

# Alterations in the amplitude and burst rate of beta oscillations impair reward-dependent motor learning in anxiety

Sebastian Sporn<sup>1,2</sup>, Thomas Hein<sup>2</sup>, Maria Herrojo Ruiz<sup>2,3\*</sup>

<sup>1</sup>School of Psychology, University of Birmingham, Birmingham, United Kingdom;

<sup>2</sup>Department of Psychology, Goldsmiths University of London, London, United Kingdom;

<sup>3</sup>Center for Cognition and Decision Making, Institute for Cognitive Neuroscience, National Research University Higher School of Economics, Moscow, Russian Federation

**Abstract** Anxiety results in sub-optimal motor learning, but the precise mechanisms through which this effect occurs remain unknown. Using a motor sequence learning paradigm with separate phases for initial exploration and reward-based learning, we show that anxiety states in humans impair learning by attenuating the update of reward estimates. Further, when such estimates are perceived as unstable over time (volatility), anxiety constrains adaptive behavioral changes. Neurally, anxiety during initial exploration increased the amplitude and the rate of long bursts of sensorimotor and prefrontal beta oscillations (13–30 Hz). These changes extended to the subsequent learning phase, where phasic increases in beta power and burst rate following reward feedback were linked to smaller updates in reward estimates, with a higher anxiety-related increase explaining the attenuated belief updating. These data suggest that state anxiety alters the dynamics of beta oscillations during reward processing, thereby impairing proper updating of motor predictions when learning in unstable environments.

**Introduction** Anxiety involves anticipatory changes in physiological and psychological responses to an uncertain future threat (*Grupe and Nitschke, 2013; Bishop, 2007*). Previous studies have established that trait anxiety interferes with prefrontal control of attention in perceptual tasks, whereas state anxiety modulates the amygdala during detection of threat-related stimuli (*Bishop, 2007; Bishop, 2009*). An emerging literature additionally identifies the dorsomedial and dorsolateral prefrontal cortex (dmPPC and dlPFC) and the dorsal anterior cingulate cortex (dACC) as central brain regions modulating sustained anxiety, both in subclinical and clinical populations (*Robinson et al., 2019*).

Computational modeling work has started to examine the mechanisms through which anxiety might impair learning, revealing that individuals with high trait anxiety do not correctly estimate the likelihood of outcomes during aversive or reward learning in uncertain environments (*Browning et al., 2015; Huang et al., 2017, Pulcu and Browning, 2019*). In the area of motor control, research has shown that stress and anxiety have detrimental effects on performance (*Baumeister, 1984; Beilock and Carr, 2001*). These results have been interpreted as anxiety interfering with information-processing resources, and as a shift towards an inward focus of attention and an increase in conscious processing of movement (*Eysenck and Calvo, 1992; Pijpers et al., 2005*). The effects of anxiety on motor learning are, however, often inconsistent, and a mechanistic understanding of these effects is still lacking. Delineating mechanisms through which anxiety influences motor learning is important to ameliorate the impact of anxiety in different settings, including in motor rehabilitation programs.

Motor variability could be one component of motor learning that is affected by anxiety; it is defined as the variation of performance across repetitions (*van Beers et al., 2004*), and is affected

\*For correspondence:

M.Herrojo-Ruiz@gold.ac.uk

**Competing interests:** The authors declare that no competing interests exist.

**Funding:** See page 58

**Received:** 29 July 2019

**Accepted:** 08 April 2020

**Reviewing editor:** Nicole C Swann, University of Oregon, United States

© Copyright Sporn et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

---

**eLife digest** Feeling anxious can hinder how well someone performs a task, a phenomenon that is sometimes called “choking under pressure”. Anxiety may also impair a person’s ability to learn a new manual task, like juggling or playing the piano; however, it remains unclear exactly how this happens.

People learn manual tasks more quickly if they can practice first, and the more someone varies their movements during these trial runs, the faster they learn afterwards. Yet, anxiety can affect movement; for example, anxious people often make repetitive motions like hand-wringing or fidgeting. There is also evidence that very anxious people may learn less from the outcomes of their actions.

To understand how anxiety may affect the learning of manual tasks, Sporn et al designed experiments where people learned to play a short sequence of notes on a piano. The main experiment involved 60 participants and was split over two phases. In the first ‘exploration’ phase, participants had to play the piano sequence using any timing they liked and were encouraged to explore different rhythms. In the second ‘learning’ phase, participants were rewarded with a higher score the closer they got to playing the notes with a certain rhythm, without being told that this was their specific goal.

To see how anxiety affected performance, the participants were split into three groups. One group were told in the initial exploration phase that they would give a public talk after they completed the piano task, which reliably made them more anxious. A second group were told about the anxiety-inducing public speaking only during the learning phase; while a third group – the controls – were not aware of any public speaking task.

People in the second group could learn the rhythm as well as the controls. Participants who were made anxious during the exploration phase, however, scored fewer points and were less likely to learn the piano sequence in the second phase. They also varied their movements less in the first phase.

As a follow-up, Sporn et al. repeated the experiment with 26 people but without the initial exploration phase. This time the anxious participants were less able to learn the piano sequence and scored fewer points. This suggests that the initial exploration in the previous experiment had enabled later anxious participants to succeed in the learning phase despite being anxious.

Finally, Sporn et al. also used a technique called electroencephalography (or EEG for short) to record brain activity and observed differences in participants with and without anxiety, particularly when they received their scores. The EEG signals showed that anxiety altered rhythmic patterns of brain activity called “sensorimotor beta oscillations”, which are known to be involved in both movement and learning.

---

by various factors including sensory and neuromuscular noise (*He et al., 2016*). As a form of action exploration, movement variability is increasingly recognized to benefit motor learning (*Todorov and Jordan, 2002; Wu et al., 2014; Pekny et al., 2015*), particularly during reward-based learning, with discrepant effects in motor adaptation paradigms (*He et al., 2016; Singh et al., 2016*). These findings are consistent with the vast amount of research on reinforcement learning that demonstrates increased learning following initial exploration (*Sutton and Barto, 1998; Olveczky et al., 2005*).

Yet contextual factors can reduce variability. For instance, an induced anxiety state leads to ritualistic behavior, characterized by movement redundancy, repetition, and rigidity (*Lang et al., 2015*). This finding resembles the reduction in behavioral variability and exploration that manifests across animal species during phasic fear in reaction to certain imminent threats (*Morgan and Tromborg, 2007*). On the basis of these results, we set out to test the hypothesis that state anxiety modulates motor learning through a reduction in motor variability.

A second component that could be influenced by anxiety is the flexibility to adapt to changes in the task structure during learning. Individuals who are affected by anxiety disorders exhibit an intolerance of uncertainty, which contributes to excessive worry and emotional dysregulation (*Ouellet et al., 2019*). Turning to non-clinical populations, computational studies have established that highly anxious individuals exhibit difficulties in estimating environmental uncertainty both in

aversive and reward-based tasks (Browning *et al.*, 2015; Huang *et al.*, 2017; Pulcu and Browning, 2019). Failure to adapt to volatile or unstable environments thus impairs learning of action-outcome contingencies in these settings. Accordingly, in the context of motor learning, and more specifically, in reward-based motor learning, we proposed that an increase in anxiety would affect individuals' estimation of uncertainty about the stability of the task structure, such as the rewarded movement.

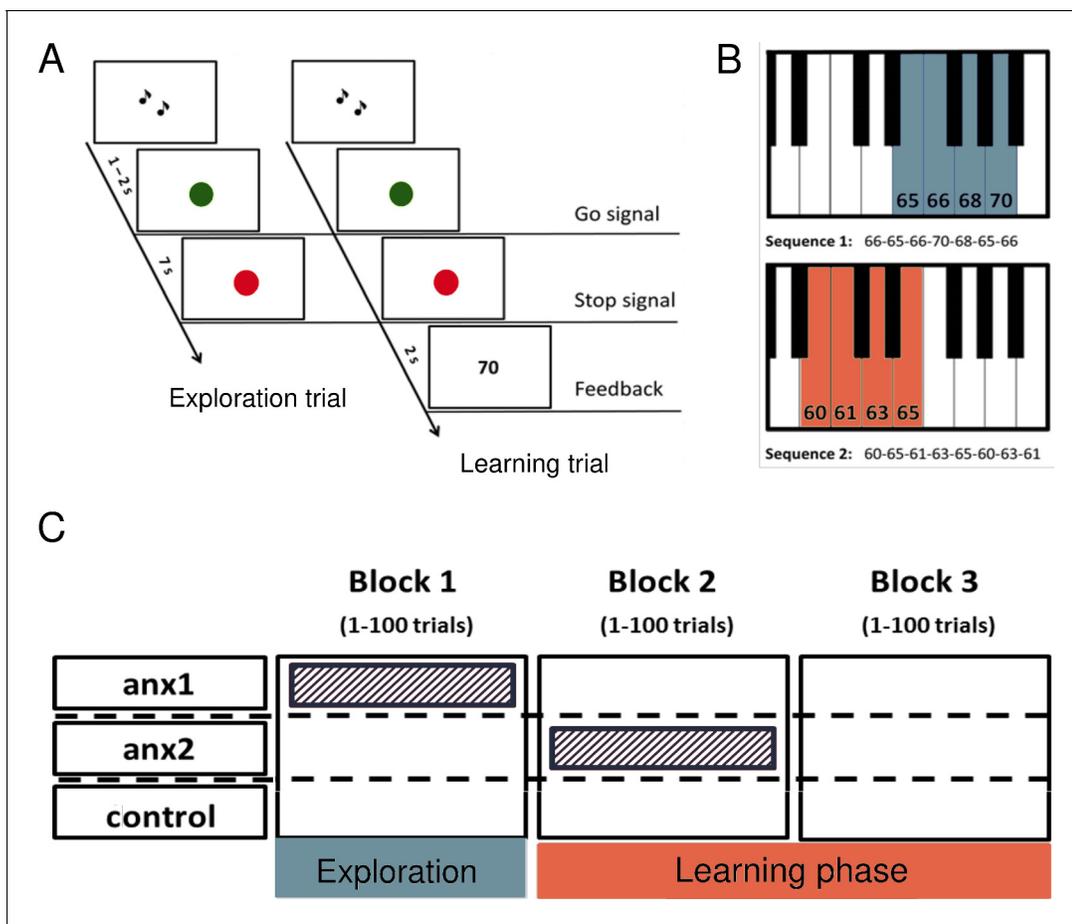
On the neural level, we posited that changes in motor variability are driven by activity in premotor and motor areas. Support for our hypothesis comes from animal studies demonstrating that variability in the primate premotor cortex tracks behavioral variability during motor planning (Churchland *et al.*, 2006). Further evidence supports the hypothesis that changes in variability in single-neuron activity in motor cortex drive motor exploration during initial learning, and reduce it following intensive training (Mandelblat-Cerf *et al.*, 2009; Santos *et al.*, 2015). In addition, the basal ganglia are crucial for modulating variability during learning and production, as shown in songbirds and, indirectly, in patients with Parkinson's disease (Kao *et al.*, 2005; Olveczky *et al.*, 2005; Pekny *et al.*, 2015).

In the present study, we analyzed sensorimotor beta oscillations (13–30 Hz) as a candidate brain rhythm associated with the modulation of motor exploration and variability. Beta oscillations are modulated with different aspects of performance and motor learning (Herrojo Ruiz *et al.*, 2014; Bartolo and Merchant, 2015; Tan *et al.*, 2014), as well as in reward-based learning (HajiHosseini *et al.*, 2012). Increases in sensorimotor beta power following movement have been proposed to signal greater reliance on prior information about the optimal movement (Tan *et al.*, 2016), which would reduce the impact of new evidence on the update of motor commands. We therefore tested the additional hypothesis that changes in sensorimotor beta oscillations mediate the effect of anxiety on belief updates and the estimation of uncertainty driving reward-based motor learning. Crucially, in addition to assessing sensorimotor brain regions, we were interested in prefrontal areas because of prior work in clinical and subclinical anxiety linking the prefrontal cortex (dmPFC and dlPFC) and the dACC to the maintenance of anxiety states, including worry and threat appraisal (Grube and Nitschke, 2013; Robinson *et al.*, 2019). Thus, beta oscillations across sensorimotor and prefrontal electrode regions were evaluated.

Traditionally, the primary focus of research on oscillations was on power changes, although there is a renewed interest in assessing dynamic properties of oscillatory activity, such as the presence of brief bursts (Poil *et al.*, 2008). Brief oscillation bursts are considered to be a central feature of physiological beta waves in motor-premotor cortex and the basal ganglia (Feingold *et al.*, 2015; Tinkhauser *et al.*, 2017; Little *et al.*, 2018). Accordingly, we assessed both the power and burst distribution of beta oscillations to capture dynamic changes in neural activity that were induced by anxiety and their link to behavioral effects. To test our hypotheses, we recorded electroencephalography (EEG) in three groups of participants while they completed a reward-based motor sequence learning paradigm, with separate phases for motor exploration (without reinforcement) and reward-based learning (using reinforcement). We manipulated anxiety by informing participants about an upcoming public speaking task (Lang *et al.*, 2015). Using a between-subject design, the anxiety manipulation targeted either the motor exploration or the reward-based learning phase. Analysis of the EEG signals aimed to assess anxiety-related changes in the power and burst distribution in sensorimotor and prefrontal beta oscillations in relation to changes in behavioral variability and reward-based learning.

