)$ is said to 'Granger-cause' $y(t)$.

Simulation work demonstrated that GC is influenced by volume conduction and source leakage, affecting its robustness [82]. To address these limitations, TRGC was developed. Haufe and colleagues [82] proposed using time-reversed data as surrogates for statistical testing when assessing GC between pairs of time series in a multivariate setting. Spurious directional flow between time series, explainable by volume conduction and corresponding to zero-lag directional influences, is expected to persist in the time-reversed analysis. By correcting the GC coefficients with those from the time-reversed GC analysis, an unbiased estimate of directional flow is obtained, one not influenced by volume conduction or source leakage. This approach has been validated through simulation work, establishing TRGC as a robust metric for inferring the correct directionality of information flow. In addition to mitigating the influence of hidden common drivers, TRGC can mitigate the influence of measurement noise [134,135].

In our study, we estimated frequency-domain TRGC to assess directional information flow between our 16 ROIs (8 bilateral ROIs). Following the recommendations of Pellegrini et al. [83], we used the source-reconstructed time series in our ROIs obtained through LCMV beamforming. This process included a dimensionality reduction step with PCA/SVD to select a singular vector associated with the direction of maximal power in each source (refer to the *Source reconstruction of MEG signals* section). We focused the TRGC analysis on the alpha and beta frequency ranges, and assessed the directionality of functional coupling during the 0.5–1 s interval of outcome processing for trials with larger unsigned pwPEs updating beliefs at level 2 (median split per participant). The TRGC analysis was implemented using the ROIconnect plugin for EEGLAB in MATLAB from ref. [83], available at <https://github.com/sccn/roiconnect>. We adapted the code for application to our LCMV output obtained in MNE-python.

Statistical analyses

To address the multiple comparisons problem, which arises in contexts such as several post-hoc analyses, we control the false discovery rate (FDR) using an adaptive linear step-up procedure [136], set to a level of $q = 0.05$. This provides an adapted threshold p-value (P_{FDR}). In the case of pairwise statistical analyses, we

provide estimates of non-parametric effect sizes for pairwise comparisons, along with associated bootstrapped confidence intervals [84,85]. The between-group effect sizes are estimated as the probability of superiority (Δ).

Supplemental Results

Demographics

Age and sex distribution were comparable across groups (Age: $P = 0.4921$, no-significant differences; $BF_{10} = 0.3472$, providing anecdotal evidence for H_0 ; Sex: Chi-squared statistic, $\chi^2 = 1.6559$, $df = 1$, $P = 0.1849$, no significant; $BF_{10} = 1.102$, no evidence for H_0 or H_1).

BD and HC groups did not significantly differ in their anxiety or depression scores (BDI score, $P_{FDR} = 0.1038$, $BF_{10} = 0.8668$, anecdotal evidence for H_0 ; State-Trait Anxiety Inventory, state subscale: $P_{FDR} = 0.4571$, $BF_{10} = 0.3833$, anecdotal evidence for H_0 ; HADS, depression subscale: $P_{FDR} = 0.4003$, $BF_{10} = 0.3824$, anecdotal evidence against group differences; anxiety, $P_{FDR} = 0.64$, $BF_{10} = 0.31$, substantial evidence for the null hypothesis). Altman's Mania scores were also similar between groups ($P_{FDR} = 0.2040$, $BF_{10} = 0.5652$). Refer to **Table 1**.

Regarding cognitive abilities, we identified significant group differences in the second part of the TMT, assessing executive functioning (independent-sample permutation test, $P_{FDR} = 0.0022$; non-parametric effect size estimator: $\Delta = 0.80$, $CI = [0.63, 0.92]$). By contrast, BD and HC participants performed similarly in the less challenging first TMT part ($P_{FDR} = 0.2216$, $BF_{10} = 0.5296$, providing anecdotal evidence for H_0). Differences in Wisconsin Card Sorting Test were not significant after FDR control ($P = 0.0364$; non-significant, as the adjusted significance level was 0.0022; $BF_{10} = 1.6560$, anecdotal evidence for H_1). Last, both samples had similar MMSE scores ($P = 0.8678$, $BF_{10} = 0.2875$, providing substantial evidence for the null hypothesis).

Altered reward-based decision dynamics in bipolar disorder during euthymia

BD showed changes in the expression of win-stay/lose-shift behaviour relative to HC (see main text). Notably, there was no between-group difference in their decision to switch strategies post-loss, based on anecdotal evidence (lose-shift rate: $P = 0.0966$, non-significant difference; $BF_{10} = 0.8905$; **Figure 2d**). Overall, when considering all types of switches, BD individuals changed their responses more than healthy participants (total switch rate, 0.16 [0.019] in HC, 0.23 [0.049] in BD; $P_{FDR} = 0.0160$; $\Delta = 0.70$, $CI = [0.55, 0.85]$). Both groups committed errors at a similar rate during the task performance (around 1%;

performance errors: $P = 0.1806$, non-significant difference; $BF_{10} = 0.5812$, anecdotal evidence for H_0 ; timeouts: $P = 0.5161$, non-significant difference; $BF_{10} = 0.3557$, anecdotal evidence for H_0).

We next examined these behavioural results further using the HGF as a computational framework. Applying Bayesian model selection to models M1-M4 in the total sample, we found that the M3 model outperformed the other models across all subjects (exceedance probability, $P_{exc} = 0.77$; expected frequency, $Freqexp = 0.45$; **Table S4**). Model M3 was also the winning model separately for HC group ($P_{exc} = .52$, $Freqexp = 0.41$), and BD group ($P_{exc} = .70$, $Freqexp = 0.43$). In the M3 response model, decisions depend on dynamic trial-by-trial estimates of log-volatility. A heightened expectation of environmental volatility on the current trial results in higher decision noise, implying a noisier belief-response mapping. Conversely, when the environment is anticipated to be more stable, the coupling between the current belief and the ensuing decision becomes more deterministic. This model also allows for individual differences in ω_2 , the tonic volatility of action-outcome associations, as well as time-invariant volatility on the third level, ω_3 ; free model parameters also include the initial values $\mu_3^{(0)}$ and $\sigma_3^{(0)}$.

Table S4. Bayesian model selection (BMS). Log model evidence was used for model comparison using random-effects Bayesian model selection [137]. The columns display the exceedance probability, P_{exc} , and expected frequency, $Freqexp$, for models M1-M4. Model M3 best explained the data in our total sample ($N = 49$), and in each group separately (HC, $N = 27$; BD, $N = 22$). BMS was implemented using code from the MACS toolbox [138].

Model, Sample	P_{exc}	$Freqexp$
M1, HC+BD	0	0.0239
M2, HC+BD	0.1920	0.3226
M3, HC+BD	0.7701	0.4484
M4, HC+BD	0.0379	0.2050
M1, HC	0	0.04
M2, HC	0.44	0.38
M3, HC	0.52	0.41
M4, HC	0.04	0.17
M1, BD	0	0.05
M2, BD	0.12	0.26
M3, BD	0.7	0.43
M4, BD	0.19	0.26

Simulations conducted to assess the accuracy of parameter estimation in the winning HGF model M3 demonstrated that parameters ω_2 and $\mu_3^{(0)}$ were estimated with high accuracy. By contrast, ω_3 and $\sigma_3^{(0)}$ were recovered with less accuracy, as previously reported [32,59]. See **Figure S4** for details. Specifically, we simulated behavioural responses of 50 agents across a 6 x 6 grid of ω_2 and ω_3 values, using the input observed for one of our participants (ID 47). For each ω_2 (ω_3) value, this resulted in 300 simulated agents, derived from 50 simulations at each of the six ω_3 (ω_2) values. To assess the estimation accuracy of parameters $\mu_3^{(0)}$ and $\sigma_3^{(0)}$, we conducted similar simulations with 50 agents across a 6 x 6 grid of $\mu_3^{(0)}$ and $\sigma_3^{(0)}$ values.