**Results** Sixty participants completed our reward-based motor sequence learning task, consisting of three blocks of 100 trials each over two phases (Figure 1): an initial motor exploration (block1, termed exploration hereafter) and a reward-based learning phase (block2 and block3: termed learning hereafter). The rationale for including a motor exploration phase in which participants did not receive trial-based feedback or reinforcement was based on findings indicating that initial motor variability (in the absence of reinforcement) can influence the rate at which participants learn in a subsequent motor task (Wu *et al.*, 2014). If state anxiety reduces the expression of motor variability during the exploration phase, subsequent motor learning would be affected.

Prior to the experimental task, we recorded 3 min of EEG at rest with eyes open in each participant. Next, on a digital piano, participants played two different sequences of seven and eight notes during the exploration and learning phases, respectively (Figure 1B). The sequence patterns were designed so that the key presses would span a range of four neighboring keys on the piano. Participants were explicitly taught the tone sequences prior to the start of the experiment, yet precise



**Figure 1.** A novel paradigm for testing reward-based motor sequence learning. (A) Schematic of the task. Participants performed sequence1 during 100 initial exploration trials, followed by 200 trials over two blocks of reward-based learning performing sequence2. During the learning blocks, participants received a performance-related score between 0–100 that would lead to monetary reward. (B) The pitch content of the sequences used in the exploration (sequence1) and reward-based learning blocks (sequence2), respectively. (C) Schematic of the anxiety manipulation. The shaded area denotes the phase in which anxiety was induced in each group, using the threat of an upcoming public speaking task, which took place immediately after that block was completed.

instructions about the timing or loudness (keystroke velocity,  $K_{vel}$ ) were not provided. The rationale for selecting two different sequences for the exploration and learning phases was to avoid carry-over effects of learning or a preferred performance pattern from the exploration period into the reward-based learning phase (following [Wu et al., 2014](#)).

During the initial exploration phase, participants were informed that they could freely change the pattern of temporal intervals between key presses (inter-keystroke intervals, IKIs) and/or the loudness of the performance in every trial, and that no reward or feedback would be provided. During learning, performance-based feedback in the form of a 0–100 score was provided at the end of each trial. Participants were informed that the overall average score would be translated into monetary reward. They were directly instructed to explore the temporal or loudness dimension (or both) and to use feedback scores to discover the unknown performance objective (which, unbeknownst to them, was related to the pattern of IKIs). The task-related dimension was therefore timing, whereas keystroke velocity was the non-task related dimension.

The performance measure that was rewarded during learning was the vector norm of the pattern of temporal differences between adjacent IKIs (see ‘Materials and experimental design’). Different combinations of IKIs could lead to the same rewarded norm of IKI-difference values, and therefore to the same score. Participants were unaware of the existence of these multiple solutions. The multiplicity in the mapping between performance and score could lead participants to perceive an

increased level of volatility in the environment (changes in the rewarded performance over time). This motivated us to assess their estimation of volatility during reward-based learning and its modulation by anxiety. In addition, we investigated whether higher initial variability would lead to higher scores during subsequent reward-based learning, independently of changes in variability during this latter phase. If initial exploration improves learning of the mapping between the actions and their sensory consequences (even without external feedback), then participants could learn better from performance-related feedback during the learning phase regardless of their use of variability in this phase. Alternatively, it could be that participants who also use more variability during learning discover the hidden goal by chance.

Participants were pseudo-randomly allocated to either a control group or to one of two experimental groups (**Figure 1C**): anxiety during exploration (anx1); and anxiety during the first block of learning (anx2). We measured changes in heart-rate variability (HRV) and heart-rate (HR) four times throughout the experimental session: resting state (3 min, prior to performance blocks); block1; block2; and block3. In addition, the state subscale from the State-Trait Anxiety Inventory (STAI, state scale X1, 20 items; *Spielberger, 1970*) was assessed four times: prior to the resting state recording and also immediately before the beginning of each block, and thus after the induction of anxiety in the experimental groups. The HRV index and STAI state anxiety subscale were able to dissociate in each experimental group between the phase targeted by the anxiety manipulation and the initial resting phase (within-group effects, see statistical results in **Figure 2**). In addition, significant between-group differences in HRV (not in STAI) further confirmed the specificity of the HRV changes in the targeted blocks (statistical details in **Figure 2**). These results confirmed that the experimental manipulation succeeded in inducing physiological and psychological responses within each experimental group that were consistent with an anxious state during the targeted phase, as reported previously (*Feldman et al., 2004*).

Statistical analysis of behavioral and neural measures focused on the separate comparison between each experimental group and the control group (contrasts: anx1 – controls, anx2 – controls). See 'Materials and methods'.

### Behavioral results Lower initial task-related variability is associated with poorer reward-based learning

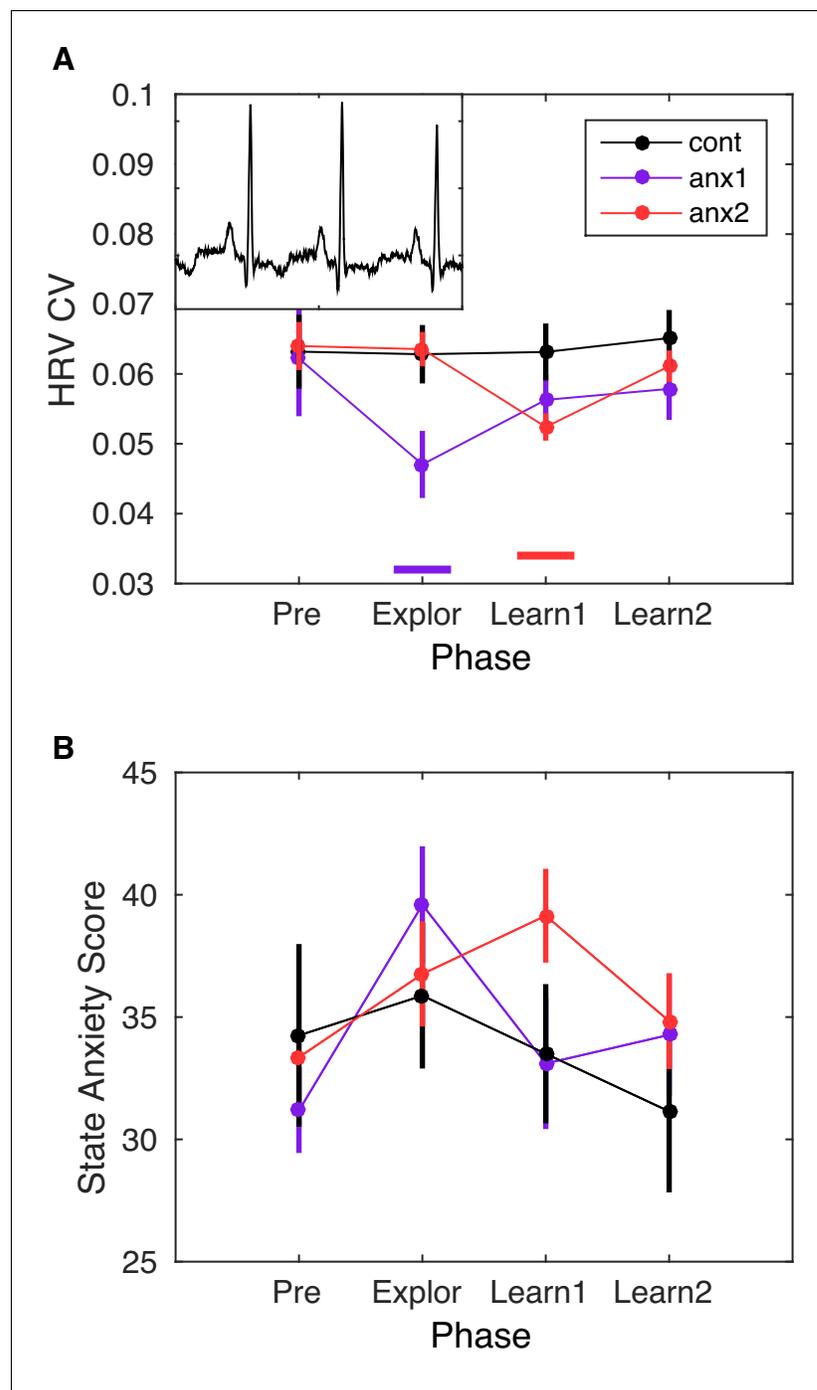
All groups of participants demonstrated significant improvement in the achieved scores during reward-based learning, confirming that they effectively used feedback to approach the hidden target performance (changes in average score from block2 to block3 — anx1:  $p=0.008$ , non-parametric effect size estimator for dependent samples,  $\Delta_{dep} = 0.93$ , confidence interval or CI = [0.86, 0.99]; anx2:  $p=0.004$ ,  $\Delta_{dep} = 0.83$ , CI = [0.61, 0.95]; controls:  $p=0.001$ ,  $\Delta_{dep} = 0.92$ , CI = [0.72, 0.98]).

Assessment of motor variability was performed separately in the task-related temporal dimension and in the non-task-related keystroke velocity dimension. Temporal variability—and similarly for Kvel variability—was estimated using the across-trials coefficient of variation of IKI (termed cvIKI hereafter; **Figure 3A–B**). This index was computed in bins of 25 trials, which therefore provided four values per experimental block. We hypothesized that in the total population, a higher degree of task-related variability during the exploration phase (that is, playing different temporal patterns in each trial), and therefore higher cvIKI, would improve subsequent reward-based learning, as this latter phase rewarded the temporal dimension. A non-parametric rank correlation analysis across the 60 participants revealed that participants who achieved higher scores in the learning phase exhibited a larger across-trials cvIKI during the exploration period (Spearman  $\rho = 0.45$ ,  $P = 0.003$ ; **Figure 3C**).

A similar result was obtained when excluding anx1 participants from the correlation analysis, supporting the hypothesis that in the subsample of 40 participants who did not undergo the anxiety manipulation during exploration there was a significant association between the level of task-related variability and the subsequent score ( $\rho = 0.41$ ,  $P = 0.04$ ). No significant rank correlation was found between the scores and cvKvel ( $P > 0.05$ ).

We also assessed whether the degree of cvIKI during learning was associated with the average score and found an inverted pattern: there was a significant negative non-parametric rank correlation between the cvIKI index and the mean score ( $\rho = -0.44$ ,  $P = 0.002$ ; **Figure 3D**). No significant effect was found for the cvKvel parameter ( $P > 0.05$ ).

Notably, the amount of variability in timing and keystroke velocity used by participants was not correlated (cvIKI and cvKvel during initial exploration:  $\rho = 0.021$ ,  $P = 0.788$ , and during learning:  $\rho = -0.030$ ,  $P = 0.844$ ). This indicates that in our task, participants could vary the temporal and



**Figure 2.** Heart-rate variability (HRV) modulation by the anxiety manipulation. (A) The average HRV measured as the coefficient of variation (CV) of the inter-beat-interval is displayed across the experimental blocks: initial resting state recording (Pre), initial exploration (Explor), first block of learning (Learn1) and, last block of learning (Learn2). Relative to Pre, there was a significant drop in HRV in anx1 participants during initial exploration (within-subject statistics with paired permutation tests,  $P < 0.05$  after controlling the false discovery rate [FDR] at level  $q = 0.05$  due to multiple comparisons, termed  $P_{FDR}$ :  $P_{FDR} < 0.05$ ,  $\Delta_{dep} = 0.81$ ,  $CI = [0.75, 0.87]$ ). In anx2 participants, the drop in HRV was found during the first learning block, which was targeted by the anxiety manipulation ( $P_{FDR} < 0.05$ ,  $\Delta_{dep} = 0.78$ ,  $CI = [0.71, 0.85]$ ). Between-group comparisons revealed that anx1, relative to the control group, exhibited a significantly lower HRV during the exploration phase ( $P_{FDR} < 0.05$ ,  $\Delta = 0.75$ ,  $CI = [0.65, 0.85]$ , purple bar at the bottom). The anx2 group manifested a significant drop in HRV relative to controls during the first learning block ( $P_{FDR} < 0.05$ ,  $\Delta = 0.71$ ,  $CI = [0.62, 0.80]$ , red bar at the bottom). These results demonstrate a group-

Figure 2 continued on next page

Figure 2 continued

specific modulation of anxiety relative to controls during the targeted blocks. The mean HR did not change within or between groups ( $P>0.05$ ). (B) STAI state anxiety score in each group across the different experimental phases. Participants completed the STAI state anxiety subscale first at the start of the experiment before the resting state recording (Pre) and subsequently again immediately before each experimental block (and right after the anxiety induction: Explor, Learn1, Learn2). There was a within-group significant increase in the score for each experimental group during the phase targeted by the anxiety manipulation (anx1: Explor relative to Pre, average score 40 [2] and 31 [2], respectively;  $P_{FDR}<0.05$ ,  $\Delta_{dep} = 0.74$ ,  $CI = [0.68, 0.80]$ ; anx2: Learn1 relative to Pre, average score 39 [2] and 34 [2], respectively;  $P_{FDR}<0.05$ ,  $\Delta_{dep} = 0.78$ ,  $CI = [0.68, 0.86]$ ). Between-group differences were non-significant.

velocity dimensions separately. On the other side, however, the generally lower cvKvel values in all blocks and groups further indicate that participants may not have been able to substantially vary this dimension. Finally, the degree of cvKI during the learning and exploration phases were not correlated ( $\rho = 0.029$ ,  $P = 0.848$ ). These findings suggest that achieving higher scores during reward-based learning in our paradigm cannot be accounted for by a general tendency towards more exploration throughout all experimental blocks. In fact, larger sustained variability during learning was detrimental to maintaining the performance close to the inferred target (**Figure 3D**).

### Anxiety during initial exploration reduces task-related variability and impairs subsequent reward-based learning

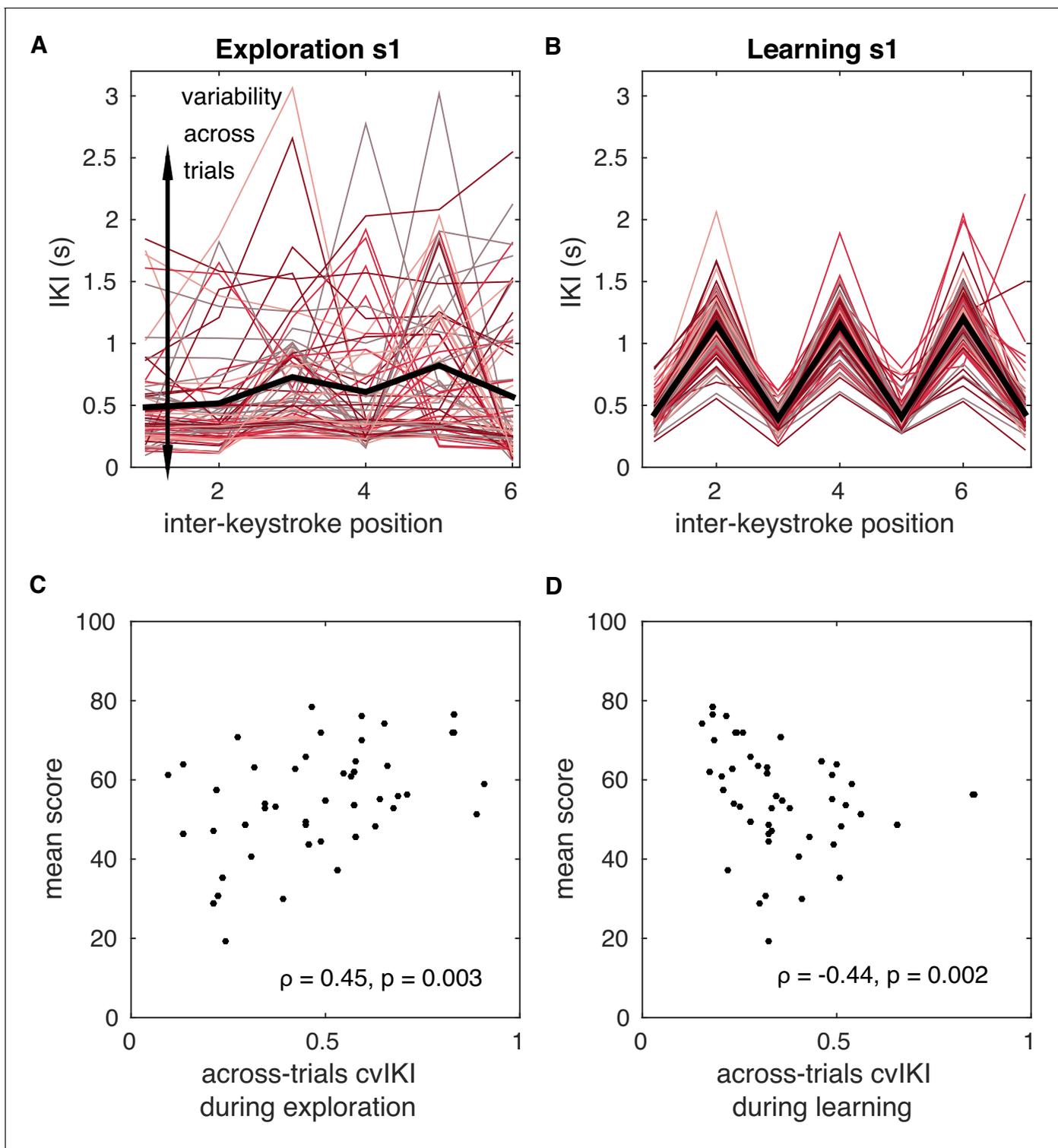
Next, we assessed pair-wise differences in the behavioral measures between the control group and each experimental group (anx1 and anx2) separately. Participants who were affected by state anxiety during initial exploration (anx1) achieved significantly lower scores in the subsequent reward-based learning phase relative to control participants (**Figure 4A**:  $P<0.05$  after controlling the false discovery rate [FDR] at level  $q = 0.05$  due to multiple comparisons, termed  $P_{FDR}$  thereafter;  $\Delta = 0.78$ ,  $CI = [0.54, 0.92]$ ). By contrast, in the anx2 group scores did not statistically differ from the scores in the control group ( $P_{FDR}>0.05$ ). A planned comparison between both experimental groups demonstrated significantly higher scores in anx2 than in anx1 ( $P_{FDR}<0.05$ ,  $\Delta = 0.67$ ,  $CI = [0.51, 0.80]$ ).

During the initial exploration block, anx1 used a lower degree of cvKI than the control group (**Figure 4B**;  $P_{FDR}<0.05$ ;  $\Delta = 0.67$ ,  $CI = [0.52, 0.85]$ ). There was no between-groups (anx1, controls) difference in cvKvel (**Figure 4C**;  $P_{FDR}>0.05$ ). Performance in anx2 in this phase did not significantly differ from performance in the control group, either for cvKI or for cvKvel ( $P_{FDR}>0.05$ ).

Subsequently, during the learning blocks, there were no significant between-group differences in cvKI or cvKvel ( $P_{FDR}>0.05$ ). In each group, there was a significant drop in the use of temporal variability from the first to the second learning block, corresponding to a transition from exploration to the exploitation of the rewarded options (significant drop in cvKI from block2 to block3 in control, anx1, and anx2 participants;  $P_{FDR}<0.05$ ; effect size —  $\Delta_{dep} = 0.77$ ,  $CI = [0.53, 0.87]$  in controls;  $\Delta_{dep} = 0.55$ ,  $CI = [0.50, 0.61]$  in anx1;  $\Delta_{dep} = 0.83$ ,  $CI = [0.62, 0.94]$  in anx2). This outcome further indicated that all groups successfully completed the reward-based learning task, although anx1 participants achieved lower scores than the reference control group.

Detailed analyses of the trial-by-trial changes in scores and performance using a Bayesian learning model and their modulation by anxiety are reported below. General performance parameters, such as the average performance tempo or the mean keystroke velocity did not differ between groups, either during initial exploration or learning ( $P>0.05$ ). Participants completed sequence1 in 3.0 (0.1) seconds on average, between 0.68 (0.05) and 3.68 (0.10) s after the GO signal (non-significant differences between groups,  $P>0.05$ ). During learning, they played sequence2 with an average duration of 4.7 (0.1) s, between 0.72 (0.03) and 5.35 (0.10) s (non-significant differences between groups,  $P>0.05$ ). The mean learned solution was not significantly different between groups, either during the first or second learning block ( $P>0.05$ ; **Figure 4—figure supplement 1**; but see trial-by-trial changes below).

These outcomes demonstrate that in our paradigm, state anxiety reduced task-related motor variability when induced during the exploration phase and this effect was associated with lower scores during subsequent reward-based learning. State anxiety, however, did not modulate task-related motor variability or the scores achieved when induced during reward-based learning. Finally, the different experimental manipulations did not affect the mean learned solution in each group. State anxiety during reward-based learning reduces learning rates if there is no prior



**Figure 3.** Temporal variability during initial exploration and during reward-based learning. (A, B) Illustration of timing performance during initial exploration (A) and learning (B) blocks for one representative participant, s1. The x-axis represents the position of the inter-keystroke interval (sequence1: seven notes, corresponding to six inter-keystroke temporal intervals; sequence2: eight notes, corresponding to seven inter-keystroke intervals). The y-axis shows the inter-keystroke interval (IKI) in ms. Black lines represent the mean IKI pattern. Red-colored traces represent the individual timing performance in each of the 100 (A) and 200 (B) trials during exploration and learning blocks, respectively. Task-related temporal variability was measured using the across-trials coefficient of variation of IKI, cvIKI. This measure was computed in successive bins of 25 trials, which allowed us to track changes in cvIKI across time. (C) Non-parametric rank correlation in the total population ( $N = 60$ ) between the across-trials cvIKI during

*Figure 3 continued on next page*

Figure 3 continued

exploration (averaged across the four 25-trial bins) and the average score achieved subsequently during learning (Spearman  $\rho = 0.45$ ,  $P = 0.003$ ). (D) Same as panel (C) but using the individual value of the across-trials cvKI from the learning phase (cvKI was averaged here across all eight 25-trial bins; Spearman  $\rho = -0.44$ ,  $P = 0.002$ ).

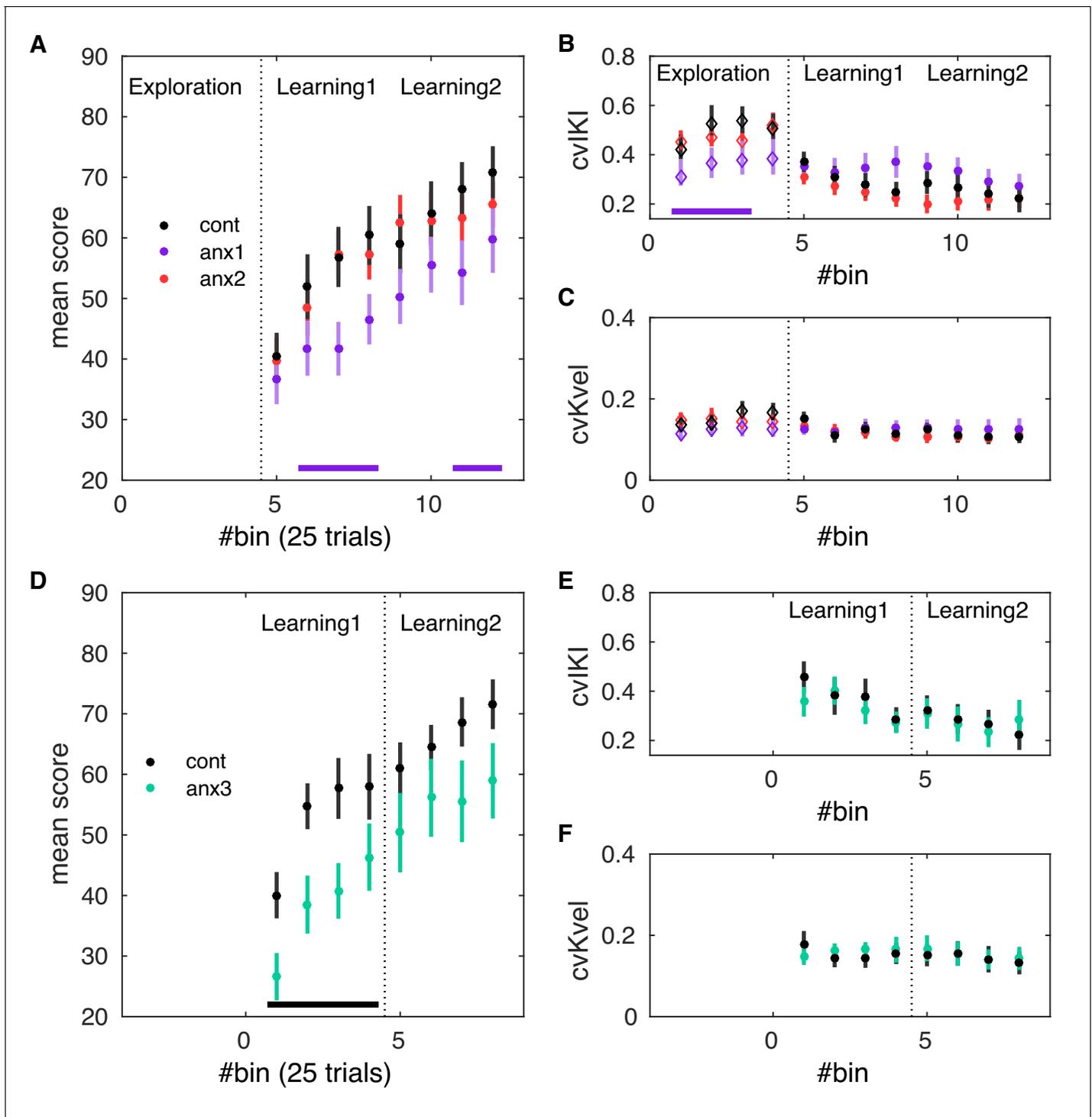
**exploration phase** Because anx2 participants performed at a level that was not significantly different from that found in control participants during learning, we asked whether the initial unconstrained motor exploration during the initial phase might have counteracted the effect of anxiety during learning blocks. Alternatively, it could be that the anxiety manipulation was not salient enough in the context of reward-based learning. To assess these alternative scenarios, we performed a control behavioral experiment with new experimental (anx3) and control groups ( $N = 13$  each, see sample size estimation in 'Materials and methods'). Participants in each group performed the two learning blocks 2 and 3 (**Figure 1C**), but without completing a preceding exploration block. In anx3, state anxiety was induced exclusively during the first learning block, as in the original experiment. We found that the HRV index was significantly reduced in anx3 relative to controls during the manipulation phase ( $P_{FDR} < 0.05$ ,  $\Delta = 0.72$ ,  $CI = [0.62, 0.83]$ ), but not during the final learning phase (block3,  $P_{FDR} > 0.05$ ). STAI state subscale scores rose during the anxiety manipulation in anx3 (but not in controls) relative to the initial scores (within-group effect,  $P_{FDR} < 0.05$ ,  $\Delta = 0.68$ ,  $CI = [0.59, 0.78]$ ).