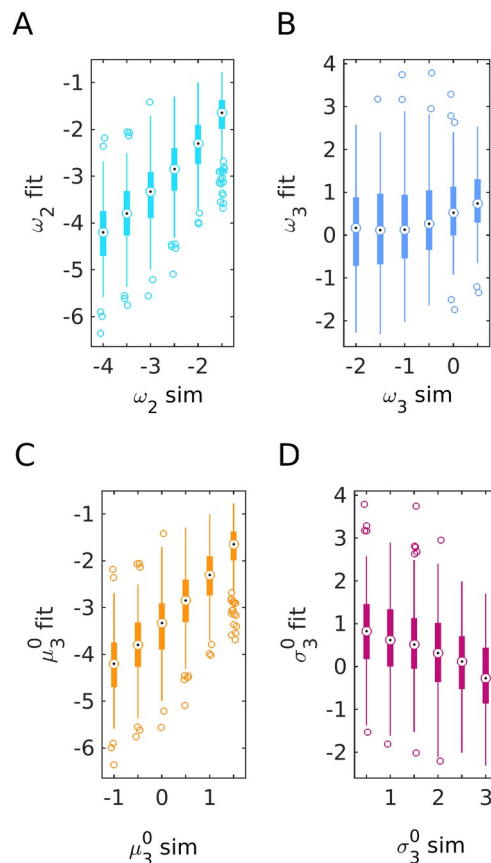


Figure S4. HGF parameter recovery. This figure illustrates the estimation of HGF parameters using the input observed from one of our participants (ID 47). Panels **a-d** display boxplots (showing the median, 25th, and 75th percentiles) of parameter estimation results for ω_2 (a), ω_3 (b), $\mu_3^{(0)}$ (c), and $\sigma_3^{(0)}$ (d). The x-axis represents the parameters set in the simulated responses (labelled "sim"), and the y-axis shows the corresponding estimated values of those parameters (labelled "fit"). Parameters ω_2 and $\mu_3^{(0)}$ were estimated with high accuracy, indicated by a significant correlation between simulated and estimated (fit) values: Pearson $R = 0.8158$, $P \ll 1 \times 10^{-6}$ for ω_2 , and $R = 0.8049$, $P \ll 1 \times 10^{-6}$ for $\mu_3^{(0)}$. By contrast, parameter ω_3 showed less accuracy in estimation: $R = 0.2761$, $P \ll 1 \times 10^{-6}$, as did $\sigma_3^{(0)}$: $R = -0.0093$, $P = 0.6920$. The prior values for ω_2 , ω_3 , $\sigma_3^{(0)}$, and $\mu_3^{(0)}$ used in the configuration file for estimating each parameter from

the simulated responses, as defined in **Table S1** for the best fitting model M3, were: $\omega_2 = -2.6$, $\omega_3 = -0.3$ (with a variance of 4 for both); $\mu_3^{(0)} = 1$, $\sigma_3^{(0)} = 1$ (with a variance of 1 for both in log space).

Using the winning model to assess between-group differences in the model parameters ω_2 , ω_3 , $\mu_3^{(0)}$, as well as in the mean log-volatility estimate over trials, μ_3 , we reported in the main text significant differences in ω_2 , ω_3 , $\mu_3^{(0)}$, μ_3 . Initial volatility and its trial-average were higher in BD than in HC, while ω_2 was lower. Parameter ω_2 influences the coupling of beliefs between levels 2 and 3 through changes in environmental uncertainty, defined as $\exp(\kappa\mu_3^{(k-1)} + \omega_2)$ [27]. Thus, despite BD patients perceiving the environment as more volatile, their belief updates about reward contingencies were not as influenced by this expectation compared to the HC group. As expected, log-volatility estimates correlated with the total switch rate, accounting for response stochasticity (**Figure S5**; Spearman $\rho = 0.79$, 95% CI [0.65, 0.88], $P = 1.7 \times 10^{-11}$). A similar result was obtained within the BD sample ($\rho = 0.69$, 95% CI [0.37, 0.86], $P_{FDR} = 3.8 \times 10^{-4}$), and HC sample ($\rho = 0.74$, 95% CI [0.49, 0.88], $P_{FDR} = 1.1 \times 10^{-5}$).

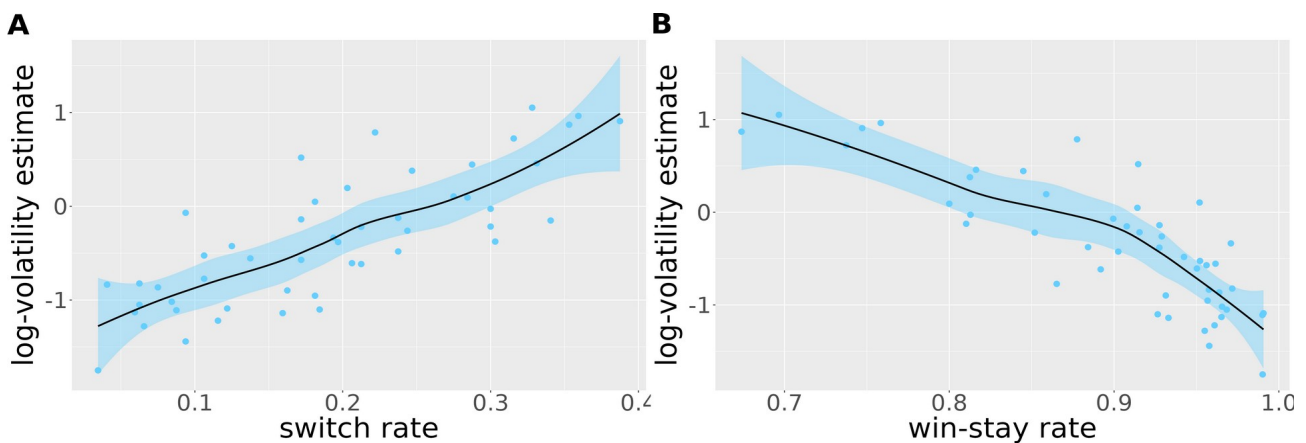


Figure S5. Non-parametric rank correlation between phasic log-volatility and relevant response rates. a. There was a significant positive Spearman rank correlation between the total switch rate and the average phasic log-volatility, μ_3 ($N = 49$ joint sample; $\rho = 0.79$, 95% CI [0.65, 0.88], $P_{FDR} = 1.7 \times 10^{-11}$). Individual participant data points are depicted as blue circles. The black line indicates the non-linear fit estimated using the Locally Weighted Scatterplot Smoothing (LOESS) method via the `geom_smooth()` function in R, employing the `ggplot2` library. The graphic also features the 95% confidence interval as a shaded blue area. **b.** Same as **(a)** but for the win-stay rate, which was found to be significantly higher in BD than HC participants in the main analysis (**Figure 2d**). There was a significant negative non-parametric rank correlation between the win-stay rate and μ_3 ($N = 49$, $\rho = -0.80$, 95% CI [-0.89, -0.67], $P_{FDR} = 1 \times 10^{-16}$). See text.

Next, in light of our key behavioural finding that BD participants were less likely to stay after winning (**Figure 2d**), we assessed whether the win-stay rate was associated with μ_3 . Non-parametric rank correlation analysis revealed that the win-stay rate and μ_3 were highly negatively correlated (Spearman $\rho = -0.80$, 95%

CI [-0.89, -0.67], $P_{FDR} = 1 \times 10^{-16}$), indicating that **individuals with higher estimates of log-volatility were less likely to repeat a response that just won**. When assessing this correlation specifically within the BD group, the results were very similar ($\rho = -0.81$, 95% CI [-0.92, -0.58], $P_{FDR} = 5.3 \times 10^{-6}$). In the HC group, the correlation results were also consistent ($\rho = -0.69$, 95% CI [-0.85, -0.41], $P_{FDR} = 1 \times 10^{-4}$). Last, there were no significant differences in tonic volatility on level 3, ω_3 ($P = 0.0730$, anecdotal evidence suggesting group differences, $BF_{10} = 1.09$).

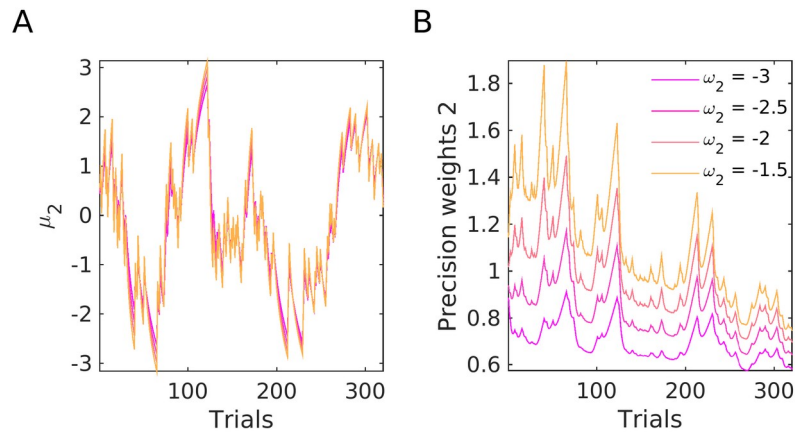


Figure S6. Simulated belief trajectories and precision weight terms on level 2. **a.** In the winning model, M3, belief trajectories for μ_2 were simulated using input data from one participant (ID 47), with priors set at $\mu_2^{(0)} = 0$, $\sigma_2^{(0)} = 0.1$, $\mu_3^{(0)} = 1$, $\sigma_3^{(0)} = 1$, $\kappa = 1$, $\omega_3 = -0.5$, but with ω_2 being modulated. This parameter, ω_2 , represents the tonic part of the variance in the Gaussian random walk for x_2 and modulates the learning rate about response outcomes at the lowest level. We demonstrate that decreasing ω_2 (represented by changes from orange to magenta lines) is associated with smaller update steps in the belief about the action-outcome contingency tendency, μ_2 . **b.** In line with (a), decreasing ω_2 also reduces the estimation uncertainty about the reward tendency, σ_2 . Since σ_2 represents the precision weights term scaling belief updating at level 2, this simulation indicates that lower tonic volatility, ω_2 , has a slowing effect on belief updating about the action-outcome contingency tendency.

Correlations between HGF parameters and residual symptoms

To assess associations between residual symptoms and HGF parameters, we conducted non-parametric Spearman rank correlations, given that residual symptoms are represented as ordinal data. Utilising the R library “correlation” and the method “spearman”, we identified a significant positive correlation between volatility estimates and trait anxiety (**Figure S7a**), as well as a negative association between mania and estimation uncertainty σ_2 (**Figure S7b**). However, no significant correlation was found between depression and σ_2 , as detailed in the main document.

Exploring whether the significant correlations in the BD sample extended to the full sample ($N = 49$ participants), we observed that log-volatility and trait anxiety were indeed significantly correlated ($\rho = 0.31$,

95% confidence interval [0.02, 0.55], $P_{FDR} = 0.033$). Mania scores and σ_2 were, however, not associated in the full sample ($N = 48$; $\rho = -0.17$ [-0.42, 0.08], $P = 0.740$, $BF_{10} = 0.740$).

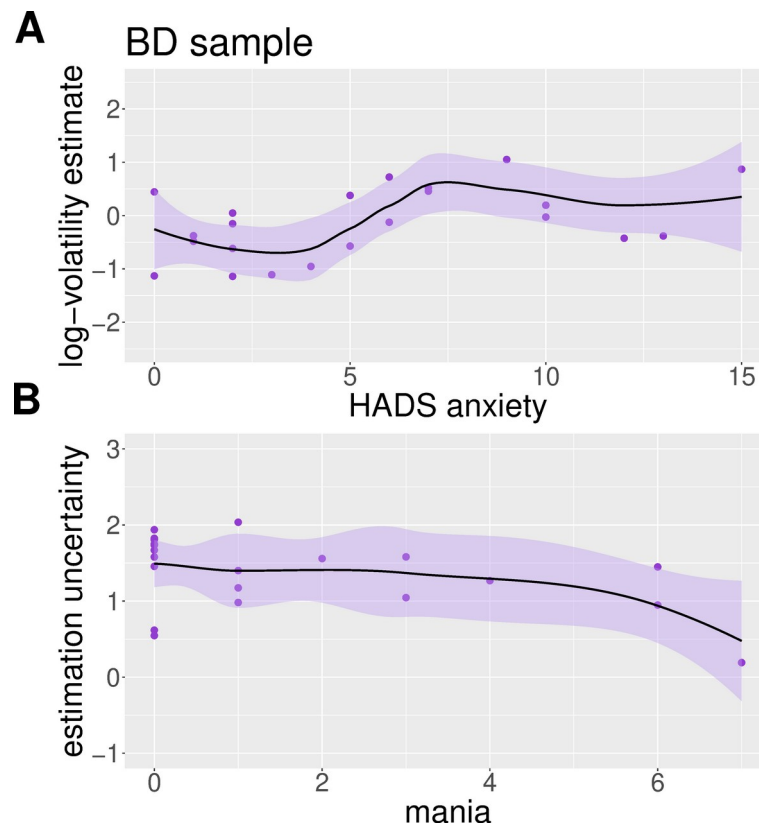


Figure S7. Non-parametric rank correlation between residual symptoms and relevant HGF variables. **a.** The Spearman rank correlation between trait anxiety levels and the mean expectation on log-volatility was positive and significant ($N = 22$; $\rho = 0.46$, 95% confidence interval [0.04, 0.75], $P_{FDR} = 0.030$). The purple circles denote individual BD values, while the black line illustrates the non-linear fit estimation (LOESS). The 95% confidence interval is represented as a shaded purple area. **b.** Same as a but variables σ_2 , estimation uncertainty, and mania scores in BD. There was a significant negative association between these variables ($N = 21$; $\rho = -0.46$ [-0.75, -0.02], $P_{FDR} = 0.037$; mania scores were not available in one patient).

Behavioural results relative to an ideal Bayesian observer

To assess whether deviations in decision-making behaviour in BD could be ascribed to larger departures from the expected learning patterns of an ideal Bayesian observer than in HC, we simulated the behavioural responses of agents observing the input of each participant in both cohorts. Using the same priors as in the primary modelling analysis, we extracted the time series of behavioural responses in each agent and computed the simulated win rates, win-stay and lose-shift rates (**Figure 2cd**).

First, we confirmed that these rates were comparable between two groups of ideal observers, representing HC and BD. This demonstrated that the pseudorandomised task structure was balanced across groups ($P = 0.2496, 0.5663, 0.07$; $BF_{10} = 0.5200, 0.3295, 0.5450$: substantial and anecdotal evidence supporting no differences for win rates, win-stay and lose-shift rates, respectively).

Subsequently, we found that both win rates and win-stay rates were significantly lower in the participant groups compared to the simulated values, as expected. This was the case for both HC and BD (HC: $P_{FDR} = 0.0004$ for win rate, 0.0002 for win-stay rate; BD: $P_{FDR} = 0.0002$ for both metrics). Lower win-stay rates in our participant groups than in the simulated agents indicated that they were more likely to change after securing a win on a trial.

In terms of lose-shift behaviour, healthy participants exhibited rates akin to those of ideal Bayesian observers ($P = 0.1076$, with the difference being non-significant; $BF_{10} = 0.6775$, offering anecdotal evidence for H_0). Conversely, BD patients demonstrated a more pronounced tendency to shift after a loss compared to their ideal Bayesian learner counterparts, which was significant ($P_{FDR} = 0.0004$).

Control of medication effects on computational variables used for correlation analyses and main between-group analyses

In our BD cohort, none of the included patients received therapy with benzodiazepines. In addition, despite the relatively high representation of antipsychotics in the therapeutic regimens, high-potency D2-receptor blockers (such as Haloperidol and Amisulprid; [139]) were absent among them, except for two participants, who were taking small doses of Zuclopenthixol and Flupentixol. The most widely used were either antipsychotics with low affinity for dopamine receptors (e.g., Quetiapine) or partial D2 receptors agonists (e.g., Aripiprazole). Taken together, the findings of our study have low risk of being explained by the effect of medications on the dopaminergic system in the BD group. However, it is important to assess this explicitly, therefore we conducted control analyses.

To assess potential medication effects on these associations, we performed exploratory analyses comparing these variables among BD participants based on their medication type. Specifically, we compared those with and without antipsychotic medication ($N = 14$ vs 8). Antipsychotic medications included drugs such as Aripiprazole, Quetiapine, Flupentixol, Cariprazine, Olanzapine, Zuclopenthixol, and Levomepromazine. We also compared BD subgroups based on the effect of the drug on dopaminergic transmission: those taking dopamine-blocking/modulating drugs ($N = 14$) and those taking drugs with no direct or some agonistic dopaminergic effects, including unmedicated participants ($N = 8$). Importantly, the subgroups were identical, whether splitting participants by antipsychotic use or dopamine-blocking/modulating effects. This is because one participant on Aripiprazole was also taking Amantadine (a dopamine modulator), and

another participant on Aripiprazole was taking Biperiden, which indirectly affects dopamine by balancing it with acetylcholine.

These control analyses follow the approach used by Pizzagalli et al. [140], who conducted similar analyses to account for potential dopaminergic-blocking effects of drugs. Our results revealed no significant differences between subgroups in any of the variables used for correlation analyses:

Table S5. Between-group statistical results assessing BD participants either taking dopamine-blocking/inhibiting drugs (N = 14) or taking drugs with no direct or some agonistic dopaminergic effects (N = 8).

Variable	with vs. without antipsychotic medication (14 vs 8) / with vs without dopamine-blocking/modulating drugs (same 14 vs 8)
μ_3	$P = 0.5441, BF_{10} = 0.4507$
σ_2	$P = 0.9508, BF_{10} = 0.3955$
HADS anxiety	$P = 0.5441, BF_{10} = 0.4365$
Mania	$P = 0.6341, BF_{10} = 0.4281$
HADS depression	$P = 0.7397, BF_{10} = 0.4091$

Reported p-values are not-corrected for multiple comparisons.