Overall, the anx3 group achieved a lower average score (and final monetary reward) than control participants ( $P = 0.0256$ ;  $\Delta = 0.64$ ,  $CI = [0.50, 0.71]$ ). In addition, anx3 participants achieved significantly lower scores than control participants during the first learning block ( $P_{FDR} < 0.05$ ,  $\Delta = 0.68$ ,  $CI = [0.54, 0.79]$ , **Figure 4D**), but not during the second learning block ( $P_{FDR} > 0.05$ ). Notably, however, the degree of cvKI or cvKvel did not differ between groups ( $P_{FDR} < 0.05$ , **Figure 4E–F**). The mean performance tempo, loudness and the mean learned solution during learning did not differ significantly between groups, as in the main experiment ( $P > 0.05$ ). Thus, removal of the initial exploration phase led to the impairment of reward-based learning by the anxiety manipulation, and this effect was not associated with a change in the use of task-related variability or in general average performance parameters.

**Bayesian learning modeling reveals the effects of state anxiety on reward-based motor learning**

To assess our hypotheses regarding the mechanisms underlying participants' performance during reward-based learning, we used several versions of a Bayesian learning model, which were based on the two-level hierarchical Gaussian filter for continuous input data (HGF; **Mathys et al., 2011; Mathys et al., 2014**). The HGF was introduced by **Mathys et al. (2011)** to model how an agent infers a hidden state in the environment (a random variable,  $x_1$ , as well as its rate of change over time ( $x_2$ , environmental volatility). This corresponds to a perceptual model, which is further coupled with a response model to generate responses based on those inferred states. In the two-level HGF, beliefs about those hierarchically related hidden states ( $x_1, x_2$ ) are continuous variables evolving as Gaussian random walks coupled through their variance. Their value ( $x_i, i = 1, 2$ ) at trial  $k$  will be normally distributed around their previous value at trial  $k - 1$ . Thus, the posterior distribution of beliefs about these states is fully determined by the sufficient statistics  $\mu_i$  (mean) and  $\sigma_i$  (variance). Beliefs are updated given new sensory input via prediction errors (PEs). In some implementations of the HGF, the series of sensory inputs are replaced by a sequence of outcomes, such as reward value in a binary lottery (**Mathys et al., 2014; Diaconescu et al., 2017**) or electric shock delivery in a one-armed bandit task (**de Berker et al., 2016**). In these cases, similarly to the case of sensory input, an agent can learn the causes of the observed outcomes and thus the likelihood that a particular event will occur. In our study, the trial-by-trial input observed by the participants was the series of feedback scores (hereafter input refers to feedback scores). Crucial to the HGF is the weighting of the PEs by the ratio between the estimation uncertainty of the current level and the lower level, or the inverse ratio when using precision (inverse variance or uncertainty of a distribution). Further details are provided in the 'Materials and methods'.

Different implementations of the HGF have recently been used in combination with neuroimaging data to investigate how the brain processes different types of hierarchically-related prediction errors (PEs) within the framework of predictive coding (**Diaconescu et al., 2017; Weber et al., 2019**). The HGF can be fit to the behavioral data from each individual participant, thus providing dynamic trial-



**Figure 4.** Effects of anxiety on behavioral variability and reward-based learning. The score was computed as a 0–100 normalized measure of proximity between the norm of the pattern of differences in inter-keystroke intervals performed in each trial and the target norm. All of the behavioral measures shown in this figure are averaged within bins of 25 trials. (A) Scores achieved by participants in the anx1 (N = 20), anx2 (N = 20), and control (N = 20) groups across bins 5:12 (trial range 101–300), corresponding to blocks 2 and 3 and the learning phase. Participants in anx1 achieved significantly lower scores than control participants ( $P_{FDR} < 0.05$ , denoted by the bottom purple line). (B) Changes in across-trials cvIKI, revealing a significant drop in task-related exploration during the initial phase in anx1 relative to control participants ( $P_{FDR} < 0.05$ ). Anx2 participants did not differ from control participants. (C) Same as panel (B) but for the across-trials cvKvel. (D–F) Control experiment: effect of anxiety on variability and learning after removal of the initial exploration phase. Panels (D–F) are displayed in the same way as panels (A–C) for experimental (N = 13) and control (N = 13) groups. Significant between-group differences are denoted by the black bar at the bottom ( $P_{FDR} < 0.05$ ,  $\Delta = 0.71$ ,  $CI = [0.64, 0.78]$ ). (F) In anx3 participants (green), there was

Figure 4 continued on next page

Figure 4 continued

a significant drop in the mean scores during the first learning block relative to control participants ( $P_{FDR} < 0.05$ ,  $\Delta = 0.77$ ,  $CI = [0.68, 0.86]$ ). Bars around the mean show  $\pm$  SEM.

The online version of this article includes the following figure supplement(s) for figure 4:

**Figure supplement 1.** Mean learned solution in each group.

wise estimates of belief updates that depend on hierarchical PEs weighted by precision (precision-weighted PE or pwPE). In predictive coding models, precision is viewed as crucial for representing uncertainty and updating the posterior expectations about the hidden states (Sedley et al., 2016). In the HGF, time-varying pwPEs reflect how participants learn stimulus-outcome or response-outcome associations and their changes over time (Mathys et al., 2014; Diaconescu et al., 2017).

Here, we adapted the HGF to model participants' estimation of quantity  $x_1$ , which represented their beliefs about the expected reward (input score, normalized 0–1) for the current trial. Beliefs about  $x_1$  on trial  $k$  were thus determined by the expectation of reward  $\mu_1^k$  (mean of the posterior distribution of  $x_1$ ) and the uncertainty about this estimate (variance,  $\sigma_1^k$ ). The model also estimated participants' beliefs about environmental volatility  $x_2$ , related to changes in the reward tendency and determined by  $(\mu_2^k, \sigma_2^k)$  on trial  $k$ . The belief trajectories about the external states  $x_1$  and  $x_2$  generated by the model were further used to estimate the most likely response corresponding with those beliefs. A schematic illustrating the model structure and the belief trajectories is shown in Figure 5. Volatility in our paradigm emerged from the multiplicity of performance-to-score mappings, as different temporal patterns of the performance with identical norm of IKI-difference values led to the same scores (e.g. IKI: [0.2, 0.5, 0.2, 0.5, 0.2, 0.5] s and [0.6, 0.3, 0.6, 0.3, 0.6, 0.3, 0.6] s).

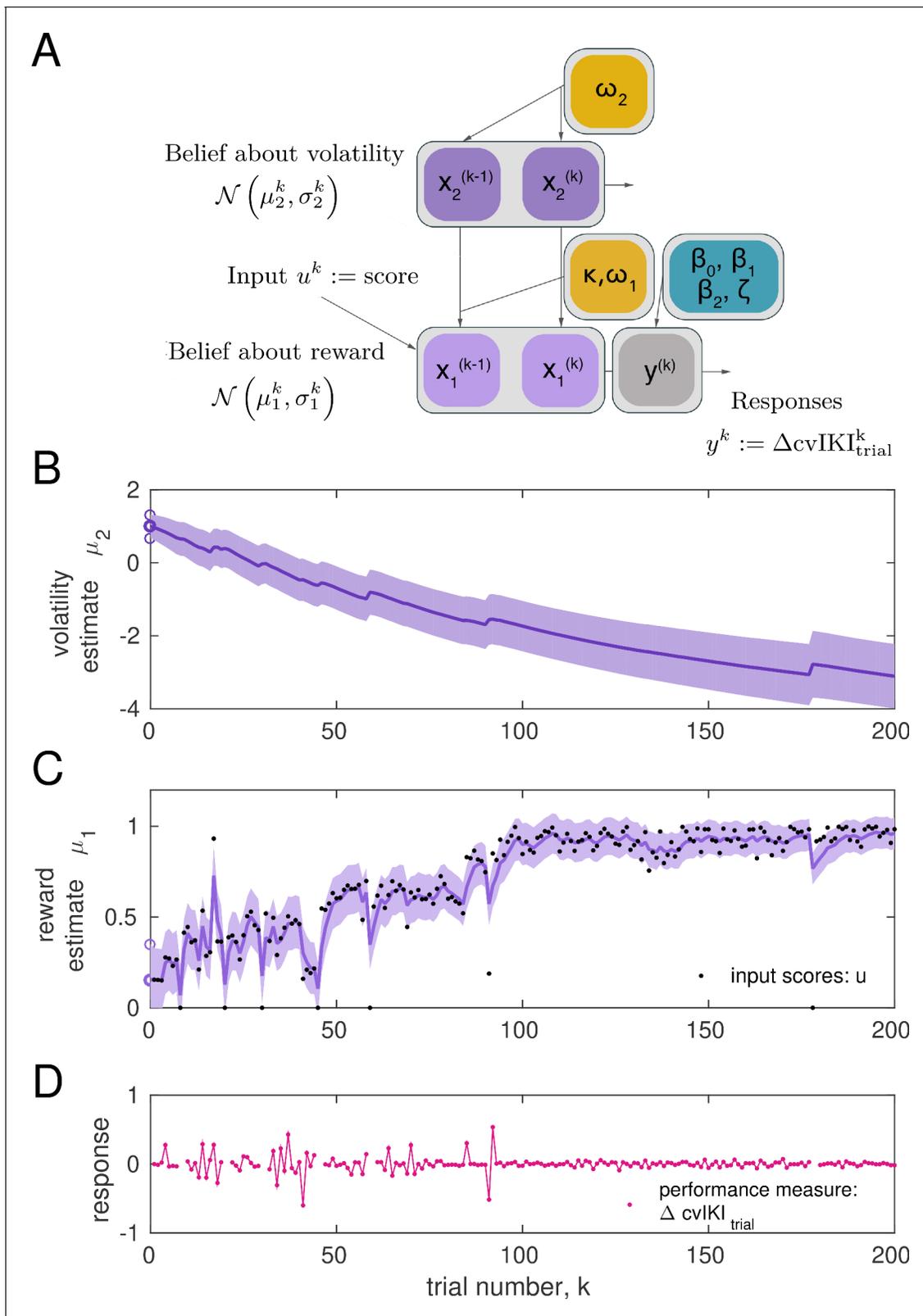
Assessment of the HGF for simulated responses revealed that the expectation of volatility (change in reward tendency) was higher in participants who modulated their performance to a greater extent across trials and thereby observed a broader range of feedback scores (see different examples for simulated performances in Figure 5—figure supplement 1).