Similar results were obtained when comparing subgroups of participants with or without antidepressants or mood stabilisers (non-significant differences between subgroups, $P \gg 0.05$ in all cases, anecdotal and substantial evidence for H_0).

In addition, if we were only to consider participants not taking antipsychotics (N = 8) and assess their differences with respect to healthy participants in the main computational variables used in the manuscript (μ_3, ω_2) we obtain that in the BD subset the expectation on volatility is higher ($P = 0.0446$, uncorrected), and tonic volatility is lower ($P = 0.0354$) as in the main analysis with the full sample (but note the small sample size of N = 8 BD participants in this cohort).

Expectation about reward probability invigorates motor performance similarly in euthymic bipolar and healthy participants

Expectations about the tendency of the action-outcome probability similarly influenced performance tempo in BD and HC groups. BML assessing the association between performance tempo (logmIKI) and $|\hat{\mu}_2^{(k)}|$ revealed that three tested models (models 5–7) explained the data equally well, as indicated by LOO-CV (Table S2).

When ELPD differences between two models are larger than four, and if the number of observations is > 100, and the model is moderately well specified, then the standard error is a good estimate of the uncertainty in the difference between models [134,141]. The smallest difference in ELPD between models 5–7 was less than twice the standard error ($2 * se_diff$), leading us to select the most parsimonious model, model 5. This model, which did not include group differences or trial-specific random effects, effectively explained performance tempo through the strength of participants' predictions about reward contingencies ($|\hat{\mu}_2^k|$), while also accounting for individual variability on the slope and intercept (subject included to model random effects). Posterior point estimates for the effects in the best model are shown in **Table S6**.

Posterior predictive checks confirmed that the observed outcome variable y (tempo: logmIKI) overlapped well with the simulated datasets y^{rep} from the posterior predictive distribution (**Figure S8a**).

Table S6. Summary parameter estimates for the winning Bayesian multilevel model assessing the effect of strength of predictions on timing performance.

Dependent Variable	Parameter (Population-level effects)	Estimate	l-95% CrI	u-95% CrI	R-hat
Performance tempo (logmIKI, in log-ms) (model5)					
	Intercept	5.78357	5.71471	5.85564	1.00
	Predictions	-0.00753	-0.01452	-0.00046	1.00
Reaction time (logRT, in log-ms) (model5)					
	Intercept	6.14696	6.06355	6.23095	1.00
	Group (BD-HC)	0.06542	-0.05844	0.18618	1.00
	Predictions	0.01003	-0.00556	0.02522	1.00
	Group:Predictions	-0.00799	-0.03083	0.01492	1.00

Estimate = posterior mean; CrI = credible interval based on quantiles. Gelman-Rubin statistics demonstrate excellent chain convergence (R-hat < 1.01). The effective sample size (ESS) was >> 400 for each parameter estimate, denoting good convergence. The predictor “Predictions” denotes the centred values of the strength of participants' predictions about reward contingencies ($|\hat{\mu}_2^k|$). Credible effects are considered when the 95% lower and upper-bound CrI around the Bayesian point estimate do not include zero (denoted by bold font).

Model 5 demonstrated that stronger predictions led to faster performance in the joint sample, confirmed by the negative slope of this association (with the 95% credible interval [CrI] not including zero, denoting a credible non-zero estimate; **Figure S9ac**).

Conversely, logRT was not modulated by the strength of predictions or other variables (**Table S6**). Model 7, the most complex for RT, emerged as the best based on ELPD criteria. Although this model showed relatively good posterior predictive accuracy, it was slightly lower than the model for tempo. It revealed no credible effects of the strength of predictions on logRT among participants, nor were there group effects, as the 95% CrI for these effects included zero (**Table S6; Figure S8b**)

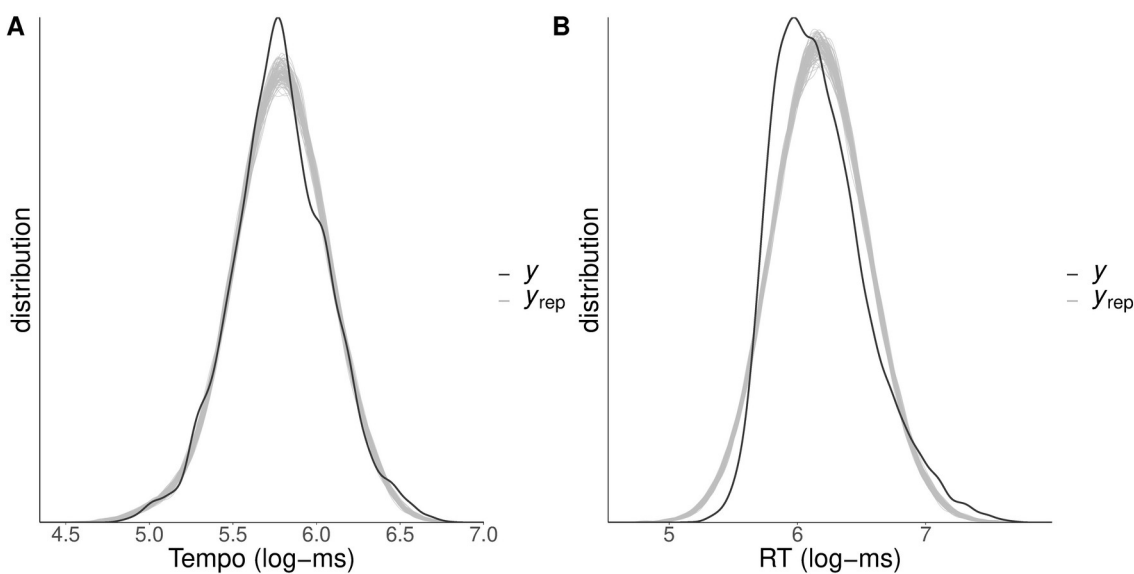


Figure S8. a. Illustration of the posterior predictive checks for logmIKI BML model 5, where the distribution of the observed outcome variable (y , in our case performance tempo, log-transformed) is compared to simulated datasets (y_{rep}) from the posterior predictive distribution (100 draws). b. same as a, but for logRT.

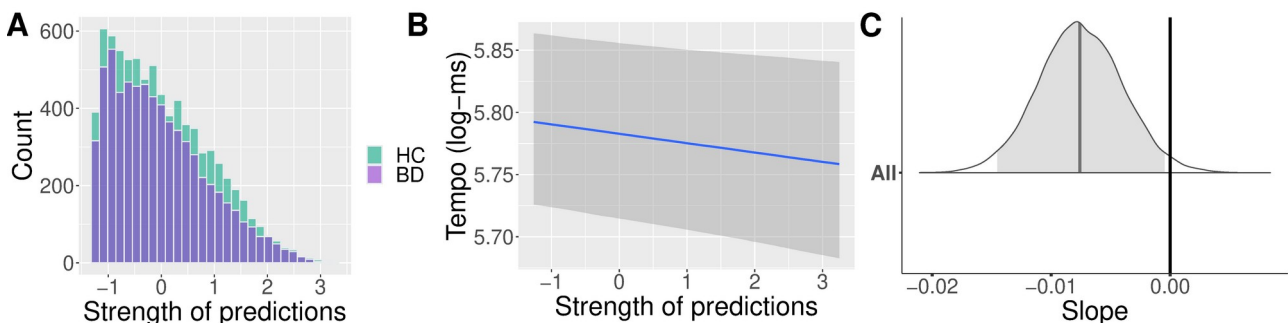


Figure S9. Effects of expectations of action-outcome contingencies on performance timing. a-c. Invigoration of performance tempo by the strength of predictions about the tendency of the action-outcome contingencies.

Performance tempo was measured as the logarithm of the mean inter-key press interval (logmIKI, in log-ms). **a.** Histogram of the strength of predictions about reward contingencies $\left| \hat{\mu}_2^{k|} \right|$ in the BD (purple) and HC (green) groups, showing similar distributions. **b.** Bayesian multilevel modelling demonstrated that performance tempo (logmIKI, in log-ms) was shorter in trials when participants had a stronger expectation on the reward tendency, as the posterior estimate of the slope of this association was negative, and the 95%-CrI did not include zero (**c**). Thus, in the joint cohort, individuals speed up sequence performance when they hold stronger predictions about the action-outcome contingencies. **c.** Illustration of the posterior distribution of the slope estimate, including the posterior point estimate (grey vertical line) and 95%-CrI (denoted by the grey area under the curve).

Control analyses: Assessing practice effects in timing performance

Control analyses evaluated practice effects in timing performance, excluding the effect of reward contingency predictions (**Table S3**). The most complex model, model 8, best described changes in timing performance (logmIKI and logRT: **Figure S10a, d**) across trials for both DVs (**Table S7**). For both DVs, the absolute mean difference in ELPD between the model 8 and the second-best fitting model (model 7) was larger than twice the standard error of the differences ($\text{elpd_diff} > 2 * \text{se_diff}$).