We implemented eight versions of the HGF with different response models. The response model defines the mapping from the trajectories of perceptual beliefs onto the observed responses of each participant. We were interested in how HGF quantities on the previous trial explained changes in performance on the subsequent trial. To assess that relationship, we considered two scenarios characterized by the choice of a different performance measure in the response model. The performance measures used were: (1) the trialwise coefficient of variation of consecutive IKI values (cv across sequence positions; termed  $cvIKI_{\text{trial}}$  to dissociate it from the measure of across-trials variability,  $cvIKI$ ); (2) the trialwise performance tempo (mean of IKI within the trial across sequence positions, termed  $mIKI_{\text{trial}}$ ; here we used the logarithm of this measure in milliseconds,  $\log(mIKI_{\text{trial}})$ ), as in Marshall et al. (2016). Accordingly, we constructed two families of models describing the link between a participant's inferred perceptual quantities on the previous trial  $k - 1$  and their changes from trial  $k - 1$  to  $k$  in one of those performance measures:

$$\Delta cvIKI_{\text{trial}}^k = cvIKI_{\text{trial}}^k - cvIKI_{\text{trial}}^{k-1}$$

$$\Delta \log(mIKI_{\text{trial}})^k = \log(mIKI_{\text{trial}}^k) - \log(mIKI_{\text{trial}}^{k-1})$$

Variable  $cvIKI_{\text{trial}}$  was chosen because it is tightly linked to the variable associated with reward: higher differences in IKI values between neighboring positions lead not only to a higher vector norm of IKI patterns but also to a higher coefficient of variation of IKI values in that trial (and indeed  $cvIKI_{\text{trial}}$  was positively correlated with the feedback score across participants, nonparametric Spearman  $\rho = 0.69$ ,  $P < 10^{-5}$ ). Alternatively, we considered the scenario in which participants would speed or slow down their performance without altering the relationship between successive intervals. Therefore, we used a performance measure related to the mean tempo,  $mIKI$ . We did not choose a performance measure associated with keystroke velocity because our results in the previous sections demonstrate that participants did not consistently modulate  $cvKvel$  across trials—either because they realized that this parameter was non-task-related or because they were not able to substantially vary the loudness of the key press. Similarly to Marshall et al. (2016), in each family of models we



**Figure 5.** Two-level Hierarchical Gaussian Filter for continuous inputs. (A) Schematic of the two-level HGF, which models how an agent infers a hidden state in the environment (a random variable),  $x_1$ , as well as its rate of change over time ( $x_2$ , environmental volatility). Beliefs about those two hierarchically related hidden states ( $x_1$ ,  $x_2$ ) at trial  $k$  are updated by the sensory input ( $u^k$ , observed feedback scores in our study) for that trial via prediction errors (PEs). The states  $x_1$  and  $x_2$  are continuous variables evolving as coupled Gaussian random walks, where the step size (variance) of the

*Figure 5 continued on next page*

Figure 5 continued

random walk depends on a set of parameters (shown in yellow boxes). The lowest level is coupled to the level above through the variance of the random walk:  $x_1^k \sim \mathcal{N}(x_1^{k-1}, \exp(\kappa x_2^{k-1} + \omega_1))$ . The posterior distribution of beliefs about these states is fully determined by the sufficient statistics  $\mu_i$  (mean) and  $\sigma_i$  (variance) for levels  $i = 1, 2$ . The equations describing how expectations ( $\mu_i$ ) change from trial  $k - 1$  to  $k$  are **Equation 6** and **Equation 10**. The response model generates the most probable response,  $y^k$ , according to the current beliefs, and is modulated by the response model parameters  $\beta_0, \beta_1, \beta_2, \zeta$ . In the winning model, the response parameter was the change between trial  $k - 1$  and  $k$  in the degree of temporal variability across keystrokes:  $y^k = \Delta \text{cvIKI}_{\text{trial}}^k$ , normalized to range 0–1. (B, C) Example of belief trajectories (mean, variance) associated with the two levels of the HGF for continuous inputs. Panel (C) displays the expectation on the first level,  $\mu_1^k$ , which represents an individual's mean estimate of the true reward values for the trial,  $x_1^k$ . Black dots represent the trial-wise input (feedback scores,  $u^k$ ). Panel (B) shows the trial-by-trial beliefs about volatility  $x_2^k$ , determined by the expectation  $\mu_2^k$  and associated variance. Shaded areas denote the variance or estimation uncertainty on that level. (D) Illustration of the performance measure used as response in the winning model,  $y^k = \Delta \text{cvIKI}_{\text{trial}}^k$ .

The online version of this article includes the following figure supplement(s) for figure 5:

**Figure supplement 1.** Trial-by-trial belief trajectories for simulated performances.

**Figure supplement 2.** Simulated trial-by-trial belief trajectories in an ideal learner.

**Figure supplement 3.**  $\beta$  coefficients of the winning response model.

**Figure supplement 4.** Example in one control participant of the association between pwPEs and performance.

**Figure supplement 5.** Example in one anx1 participant of the association between pwPEs and performance.

**Figure supplement 6.** Grand-average trialwise residuals.

defined four types of response models to explain the performance measure as a linear function of relevant HGF perceptual parameters on the previous trial, such as the expectation of reward ( $\mu_1$ ) or volatility ( $\mu_2$ ) and the pwPEs on these estimates (labeled  $\epsilon_1$  and  $\epsilon_2$ , respectively; see **Equation 14** and **Equation 15**). One example is illustrated here:

$$\Delta \text{cvIKI}_{\text{trial}}^k = \beta_0 + \beta_1 \mu_1^{k-1} + \beta_2 \epsilon_1^{k-1} + \zeta \quad (1)$$

where  $\beta_0$  represents a constant value (intercept) and  $\zeta$  is a Gaussian noise variable. Details on the alternative models are provided in the 'Materials and methods' section.

In each model, the feedback scores and the performance measure at each trial  $k$  were used to update model parameters, and the log model-evidence was used to optimize the model fit (**Diaconescu et al., 2017; Soch and Allefeld, 2018**). More details on the modeling approach can be found in the 'Materials and methods' section and in **Figure 5**.

Between-group comparison focused on four variables, the mean trajectories of perceptual beliefs ( $\mu_1$  and  $\mu_2$ , means of the posterior distributions for  $x_1$  and  $x_2$ ; **Figure 5**), and the uncertainty about those beliefs (variance of the posterior distributions,  $\sigma_1$  and  $\sigma_2$ ; note that the inverse variance is the precision, termed  $\pi_1$  and  $\pi_2$ , corresponding with the confidence placed on those beliefs). As indicated above, volatility estimates are related to the rate of change in reward estimates, and accordingly we predicted a higher expectation of volatility  $\mu_2$  for participants exhibiting more variation in  $\mu_1$  values. In addition, the perceptual model parameters  $\omega_1$  and  $\omega_2$ , which characterize the learning style of each participant (see **Figure 5—figure supplement 2**), and the parameters  $\beta_0, \beta_1, \beta_2, \zeta$ , characterizing the response model, were contrasted between groups.

Random Effects Bayesian Model Selection (BMS) was used to assess at the group level ( $N = 60$ ) the different models of learning (**Stephan et al., 2009**); code freely available from the MACS toolbox, (**Soch and Allefeld, 2018**). First, the models were grouped into two families corresponding with each performance measure ( $\Delta \text{cvIKI}_{\text{trial}}$  and  $\Delta \log(\text{mIKI}_{\text{trial}})$ ). The log-family evidence (LFE) was calculated from the log-model evidence (LME). BMS then determined which family of models provided more evidence. In the winner family, additional BMS determined the final optimal model. BMS provided stronger evidence for the family of models defined for  $\Delta \text{cvIKI}_{\text{trial}}$ , with an exceedance probability of 1, and an expected frequency of 0.9353 (similar values in experimental and control groups). Next, among all four models in that family, the winning model (exceedance probability 1, model frequency 0.8614) explained the performance measure  $\Delta \text{cvIKI}_{\text{trial}}$  as a linear function of the pwPE relating to reward,  $\epsilon_1$ , and volatility,  $\epsilon_2$ , on the previous trial:

$$\Delta \text{cvIKI}_{\text{trial}}^k = \beta_0 + \beta_1 \epsilon_1^{k-1} + \beta_2 \epsilon_2^{k-1} + \zeta \quad (2)$$

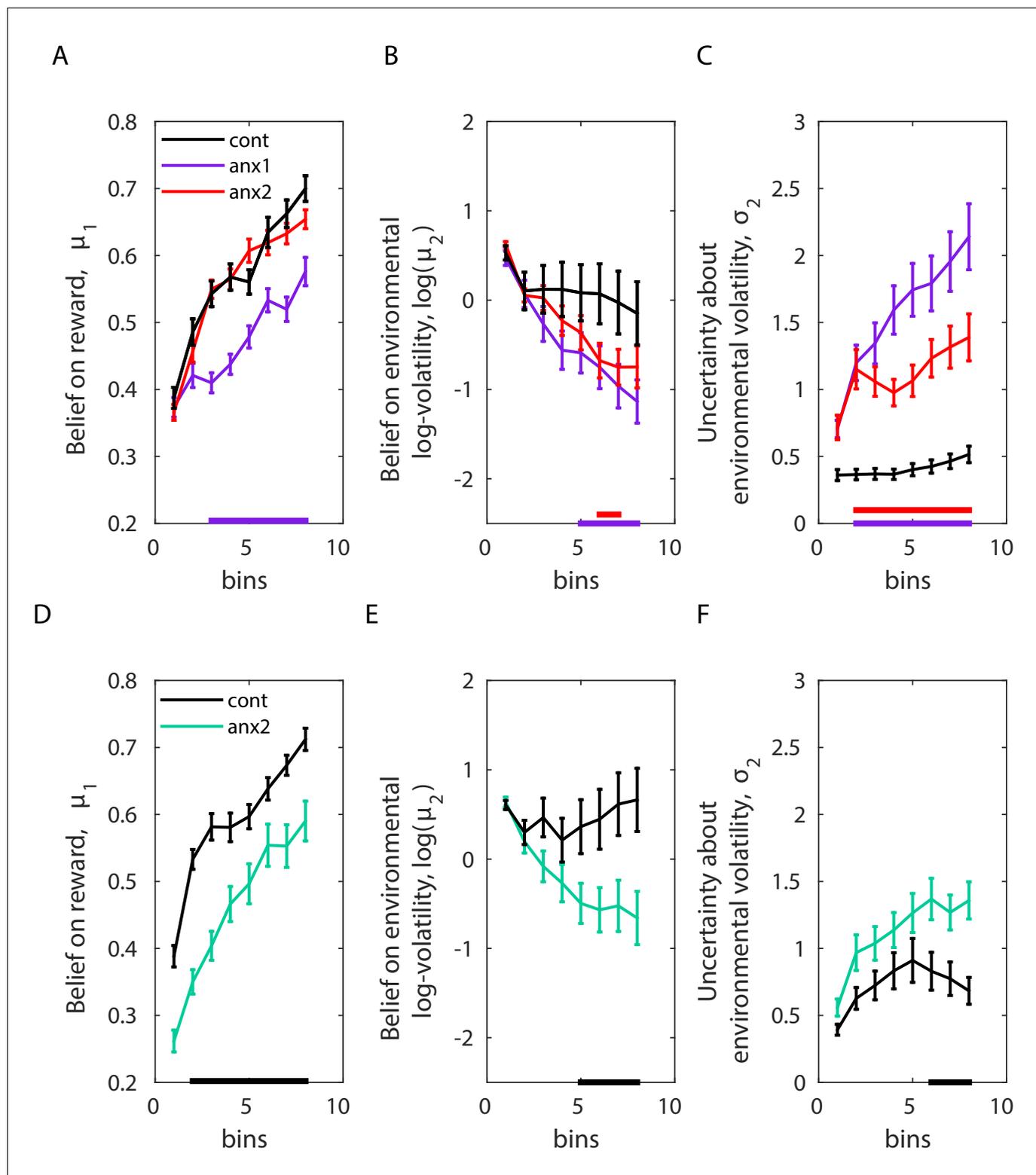
The  $\beta_0$  and  $\beta_1$  coefficients were significantly different than zero in each experimental and control

group ( $P_{FDR} < 0.05$ , controlled for multiple comparisons arising from three group tests; **Figure 5—figure supplement 3**). On average,  $\beta_0$  was positive, and  $\beta_1$  was negative. By contrast,  $\beta_2$  was positive in the control group yet negative in the anx1 and anx2 groups ( $P_{FDR} < 0.05$ ). Because pwPEs directly modulate the update in the expectation of beliefs, these findings imply that smaller pwPEs relating to reward on the previous trial (smaller update in the expectation of reward at  $k - 1$ ) were associated in all groups with increases in  $cvIKI_{\text{trial}}^k$  for the next trial. On the other side, a negative  $\beta_1$  also indicates that larger pwPE for reward on the previous trial decreased changes in the performance variable on the following trial. In addition, exclusively in control participants, there was a positive association between larger pwPE relating to volatility at  $k - 1$  (greater update in the expectation on volatility on the last trial) and a follow-up increment in  $cvIKI_{\text{trial}}^k$ . In anx1 and anx2 participants, however, trials of larger pwPE driving updates on volatility were followed by reduced changes in trial-wise temporal variability. The results imply that a larger increase in the expectation of volatility on the previous trial promoted larger subsequent changes in the relevant performance variable in control participants (**Figure 5—figure supplement 4**), whereas in anx1 and anx2, it led to reductions in task-related behavioral changes (**Figure 5—figure supplement 5**).

The HGF and the winning response model provided a good fit to the behavioral data from each group, as shown in the examination of the residuals (**Figure 5—figure supplement 6**). Further, there were no systematic differences in the model fits across groups (trial-averaged residuals were compared between each experimental and control group with permutation tests;  $P > 0.05$  in both comparisons;  $P = 0.1598$  for anx1 and control groups;  $P = 0.5646$  for anx2 and control groups). The low mean residual values further indicate that the model captured the fluctuations in data well (trial-averaged residuals and SEM: 0.0004[0.00095] in controls;  $-0.001[0.0013]$  in anx1; and 0.0003[0.0003] in anx2).

Using the winning model, we next evaluated between-group differences in the mean trajectories of perceptual beliefs and their uncertainty throughout learning (**Figure 6A–C**). Participants in the anx1 relative to the control group had a lower estimate of the mean tendency for  $x_1$  ( $P_{FDR} < 0.05$ ,  $\Delta = 0.75$ ,  $CI = [0.59, 0.89]$ ). This indicates a lower expectation of reward in the current trial. Note that this outcome could be anticipated from the behavioral results shown in **Figure 4A**. The expectation on log-volatility was significantly smaller in anx1 than in control participants ( $P_{FDR} < 0.05$ ,  $\Delta = 0.71$ ,  $CI = [0.60, 0.81]$ ). This quantity was also partly reduced in the anx2 group relative to the control group ( $P_{FDR} < 0.05$ ,  $\Delta = 0.69$ ,  $CI = [0.53, 0.75]$ ). In addition, the uncertainty about environmental volatility,  $\sigma_2$ , was larger in the anx1 and anx2 participants when compared to control participants (control relative to anx1,  $P_{FDR} < 0.05$ ,  $\Delta = 0.71$ ,  $CI = [0.65, 0.89]$ ; control relative to anx2,  $P_{FDR} < 0.05$ ,  $\Delta = 0.65$ ,  $CI = [0.52, 0.86]$ ). Because larger estimation uncertainty on the current HGF level contributes toward larger steps in the update equations for that level (due to larger precision weights on the PEs, **Equation 5**), this last outcome suggests that anx1 and anx2 participants updated their estimates of environmental volatility with larger steps (albeit in a negative direction as indicated by the negative slope of the underlying trends in **Figure 6C**, reducing  $\mu_2$ ). No differences between anx2 and control participants in the  $\mu_1$  estimates were found. Neither did we obtain between-group differences in  $\sigma_1$ .

To understand why anx2 did not substantially differ from the control group in their expectation of reward yet had significantly lower volatility estimates (resembling those of the anx1 group), we looked more closely at **Figure 5—figure supplement 1**. This figure shows the HGF trajectories for perceptual beliefs and related quantities for a series of simulated responses. The results indicate that lower expectation of volatility can result from a smaller variance in the distribution of observed feedback scores, but also from a behavior characterized by smaller changes from trial to trial in the performance variable ( $\Delta cvIKI_{\text{trial}}$ ). Accordingly, as a post-hoc analysis, we tested whether anx2 participants had smaller variance in the distribution of feedback scores when compared to control participants. This was the case (means [SEM]) were 0.064 [0.004] in control participants and 0.052 [0.003] in anx2,  $P_{FDR} < 0.05$ ). Anx1 participants also contributed to a similar effect (means [SEM]) were 0.051 [0.002],  $P_{FDR} < 0.05$ , smaller in anx1 than in the control group). Furthermore, anx2 participants had, on average, smaller  $\Delta cvIKI_{\text{trial}}$  values than the control group (means [SEM]) were 0.005 [0.0011] in controls and 0.0032 [0.0007] in anx2,  $P_{FDR} < 0.05$ ). The same results were obtained for the anx1 group (0.0013 [0.0009],  $P_{FDR} < 0.05$ ). Thus, anx2 participants achieved high scores, as did control participants, yet they observed a reduced set of scores. In addition, their task-related behavioral



**Figure 6.** Computational modeling analysis. Data shown as mean and  $\pm$  SEM. (A) In the main experiment, anx1 participants underestimated the tendency for  $x_1$  (meaning their expectation on reward in the current trial was lower;  $P_{FDR} < 0.05$ ,  $\Delta = 0.75$ ,  $CI = [0.59, 0.89]$ , purple bar at the bottom). (B) In addition, the expectation on environmental (phasic) log-volatility ( $\log(\mu_2)$ ) was significantly smaller in anx1 participants than in control participants ( $P_{FDR} < 0.05$ ,  $\Delta = 0.71$ ,  $CI = [0.60, 0.81]$ ). Similar results were obtained in the anx2 group as compared to the control group ( $P_{FDR} < 0.05$ ,  $\Delta = 0.69$ ,  $CI = [0.53, 0.75]$ ). (C) The uncertainty about environmental volatility was higher in anx1 and anx2 relative to control participants. Figure 6 continued on next page

Figure 6 continued

(anx1:  $P_{FDR} < 0.05$ ,  $\Delta = 0.71$ ,  $CI = [0.65, 0.89]$ ; anx2:  $P_{FDR} < 0.05$ ,  $\Delta = 0.65$ ,  $CI = [0.52, 0.86]$ ). Larger  $\sigma_2$  in the anx1 and anx2 groups contributed to the larger update steps of the estimate  $\mu_2$ , shown in panel (B). (D–F) Same as panels (A–C) but in the separate control experiment. (D) The expectation on the reward tendency,  $\mu_1$ , was lower for anx3 participants relative to control participants ( $P_{FDR} < 0.05$ ,  $\Delta = 0.80$ ,  $CI = [0.68, 0.95]$ , denoted by the black bar at the bottom). (E) Same as panel (B): anx3 participants had a reduced expectation of environmental volatility ( $P_{FDR} < 0.05$ ,  $\Delta = 0.67$ ,  $CI = [0.55, 0.76]$ ). (F) Anx3 participants were also more uncertain about their phasic volatility estimates relative to control participants ( $P_{FDR} < 0.05$ ,  $\Delta = 0.65$ ,  $CI = [0.51, 0.77]$ ). Thus, the anxiety manipulation in the control experiment biased participants to make larger updates of their expectation of phasic volatility. The online version of this article includes the following figure supplement(s) for figure 6:

**Figure supplement 1.** Correlation between HGF volatility estimates and the variance of the distribution of feedback scores.

changes from trial to trial were more constrained. These smaller trial-to-trial behavioral changes in anx2 indicated a tendency to exploit their inferred optimal performance, leading to consistently high scores. This different strategy of successful performance ultimately accounted for the reduced estimation of environmental volatility in this group, and contrasted with the higher  $\mu_2$  values obtained in control participants.

As an additional post-hoc analysis, and based on the insights obtained from **Figure 5—figure supplement 1**, we assessed in the total population whether volatility estimates were associated with the change in performance variable  $\Delta cvIKI_{\text{trial}}$  or with the variance of the distribution of feedback scores. There was only a small yet significant non-parametric correlation between the HGF log-volatility estimates  $\log(\mu_2)$  and the variance of the distribution of feedback scores across the 200 trials (Spearman  $\rho = 0.3029$ ,  $P < 0.0190$ , **Figure 6—figure supplement 1**). This outcome suggests that participants who encountered more variable feedback scores in association with their performance also had a higher expectation of volatility.

Along with the above-mentioned group effects on relevant expectation and uncertainty trajectories, we found significant differences between anx1 and control participants in the perceptual parameter  $\omega_2$  (mean and SEM values:  $-5.2 [0.50]$  in controls,  $-3.6 [0.49]$  in anx1;  $P_{FDR} < 0.05$ ), but not in  $\omega_1$  ( $-4.8 [0.72]$  in controls,  $-4.8 [0.52]$  in anx1;  $P > 0.05$ ). Parameter  $\omega_2$  modulates the rate at which volatility changes, with higher values—as obtained in anx1 participants—leading to sharper and more pronounced steps of update in volatility (**Figure 5—figure supplement 2C**). This can also be described as a different learning style (**Weber et al., 2019**). Participants in the anx2 group did not differ from control participants in  $\omega_1$  ( $-4.1 [0.47]$ ,  $P > 0.05$ ) or  $\omega_2$  ( $-4.0 [0.74]$ ,  $P > 0.05$ ).

In the second experiment, in which anx3 participants demonstrated a pronounced drop in scores relative to those of control participants during the anxiety manipulation, we found that on the group level, the winning family of models was also the one associated with the performance parameter  $\Delta cvIKI_{\text{trial}}$  (model frequency 0.8747 and exceedance probability of 1). Further, the best individual model within that family was the one that explained  $\Delta cvIKI_{\text{trial}}^k$  as a function of  $\epsilon_1^{k-1}$  and  $\epsilon_2^{k-1}$  (exceedance probability of 1, and model frequency of 0.9051). Between-group comparisons in relevant model parameters demonstrated that, like anx1 participants in the main study, anx3 participants in this control experiment had a lower estimate of the mean tendency for  $x_1$  ( $P_{FDR} < 0.05$ ,  $\Delta = 0.80$ ,  $CI = [0.68, 0.95]$ ; **Figure 6D–F**), and also had a reduced expectation on environmental volatility ( $P_{FDR} < 0.05$ ,  $\Delta = 0.67$ ,  $CI = [0.55, 0.76]$ ). In addition, the anxiety manipulation led participants to have higher uncertainty about their phasic volatility estimates relative to control participants ( $P_{FDR} < 0.05$ ,  $\Delta = 0.65$ ,  $CI = [0.51, 0.77]$ ). No differences in the uncertainty about estimates for  $x_1$  were found. The perceptual parameters  $\omega_1$  and  $\omega_2$  did not differ between groups ( $P > 0.05$ ; average values of  $\omega_1$  and  $\omega_2$  were  $-4.9 [SEM 0.32]$  and  $-3.4 [0.41]$  in the control group, and  $-5.6 [0.39]$  and  $-4.4 [0.44]$  in the anx3 group). Last, among all response parameters,  $\beta_0, \beta_1, \beta_2, \zeta$ , we found that exclusively  $\beta_2$  (modulating the impact of  $\epsilon_2^{k-1}$  on  $\Delta cvIKI_{\text{trial}}^k$ ) was significantly different between groups (larger in control participants;  $P = 0.041$ ,  $\Delta = 0.68$ ,  $CI = [0.55, 0.76]$ ). Converging with the main experiment, parameters  $\beta_0$  and  $\beta_1$  were on average positive and negative, respectively, in each group.

**Electrophysiological analysis** The analysis of the EEG signals focused on sensorimotor and prefrontal (anterior) beta oscillations and aimed to assess separately (i) tonic and (ii) phasic (or event-related) changes in spectral power and burst rate. Tonic changes in average beta activity would be an indication that the anxiety manipulation had an effect on the general modulation of underlying

beta oscillatory properties. Complementing this analysis, assessment of the phasic changes in the measures of beta activity during trial performance and following feedback presentation allowed us to investigate the neural processes that drive reward-based motor learning and their alteration by anxiety. These analyses focused either on all channels (tonic changes) or on a subset of channels across contralateral sensorimotor cortices and anterior regions (phasic changes; see statistical analysis details in 'Materials and methods').

**State anxiety prolongs beta bursts and enhances beta power during exploration** We first looked at the general averaged properties of beta activity in this phase and their modulation by anxiety. The first measure we used was the standard averaged normalized power spectral density (PSD) of beta oscillations. Normalization of the raw PSD into decibels (dB) was carried out using the average PSD from the initial rest recordings (3 min) as reference. This analysis revealed a significantly higher beta-band power in a small contralateral sensorimotor region in anx1 participants relative to that in control participants during initial exploration ( $P < 0.025$ , two-sided cluster-based permutation test, FWE-corrected; **Figure 7—figure supplement 1**). In anx2 participants, the beta power in this phase was not significantly different than that in controls (**Figure 7—figure supplement 1**,  $P > 0.05$ ). No significant between-group changes in PSD were found in lower ( $< 13\text{Hz}$ ) or higher ( $> 30\text{Hz}$ ) frequency ranges ( $P > 0.05$ ).