This model described changes in motor performance (either logmIKI or logRT) across trials for the reference group of healthy controls, captured by the fixed effect of the trial. The modulation of these practice effects by group was represented by the interaction term $\text{group} * \text{trial}$. This interaction reflects the differences in slopes between BD and HC groups, illustrating how changes in timing across trials varied between groups. In addition, the inclusion of the random effects term ($1 + \text{trial} | \text{subject}$) in the structure of this model allows for both the intercept (baseline DV) and the slope (change in DV across trials) to vary by subject, directly capturing the individual differences in how practice affects tempo or RT.

Posterior predictive checks indicated that the best model for tempo demonstrated robust predictive accuracy across the range of the DV (**Figure S11a**). The predictive strength of the best model for RT was, however, lightly diminished (**Figure S11b**). **Table S7** presents a summary of the posterior point estimates for the winning model.

Table S7. Summary parameter estimates for the winning Bayesian multilevel model assessing practice effects.

Dependent Variable	Parameter (Population-level effects)	Estimate	l-95% CrI	u-95% CrI	R-hat
Performance tempo (logmIKI, in log-ms) (model8)					
	Intercept	5.79014	5.68824	5.89292	1.02

	Group (BD - HC)	0.08536	-0.04126	0.21718	1.02
	Trial	-0.00008	-0.00012	-0.00003	1.00
	Group:Trial	-0.00022	-0.00029	-0.00016	1.00
Reaction time (logRT, in log-ms) (model8)					
	Intercept	6.21087	6.11646	6.30298	1.00
	Group (BD - HC)	0.11165	-0.02092	0.24376	1.00
	Trial	-0.00037	-0.00063	-0.00013	1.00
	Group:Trial	-0.00031	-0.00067	0.00005	1.00

Estimate = posterior mean; CrI = credible interval based on quantiles. Gelman-Rubin statistics demonstrate excellent chain convergence in most parameters ($R\text{-hat} < 1.01$; [123]). The effective sample size (ESS) was $\gg 400$ for each parameter estimate, denoting good convergence.

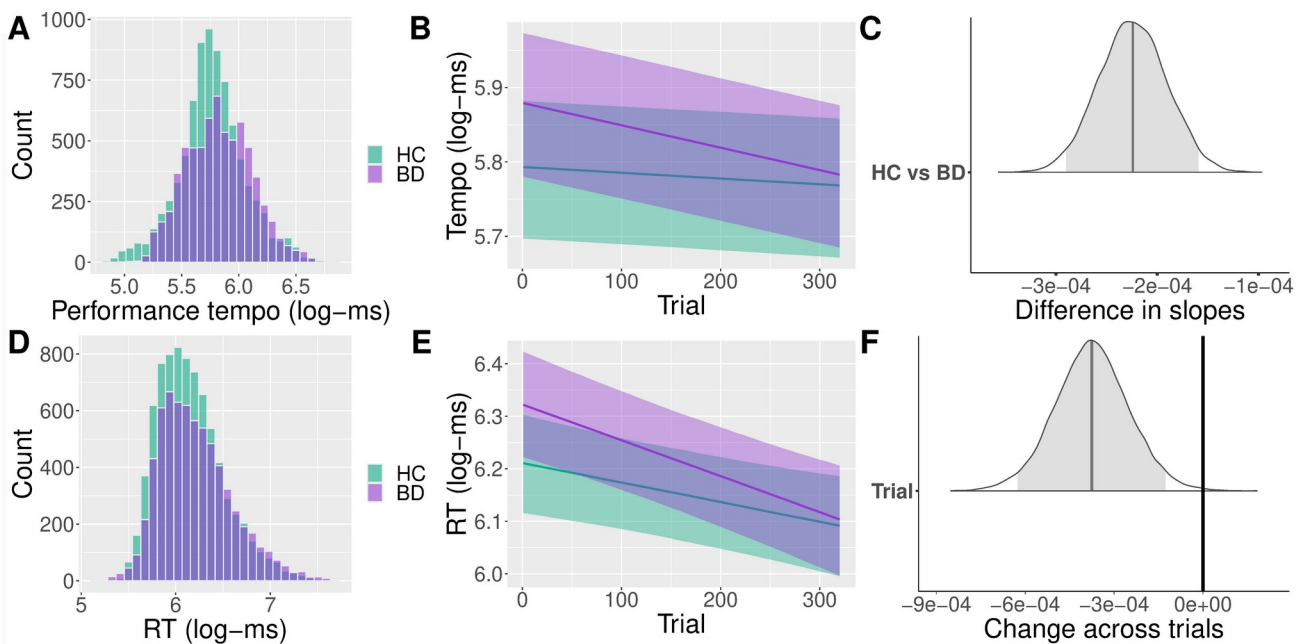


Figure S10. Effects of practice on performance timing. a-c. Practice effects on performance tempo, measured as the logarithm of the mean inter-key press interval (logmIKI, in log-ms). a. Distribution of logmIKI (“tempo” in the graphics) in the BD (purple) and HC (green) cohorts. b. Output of the Bayesian multilevel model that provided a better fit to the data (see main text). The graphic represents the association between trial number and performance tempo, represented as logmIKI, in each cohort. The bold lines and shaded areas denote the posterior mean and the 95% CrI (shaded areas) in each group. In the reference HC group, participants sped up their performance across trials. c. Illustration of the posterior distribution of the between-group difference in slopes, including the posterior point

estimate (grey vertical line) and 95%-CrI (denoted by the grey area under the curve). The BD minus HC slope difference was negative, and the 95%-CrI did not overlap with zero, supporting that performance in the BD group sped up more throughout the task. **d-e**. Same as a-c but for the logarithm of reaction time, logRT (in log-ms). The reference HC group also demonstrated a reduction in RT over trials (panel e). The 95% CrI of the between-group slope difference included zero, indicating a credible effect for a similar change over time in RT in both groups. **f**. Posterior distribution of the fixed effect of trials, showing negative changes in RT over time. The black vertical line represents a zero-change difference, which did not overlap with the 95% CrI of the estimate, suggesting a 95% probability of a negative change in RT over trials (speeding up).

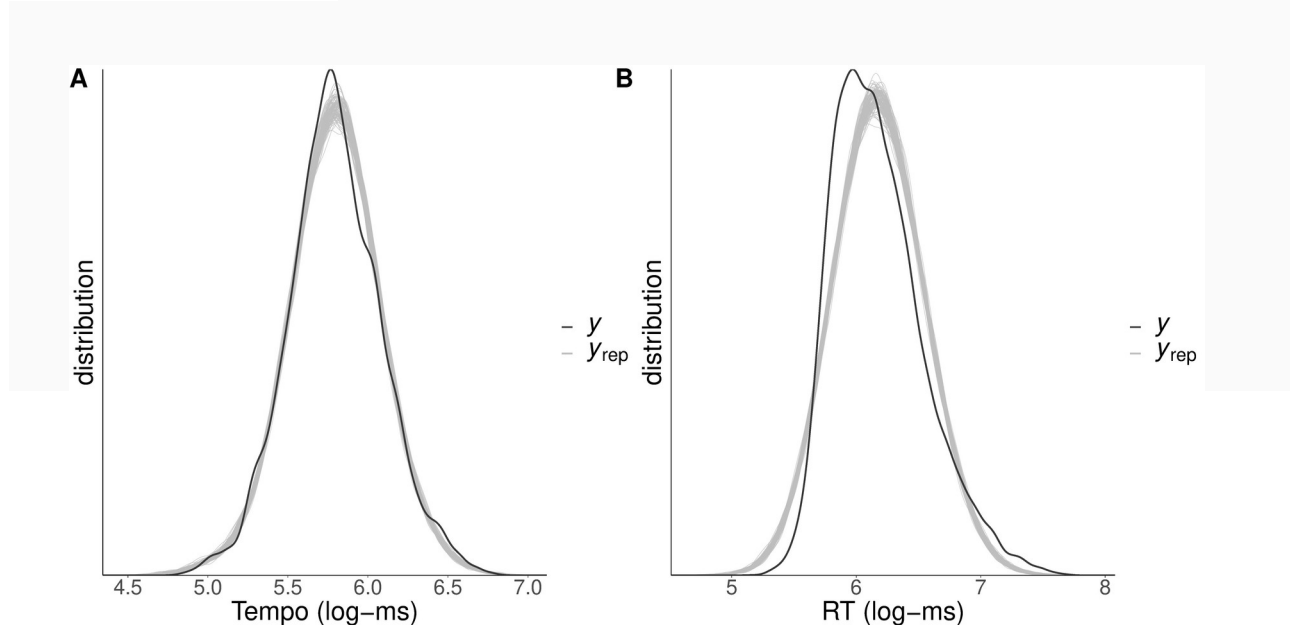


Figure S11. a. Illustration of the posterior predictive checks where the distribution of the observed outcome variable (y , in our case performance tempo, log-transformed) is compared to simulated datasets (y_{rep}) from the posterior predictive distribution (100 draws). **b.** same as a, but for logRT.