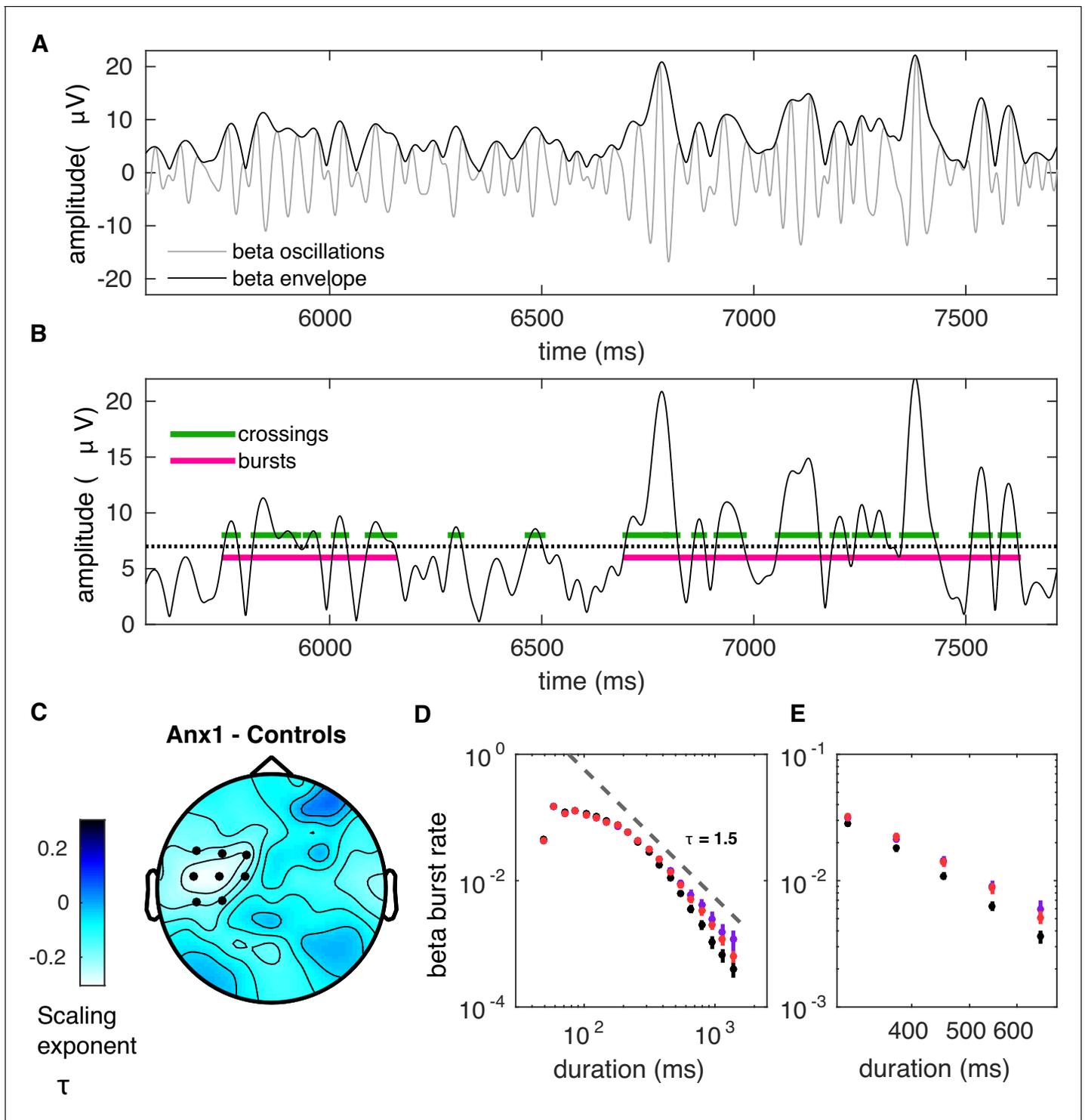
Next, we analyzed the between-group differences in the distribution of beta bursts extracted from the amplitude envelope of beta oscillations during initial exploration (**Figure 7A**). This analysis was motivated by evidence from recent studies suggesting that differences in the duration, rate, and onset of beta bursts could account for the association between beta power and movement in humans (**Little et al., 2018; Torrecillos et al., 2018**). To identify burst events and to assess the distribution of their duration, we applied an above-threshold detection method, which was adapted from previously described procedures (**Poil et al., 2008; Tinkhauser et al., 2017; Figure 7B**). In this analysis, we selected epochs locked to the GO signal at 0 s and extending up to 11 s. This interval included the STOP signal at 7 s and—in reward-based learning trials only—the feedback score at 9 s. Bursts extending for at least one cycle were selected. Using a double-logarithmic representation of the probability distribution of burst durations, we obtained a power law and extracted the (absolute) slope,  $\tau$ , also termed the 'life-time' exponent (**Poil et al., 2008**). Modeling work has revealed that a power law in the burst-duration distribution, reflecting the fact that the oscillation bursts have no characteristic scale, indicates that the underlying neural dynamics operate in a state close to criticality, and thus are beneficial for information processing (**Poil et al., 2008; Chialvo, 2010**).

Crucially, because the burst duration, rate, and slope provide complementary information, we focused our statistical analysis of the tonic beta burst properties on the slope or life-time exponent,  $\tau$ . A smaller slope corresponds to a burst distribution that is biased towards more frequent long bursts.

In all of our participants, the double-logarithmic representation of the distribution of burst duration followed a decaying power-law with slope values  $\tau$  in the range 1.4–1.9. The life-time exponents were smaller in the anx1 group than in the control group at left sensorimotor electrodes, corresponding with a long-tailed distribution (1.43 [0.30]; 1.70 [0.15];  $P_{FDR} < 0.05$ ,  $\Delta = 0.81$ ,  $CI = [0.75, 0.87]$ ). No differences in slope values  $\tau$  were found between anx2 and control participants. The smaller life-time exponents in anx1 in sensorimotor electrodes were also reflected in a longer mean burst duration: 182 (10) ms in the anx1 group, 153 (2) ms in control participants (166 [6] ms in anx2 participants). The differences in slope in the distribution of burst duration in anx1 reflected the more frequent presence of long bursts ( $> 500$  ms) and the less frequent brief bursts in this group relative to control participants (**Figure 7D–E**).

We next turned to our main goal and asked whether there were between-group differences in the beta oscillatory properties at specific periods throughout the initial exploration trials, above and beyond the general block-averaged changes reported above. The results in **Figure 4** establish that state anxiety during the initial exploration phase reduced task-related motor variability, but also subsequently led to impaired reward-based learning. We therefore sought to assess whether the anxiety-related reduction in motor variability during exploration was associated with altered dynamics in beta-band oscillatory activity at specific time intervals during trial performance.

In anx1 participants, the mean beta power increased after completion of the sequence performance and further following the STOP signal, and these changes were significantly more pronounced than in control participants ( $P_{FDR} < 0.05$ ,  $\Delta = 0.72$ ,  $CI = [0.63, 0.80]$ ; **Figure 8A**). This



**Figure 7.** Anxiety during initial exploration prolongs the life-time of sensorimotor beta-band oscillation bursts. (A) Illustration of the amplitude of beta oscillations (gray line) and the amplitude envelope (black line) for one representative subject and channel. (B) Schematic overview of the threshold-crossing procedure used to detect beta oscillation bursts. A threshold of 75% of the beta-band amplitude envelope was selected and beta bursts extending for at least one cycle were accepted. Windows of above-threshold amplitude crossings detected in the beta-band amplitude envelope (black line) are denoted by the green lines, whereas the windows of the associated bursts are marked by the magenta lines. (C) Scalp topography for between-group changes in the scaling exponent  $\tau$  during initial exploration. A significant positive cluster was found in an extended cluster of left sensorimotor electrodes, resulting from a smaller life-time exponent in anx1 than in control participants. (Black dots indicate significant electrodes, two-tailed cluster-based permutation test,  $P_{FWE} < 0.025$ .) (D) Probability distribution of beta-band oscillation-burst life-times within the 50–2000 ms range for Figure 7 continued on next page

Figure 7 continued

each group during initial exploration. The double-logarithmic representation reveals a power law within the fitted range (first duration bin excluded from the fit, as in [Poil et al. \[2008\]](#)). For each power law, we extracted the slope,  $\tau$ , also termed the life-time exponent. The dashed line illustrates a power law with  $\tau = 1.5$ . The smaller scaling exponent found in anx1 participants, as compared to control participants, was associated with long-tailed distributions of burst duration, reflecting the presence of frequent long bursts. Anx2 participants did not differ from control participants in the scaling exponent. Data are shown as mean and  $\pm$  SEM in the electrodes pertaining to the significant cluster in panel (C). (E) Enlarged display of panel (D) showing that bursts of duration 500 ms or longer were more frequent in anx1 than in control participants.

The online version of this article includes the following figure supplement(s) for figure 7:

**Figure supplement 1.** Sensorimotor beta power is modulated by anxiety during initial exploration.

significant effect was localized to contralateral sensorimotor and right prefrontal channels. As a post-hoc analysis, the time course of the burst rate was assessed separately in beta bursts of shorter (<300 ms) and longer (>500 ms) duration, following the results from [Figure 7](#) showing a pronounced dissociation between longer and brief bursts in the experimental and control groups. In addition, this split was motivated by previous studies linking longer beta bursts to detrimental performance (e.g. beta bursts longer than 500 ms in the basal ganglia of Parkinson's disease patients are associated with worse motor symptoms; [Tinkhauser et al., 2017](#)).

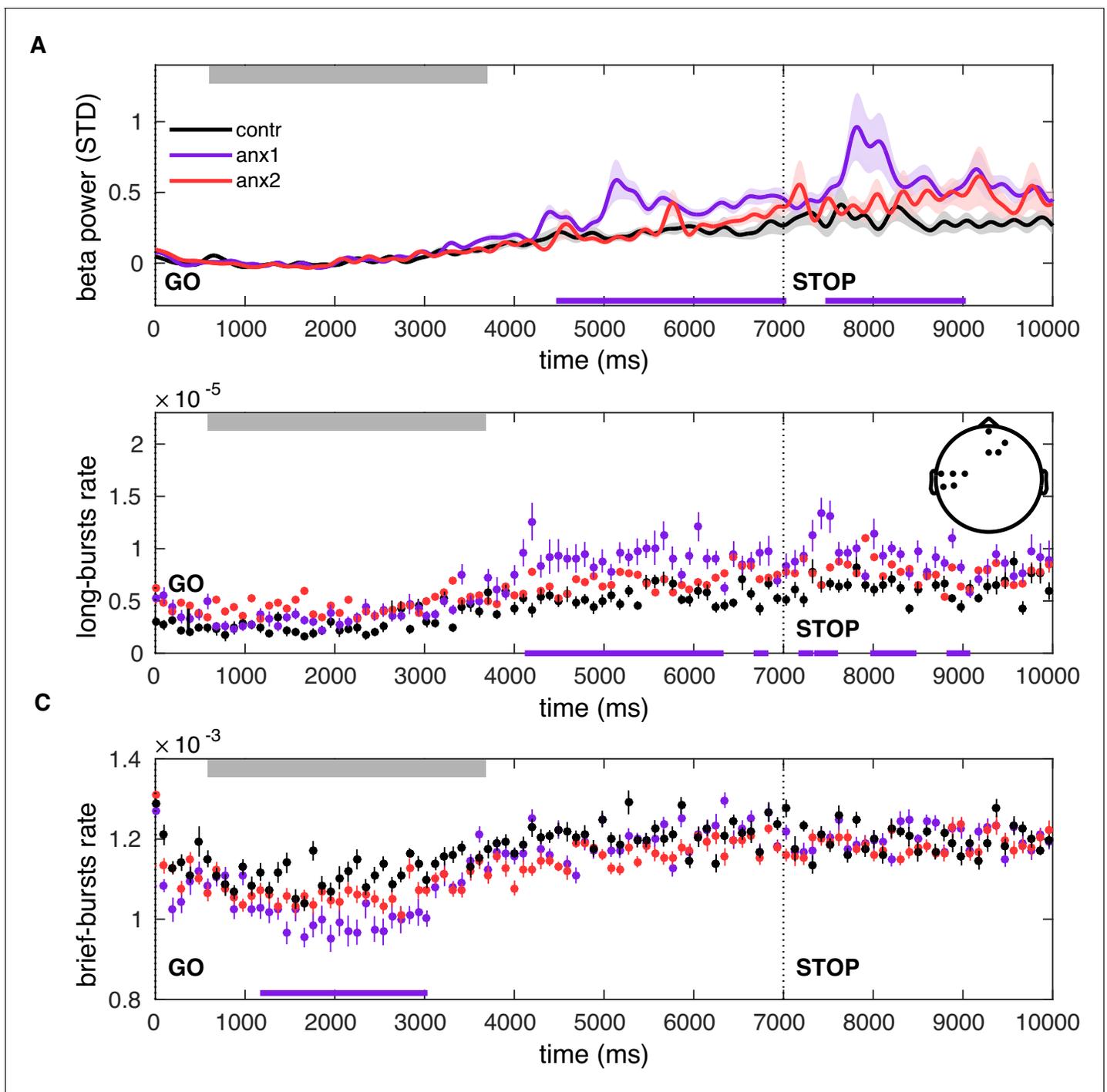
The rate of long oscillation bursts displayed a similar time course and topography to those of the power analysis, with an increased burst rate after movement termination and after the STOP signal in anx1 participants relative to control participants ( $P_{FDR} < 0.05$ ,  $\Delta = 0.69$ ,  $CI = [0.61, 0.78]$ ; [Figure 8B](#)). By contrast, brief burst events were less frequent in anx1 than in control participants, albeit exclusively during performance ( $P_{FDR} < 0.05$ ,  $\Delta = 0.74$ ,  $CI = [0.65, 0.82]$ ; [Figure 8C](#)). No significant effects were found when comparing any of these measures between anx2 and control participants.

Additional post-hoc control analyses were carried out to dissociate the separate effects of anxiety and motor performance on the time course of the beta-band oscillation properties during initial exploration. These analyses demonstrated that, when controlling for changes in motor variability, anxiety alone could explain the findings of larger post-movement beta-band PSD and rate of longer bursts, while also explaining the reduced rate of brief bursts during performance ([Figure 8—figure supplement 1](#)). Similar outcomes were found when controlling for changes in the mean total duration of the sequence ([Figure 8—figure supplement 2](#)), the variability of the sequence length (the coefficient of variation of sequence duration; [Figure 8—figure supplement 3](#)), and mean keystroke velocity ([Figure 8—figure supplement 4](#)).

Motor variability did also partially modulate the beta power and burst measures, after excluding anxious participants. This effect, however, had a small effect size and was limited to contralateral sensorimotor electrodes ([Figure 8—figure supplement 5](#)). In a last post-hoc analysis, we found that the average beta power following the STOP signal in those same significant sensorimotor electrodes was negatively correlated with the across-trials temporal variability, such that participants with a smaller increase in sensorimotor beta power after the STOP signal had a larger expression of task-related variability in this initial block (Spearman  $\rho = -0.4397$ ,  $P = 0.0001$ ; [Figure 8—figure supplement 6](#)).