The results of this analysis revealed that both groups started with comparable timing (logmIKI and logRT: the 95% credible interval, CrI, for the difference in intercept estimates overlapped with zero). In addition, for both DVs there was a credible fixed effect of trial number on changes in timing performance (HC as reference group; negative slope: the 95% CrI of the estimated slope for HC did not include zero, denoting a meaningful effect). This means that performance timing (tempo and RT) improved across trials in HC.

Notably, the BD group exhibited more rapid improvements in their performance tempo across trials (steeper slope, negative slope difference: **Figure S10bc**) than their healthy counterparts (the 95% CrI for the estimate of slope differences did not encompass zero). This was not the case for RT (no credible effect for slope differences between groups; **Figure S10ef; Table S7**).

Complementing these findings, baseline motor performance in bipolar participants was slower compared to healthy individuals when playing simple sequences upwards and downwards (the average performance

tempo was 266 [12.2] ms in HC, 323 [19.1] ms in BD, $P = 0.02$; $\Delta = 0.73$, CI = [0.54, 0.87]). This outcome suggests that when playing a non-rewarded simple motor task, BD participants were slower than HC.

In conclusion, while baseline motor performance varied between groups, the timing of actions during the main motor decision-making task was similar across both groups. Nevertheless, BD participants displayed a more pronounced slope for practice effects on tempo.

Neural representation of pwPE updating beliefs about the action-outcome contingencies

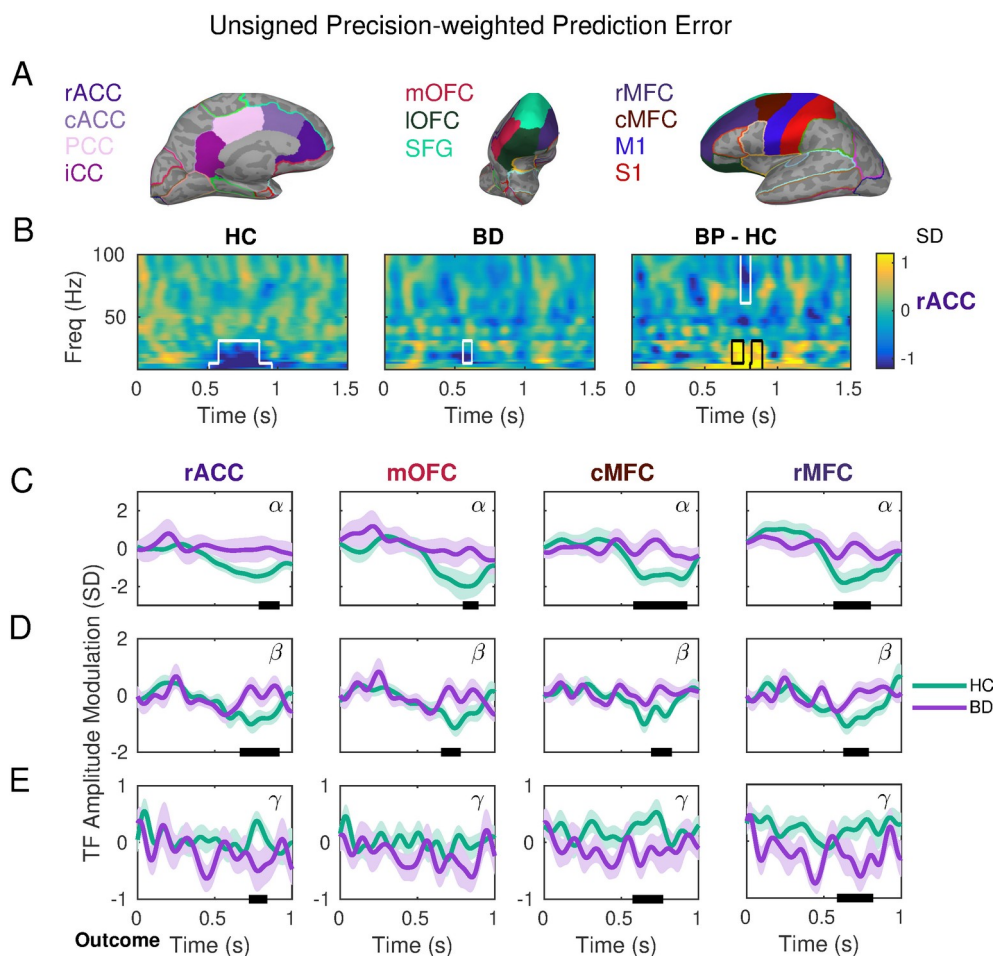


Figure S12. Attenuated gamma increase and alpha-beta suppression during encoding unsigned precision-weighted prediction errors about stimulus outcomes in bipolar disorder. Same as Figure 3 but in additional ROIs where between-subject differences were observed after FWER control at 0.05 (two-sided tests, effects considered if $P_{FWER} < 0.025$): rACC, mOFC, caudal and rostral MFG. Labels denote the rostral anterior cingulate cortex, rACC; caudal ACC, cACC; superior frontal gyrus, SFG; lateral and medial orbitofrontal cortex, IOFC and mOFC; primary motor cortex, M1; caudal and rostral middle frontal gyrus, cMFG, rMFC.

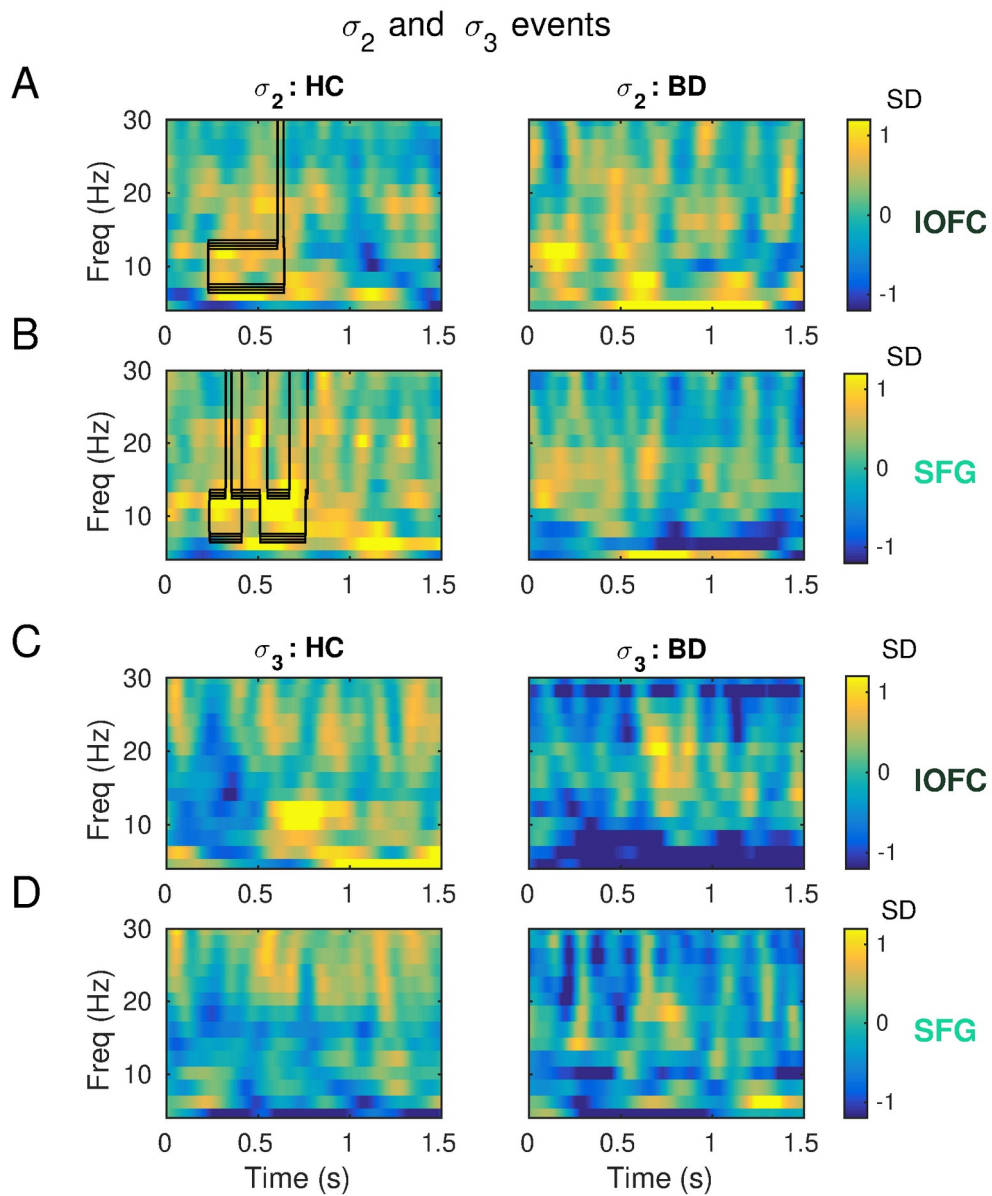


Figure S13. Modulation of alpha and beta activity by uncertainty regressors in each group. a-d. Same as Figure 3, but showing alpha and beta activity modulation by uncertainty regressors σ_2 and σ_3 in healthy control (HC, left) and bipolar disorder (BD, right) participants. For HC participants, significant increases in 8-30Hz activity were observed with estimation uncertainty, σ_2 , in regions cACC, rACC, cMFG, rMFG, IOFG, M1, and SFG ($P_{\text{FWER}} = 0.0130$, effects in regions IOFG and SFG are illustrated in panels a-b). No significant within-subject effects were found in BD, and there were no significant between-group differences after FWER control either ($P = 0.1928, 0.2358$ for σ_2 and σ_3 , respectively). c-d. Regressor σ_3 was not associated with any significant within-subject or between-group differences after FWER control.

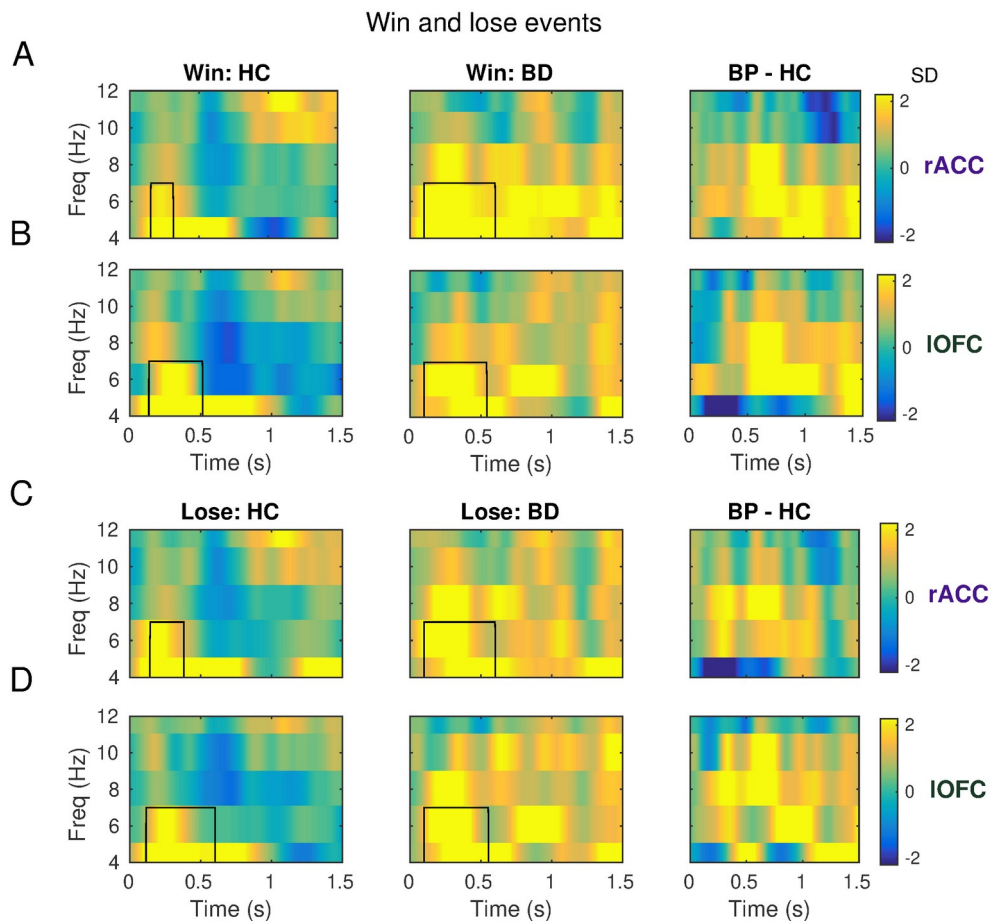


Figure S14. Modulation of theta activity by win and lose events in bipolar and healthy control groups. a-d. Same as **Figure 3** but for the discrete win and lose events. As expected, the discrete win (**ab**) and lose (**cd**) regressors elicited robust and significant increases in theta activity in both participant groups from 0.1 to 0.4 s ($P_{FWER} = 0.0240, 0.0010$ for win events and HC and BD, respectively; $P_{FWER} = 0.0070, 0.0010$ for lose events in each group). These effects were more pronounced in the caudal and rostral ACC, caudal and rostral MFG, IOFC, and mOFC, and extended for at least 200 ms (one cycle at 5 Hz) in these regions. The effects lasted longer in the bipolar group, extending from approximately 0.1 to 0.6 s. Bipolar and healthy control groups did not differ significantly with regard to the increase in theta activity in response to win or lose events ($P = 0.1808, 0.3397$ for win and lose regressors, respectively).

Changes in Raw Power Spectral Density During the Inter-Trial Interval

The convolution modelling results from outcome processing revealed that BD patients exhibited a less pronounced suppression in the alpha and beta bands when encoding pwPEs. In a post-hoc exploratory analysis, we aimed to determine whether this attenuated suppression of alpha and beta activity might be associated with a smaller dynamic range of spectral power in these frequency bands in BD during task performance, namely, a reduced alpha and beta raw power.

To investigate this, we analysed the raw power spectral density (PSD, in fT) during the inter-trial intervals (ITI), a period when participants were at rest, awaiting the next stimulus to make a new choice and perform

the corresponding action sequence. We extracted ITI epochs between 1.7 and 0 seconds preceding the stimulus presentation and conducted LCMV as previously described. The data covariance matrix was estimated from -1.5 to 0 s, and the noise covariance matrix from -1.7 to -1.5 s.

The analysis revealed a significant group effect on the ITI PSD, attributed to a pronounced reduction in the low beta PSD in BD compared to HC (relative suppression in the 13–20 Hz range, $P_{FDR} = 0.006$). This effect was observed across the lateral and medial OFC, and cACC and rMFG.

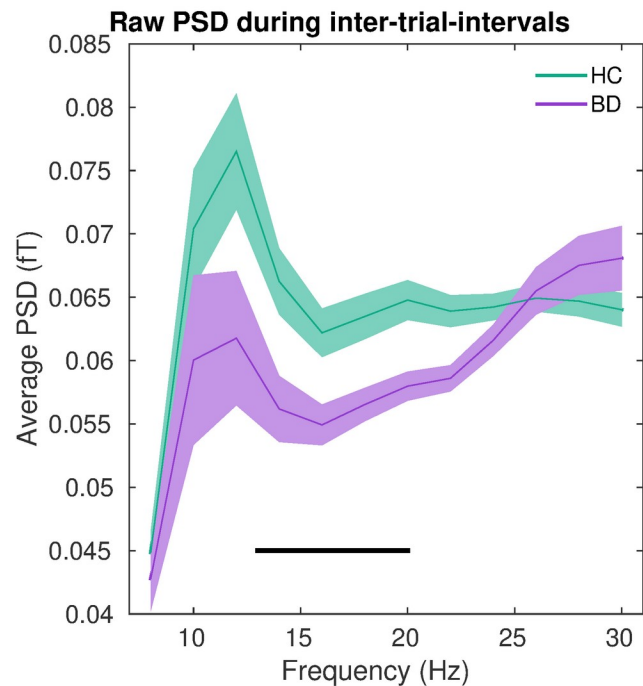


Figure S15. Power spectral density (PSD) during inter-trial-intervals. Grand-average of the raw PSD (in fT) during inter-trial-intervals in healthy control (HC, green) and bipolar disorder (BD, purple) participants. A significant between-group difference was obtained within 13–20 Hz ($P_{FDR} = 0.006$). This effect emerged in most ROIs where the pwPE effect was expressed (illustrated in **Figure 3** and **Figure S12**). The cluster-average in each group is represented by the continuous lines, while shaded areas denote SEM.

Frequency-domain connectivity patterns during pwPE processing

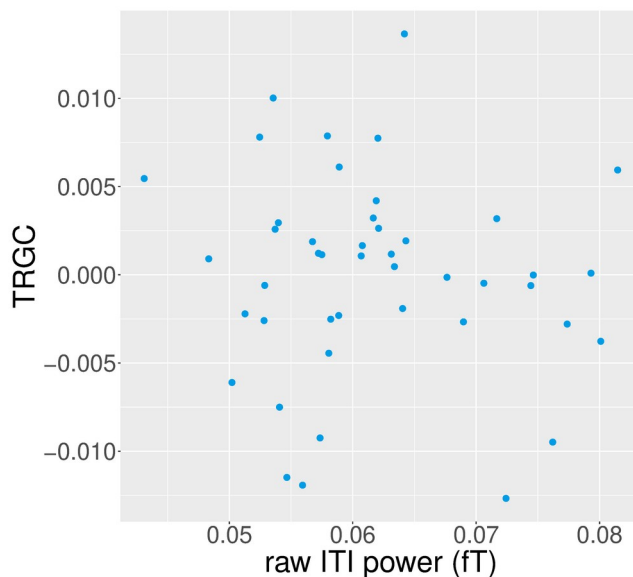


Figure S16. Non-parametric rank correlation between average power spectral density (PSD) during inter-trial-intervals and functional connectivity metrics. Our analyses in **Figure S15** had shown a reduction in the raw power spectral density (PSD, in fT) during inter-trial-intervals (ITI) within 13-20 Hz in BD compared to HC, denoting reduced signal-to-noise ratio (SNR) in the beta band in that time window in the bipolar group. Because SNR can influence the estimation of time-reversed Granger causality (TRGC) metrics [83], a measure of directional functional connectivity, we examined the association between beta-band ITI PSD and beta-band TRGC metrics. A BF-based Spearman rank correlation provided evidence for a lack of association between both variables ($N = 49$; $\rho = -0.04$ [-0.30, 0.24], $P = 0.644$; $BF_{10} = 0.355$, supporting H_0 based on anecdotal evidence). Thus, the main result of a significant beta-band TRGC increase in the bipolar group relative to HC (**Figure 4**) cannot be explained by an association between raw ITI PSD and functional connectivity metrics.

Supplementary References

105. Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clin Psychol Rev* 1988; 8(1): 77-100.
106. Altman EG, Hedeker D, Peterson JL, Davis JM. The Altman Self-Rating Mania Scale. *Biol Psychiatry* 1997; 42(10): 948-955.
107. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand* 1983; 67(6): 361-370.
108. Spielberger CD, Gorsuch RL, Lushene PR, Vagg PR, Jacobs GA. Manual for the state-trait anxiety inventory. Consult Psychol Press 1983.
109. Reitan RM. The relation of the Trail Making Test to organic brain damage. *J Consult Psychol* 1955; 19(5): 393-394.

110. Arnett JA, Labovitz SS. Effect of physical layout in performance of the trail making test. *Psychol Assess* 1995; 7(2): 220-221.
111. Berg EA. A Simple Objective Technique for Measuring Flexibility in Thinking. *J Gen Psychol* 1948; 39(1): 15-22.
112. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; 12(3): 189-198.
113. Yu AJ, Dayan P. Uncertainty, neuromodulation, and attention. *Neuron* 2005; 46(4): 681-692.
114. Weber LA, Waade PT, Legrand N, Møller AH, Stephan KE, Mathys C. The generalized Hierarchical Gaussian Filter. ArXiv:2305.10937. Preprint at <https://arxiv.org/abs/2305.10937> (2023).
115. Weber LA, Diaconescu AO, Mathys C, et al. Ketamine Affects Prediction Errors about Statistical Regularities: A Computational Single-Trial Analysis of the Mismatch Negativity. *J Neurosci* 2020; 40(29): 5658-5668.
116. Mazzoni P, Hristova A, Krakauer JW. Why Don't We Move Faster? Parkinson's Disease, Movement Vigor, and Implicit Motivation. *J Neurosci* 2007; 27(27): 7105-7116.
117. Yoshida J, Oñate M, Khatami L, Vera J, Nadim F, Khodakhah K. Cerebellar Contributions to the Basal Ganglia Influence Motor Coordination, Reward Processing, and Movement Vigor. *J Neurosci* 2022; 42(45): 8406-8415.
118. Jang HJ, Ward RM, Golden CEM, Constantinople CM. Acetylcholine demixes heterogeneous dopamine signals for learning and moving. *BioRxiv*. 2024:2024-05. Preprint at <https://www.biorxiv.org/content/10.1101/2024.05.03.592444v1> (2024).
119. Merikangas KR, Swendsen J, Hickie IB, Lihong C, Haochang S, Merikangas AK et al. Real-time Mobile Monitoring of the Dynamic Associations Among Motor Activity, Energy, Mood, and Sleep in Adults With Bipolar Disorder. *JAMA psychiatry* 2019; 76(2): 190-198.
120. Raudenbush, S. W. & Bryk, A. S. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Vol 1. 2nd ed. (Sage Publications, 2002).
121. Bürkner PC. brms: An R Package for Bayesian Multilevel Models Using Stan. *J Stat Software* 2017; 80: 1-28.
122. Bürkner PC. Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal* 2018; 10(1): 395-411.
123. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. *Stat Sci* 1992; 7(4): 457-472.
124. Vehtari A, Gelman A, Simpson D, Carpenter B, Burkner PC. Rank-normalization, folding, and localization: An improved R-hat for assessing convergence of MCMC. *Bayesian Anal* 2021; 16(2): 667-718.

125. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* 2017; 27(5): 1413-1432.
126. Taulu S, Hari R. Removal of magnetoencephalographic artifacts with temporal signal-space separation: Demonstration with single-trial auditory-evoked responses. *Hum Brain Mapp* 2009; 30(5): 1524-1534.
127. Groman SM, Keistler C, Keip AJ, Hammarlund E, DiLeone RJ, Pittenger C et al. Orbitofrontal Circuits Control Multiple Reinforcement-Learning Processes. *Neuron* 2019; 103(4): 734-746.
128. Rolls ET, Cheng W, Feng J. The orbitofrontal cortex: reward, emotion and depression. *Brain communications*. 2020; 2(2): fcaa196.
129. Morey RA, Haswell CC, Hooper SR, Bellis MD De. Amygdala, Hippocampus, and Ventral Medial Prefrontal Cortex Volumes Differ in Maltreated Youth with and without Chronic Posttraumatic Stress Disorder. *Neuropsychopharmacology* 2016; 41(3): 791-801.
130. Yuan Y, Cassano P, Pias M, Fang Q. Transcranial photobiomodulation with near-infrared light from childhood to elderliness: simulation of dosimetry. *Neurophotonics* 2020; 7(1):015009.
131. Rushworth MFS, Noonan MP, Boorman ED, Walton ME, Behrens TE. Frontal Cortex and Reward-Guided Learning and Decision-Making. *Neuron* 2011; 70(6): 1054-1069.
132. Granger CWJ. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* 1969; 37(3): 424-438.
133. Geweke J. Measurement of Linear Dependence and Feedback between Multiple Time Series. *J Am Stat Assoc* 1982; 77(378): 304-313.
134. Winkler I, Panknin D, Bartz D, Muller KR, Haufe S. Validity of Time Reversal for Testing Granger Causality. *IEEE Trans Signal Process* 2016; 64(11): 2746-2760.
135. Anzolin A, Presti P, Steen F Van De, Astolfi L, Haufe S, Marinazzo D. Quantifying the Effect of Demixing Approaches on Directed Connectivity Estimated Between Reconstructed EEG Sources. *Brain Topogr* 2019; 32(4): 655-674.
136. Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 2006; 93(3): 491-507.
137. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. Bayesian model selection for group studies. *Neuroimage* 2009; 46(4): 1004-1017.
138. Soch J, Allefeld C. MACS – a new SPM toolbox for model assessment, comparison and selection. *J Neurosci Methods* 2018; 306: 19-31.
139. Siafis S, Wu H, Wang D, Burschinski A, Nobura N, Takeuchi H et al. Antipsychotic dose, dopamine D2 receptor occupancy and extrapyramidal side-effects: a systematic review and dose-response meta-analysis. *Mol Psychiatry* 2023; 28(8): 3267-3277.

140. Pizzagalli DA, Goetz E, Ostacher M, Iosifescu D V., Perlis RH. Euthymic patients with bipolar disorder show decreased reward learning in a probabilistic reward task. *Biol Psychiatry* 2008; 64(2): 162-168.
141. Sivula T, Magnusson M, Vehtari A. Unbiased estimator for the variance of the leave-one-out cross-validation estimator for a Bayesian normal model with fixed variance. *Commun Stat - Theory Methods* 2023; 52(16): 5877-5899.