**Reduced beta power and the presence of long beta bursts during feedback processing promotes the update of beliefs about reward** During learning, the general average level of PSD did not differ between groups ( $P_{FDR} < 0.05$ ; [Figure 9—figure supplement 1A–C](#)), neither was there a significant between-group difference in the scaling exponent of the distribution of beta-band oscillation bursts ( $P_{FDR} > 0.05$ , [Figure 9—figure supplement 1D–E](#); mean  $\tau$  across contralateral and prefrontal electrodes: 1.78 [0.06] in control, 1.61 [0.10] in anx1, 1.70 [0.06] in anx2 group). The lack of significant between-group differences in these measures indicated that during reward-based motor learning, there were no pronounced tonic changes in average beta activity induced by the previous (anx1) or concurrent (anx2) anxiety manipulation.

[Figure 4](#) had established that motor variability (or other motor output variables) did not differ in learning blocks between experimental and control groups, and therefore could not explain the significant and pronounced drop in scores in anx1 participants. Accordingly, we next aimed to assess whether alterations in the beta-band measures over time during trial performance or in feedback processing could account for that effect. In the anx1 group, the mean beta power increased towards the end of the sequence performance more prominently than in control participants, and this effect



**Figure 8.** Time course of the beta power and burst rate during trials in the exploration block. (A) The time representation of the beta power throughout trial performance shows two distinct time windows of increased power in participants affected by the anxiety manipulation: following sequence performance and after the STOP signal ( $P_{FDR} < 0.05$ ,  $\Delta = 0.72$ ,  $CI = [0.63, 0.80]$ ; black bars at the bottom indicate the windows of significant differences). Shaded areas indicate the SEM around the mean. Performance of sequence1 was completed on average between 680 (50) and 3680 (100) ms, denoted by the gray rectangle at the top. The STOP signal was displayed at 7000 ms after the GO signal, and the trial ended at 9000 ms. (B) The rate of oscillation bursts of longer duration ( $> 500$  ms) exhibited a similar temporal pattern, with increased burst rate in anx1 participants following movement and the STOP signal ( $P_{FDR} < 0.05$ ,  $\Delta = 0.69$ ,  $CI = [0.61, 0.78]$ ). (C) In contrast to the rate of long bursts, the rate of brief oscillation bursts was reduced in anx1 relative to control participants, albeit during performance ( $P_{FDR} < 0.05$ ,  $\Delta = 0.74$ ,  $CI = [0.65, 0.82]$ ). All averaged values in panels (A–C) are estimated across the significant sensorimotor and prefrontal electrodes shown in the inset in panel (B).

The online version of this article includes the following figure supplement(s) for figure 8:

Figure 8 continued on next page

Figure 8 continued

**Figure supplement 1.** Post-movement increases in the beta-band amplitude and burst rate can be explained by state anxiety after matching participants on temporal variability.

**Figure supplement 2.** Post-movement increases in the beta-band amplitude and burst rate can be explained by state anxiety after matching participants on the sequence duration.

**Figure supplement 3.** Post-movement increases in the beta-band amplitude and burst rate can be explained by state anxiety after matching participants on the variability of the total sequence duration.

**Figure supplement 4.** Post-movement increases in the beta-band amplitude and burst rate can be explained by state anxiety after matching participants on the mean keystroke velocity.

**Figure supplement 5.** Changes in motor variability without concurrent changes in state anxiety only partially account for the observed alterations in post-movement beta amplitude and burst rate.

**Figure supplement 6.** Correlation between average beta power and the degree of task-related behavioral variability across trials during the exploration phase.

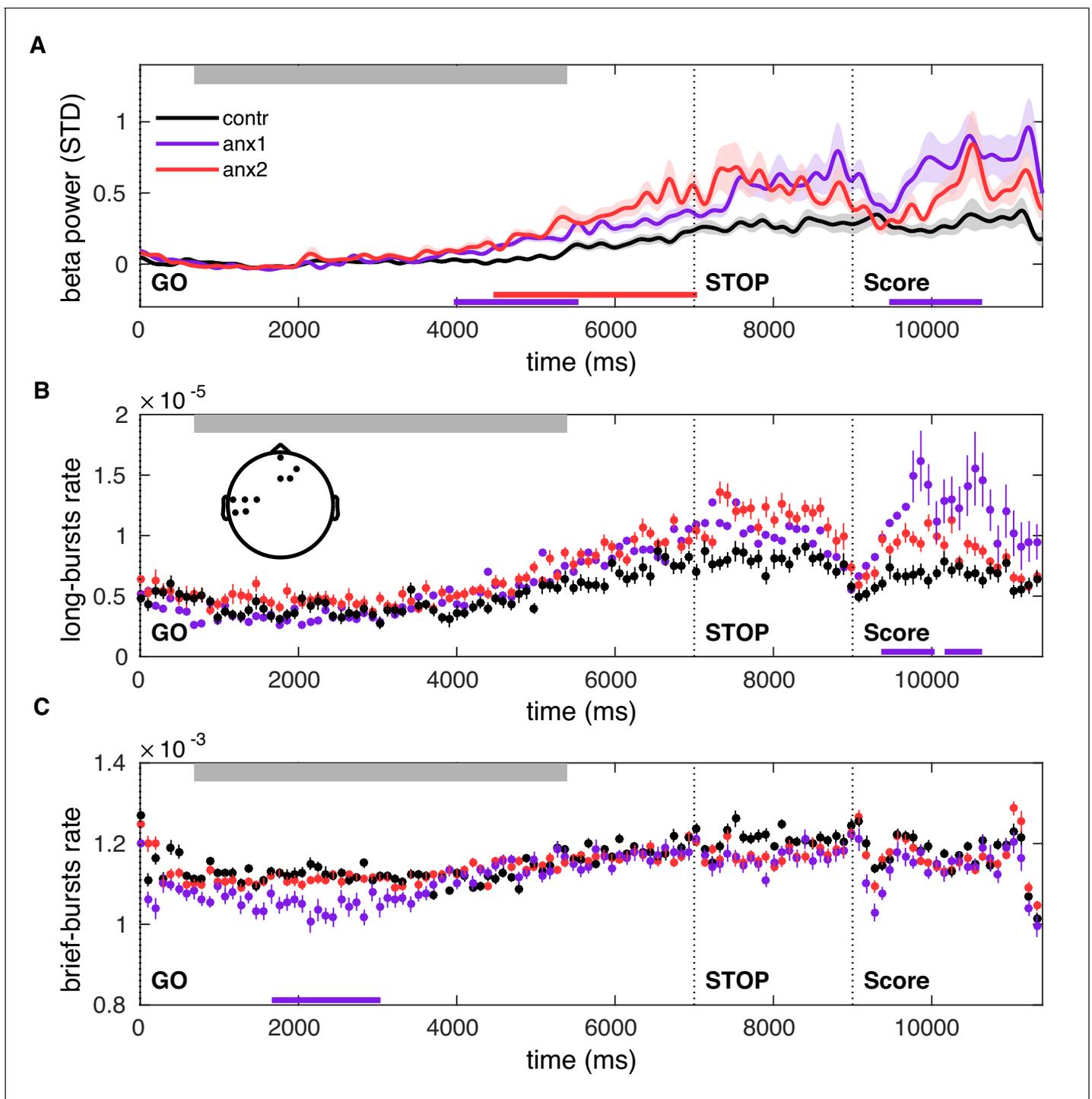
was significant in sensorimotor and prefrontal channels ( $P_{FDR} < 0.05$ ,  $\Delta = 0.67$ ,  $CI = [0.56, 0.78]$ ; **Figure 9A**). A significant increase with similar topography and latency was observed in the anx2 group relative to control participants ( $P_{FDR} < 0.05$ ,  $\Delta = 0.61$ ,  $CI = [0.56, 0.67]$ ). An additional and particularly pronounced enhancement in beta power appeared in anx1 and anx2 participants within 400–1600 ms following presentation of the feedback score. This post-feedback beta increase was significantly larger in anx1 than in the control group ( $P_{FDR} < 0.05$ ,  $\Delta = 0.65$ ,  $CI = [0.55, 0.75]$ ; no significant effect in anx2,  $P > 0.05$ ).

Further, we found that the time course of the beta burst rate exhibited a significant increase in anx1 participants relative to that in control participants within 400–1600 ms following feedback presentation, similar to the power results (**Figure 9B**;  $P_{FDR} < 0.05$ ,  $\Delta = 0.82$ ,  $CI = [0.70, 0.91]$ ). The rate of brief oscillation bursts was, by contrast, smaller in anx1 than in control participants, albeit exclusively during performance and not during feedback processing (**Figure 9C**;  $P_{FDR} < 0.05$ ,  $\Delta = 0.70$ ,  $CI = [0.56, 0.84]$ ). The significant effects in anx1 participants were observed in left sensorimotor and right prefrontal electrodes. There were no significant differences between anx2 and control groups in the rate of brief or long bursts throughout the trial ( $P > 0.05$ ).

To rule out the possibility that the feedback-related changes in beta activity were accounted for by concurrent movement-related artifacts (e.g. larger artifacts in anx1 than in control participants), we performed a control analysis of higher gamma band activity, which has been consistently associated with muscle artifacts in previous studies (**Muthukumaraswamy, 2013**). This control analysis found no evidence for movement artifacts affecting differently anx1 or control groups (**Figure 9—figure supplement 2**).

Having established that, relative to control participants, anx1 participants exhibited a phasic increase in beta activity and an increase in the rate of long bursts 400–1600 ms following feedback presentation, we next investigated whether these post-feedback beta changes could account for the altered reward and volatility estimates in the anx1 group (**Figure 6**). In the proposed predictive coding framework, superficial pyramidal cells encode PEs weighted by precision (precision-weighted PEs or pwPEs), and these are also the signals that are thought to dominate the EEG (**Friston and Kiebel, 2009**). A dissociation between high (gamma >30 Hz) and low (beta) frequency of oscillations has been proposed to correspond with the encoding of bottom-up PEs and top-down predictions, respectively (**Arnal and Giraud, 2012**). Operationally, however, beta oscillations have been associated with the change in predictions or expectations ( $\Delta\mu_i$ ) rather than with predictions themselves (**Sedley et al., 2016**). In the HGF, the update equations for  $\mu_1$  and  $\mu_2$  are determined exclusively by the pwPE term in that level, such that the change in predictions,  $\Delta\mu_i$ , is equal to pwPE (see **Equation 14** and **Equation 15**). Accordingly, we assessed whether the trialwise feedback-locked beta power or burst rate represented the magnitude of pwPEs in that trial that serve to update expectations on reward ( $\mu_1$ ) and environmental volatility ( $\mu_2$ ).

For each participant, we assessed simultaneously the effect of  $\epsilon_1$  and  $\epsilon_2$  on the trial-by-trial feedback-locked beta activity by running a multiple linear regression. These two regressors were not linearly correlated with each other (Pearson  $r$  coefficient in the total population was 0.1 on average [median = 0.1], and individual correlation  $p$ -values were  $P > 0.05$  in 80% of all participants). For the multiple linear regression analysis, trial-wise estimates of beta power (or burst rate) were averaged



**Figure 9.** Time course of the beta power and burst rate throughout trial performance and following reward feedback. (A) Time course of the feedback-locked beta power during sequence performance in the learning blocks, shown separately for anx1, anx2 and control groups. Average across sensorimotor and prefrontal electrode regions as in panel (B). Shaded areas indicate the SEM around the mean. Participants completed sequence2 on average between 720(30) and 5350 (100) ms, denoted by the top gray box. The STOP signal was displayed 7000 ms after the GO signal, and was followed at 9000 ms by the feedback score. This representation shows two distinct time windows of significant differences in beta activity between the anx1 and control groups: at the end of the sequence performance and subsequently following feedback presentation ( $P_{FDR} < 0.05$ ,  $\Delta = 0.65$ ,  $CI = [0.55, 0.75]$ ), respectively, denoted by the purple bar at the bottom). Anx2 participants also exhibited an enhanced beta power towards the end of the sequence performance ( $P_{FDR} < 0.05$ ,  $\Delta = 0.61$ ,  $CI = [0.56, 0.67]$ ). (B) Time course of the rate of longer (>500 ms) oscillation bursts during sequence performance in the learning blocks. Anx1 participants exhibited a prominent rise in the burst rate 400–1600 ms following the feedback. Figure 9 continued on next page

Figure 9 continued

score, which was significantly larger than the rate in control participants ( $P_{FDR} < 0.05$ ,  $\Delta = 0.82$ ,  $CI = [0.70, 0.91]$ ). Data display the mean and  $\pm$  SEM. The topographic map indicates the electrodes of significant effects for panels (A–C) ( $P_{FDR} < 0.05$ ). (C) Same as panel (B) but showing the rate of shorter beta bursts (<300 ms) during sequence performance in the learning blocks. Between-group comparisons demonstrated a significant drop in the rate of brief oscillation bursts in anx1 participants relative to control participants at the beginning of the performance ( $P_{FDR} < 0.05$ ,  $\Delta = 0.70$ ,  $CI = [0.56, 0.84]$ ), but not after the presentation of the feedback score. In all panels, the traces of the mean power and burst rates were displayed after averaging across the significant sensorimotor and prefrontal electrodes shown in the inset in panel (B).

The online version of this article includes the following figure supplement(s) for figure 9:

**Figure supplement 1.** Beta power spectral density and burst rate during reward-based learning.

**Figure supplement 2.** Higher gamma band activity analysis rules out an explanation in which muscle artifacts influence feedback-related changes in power.

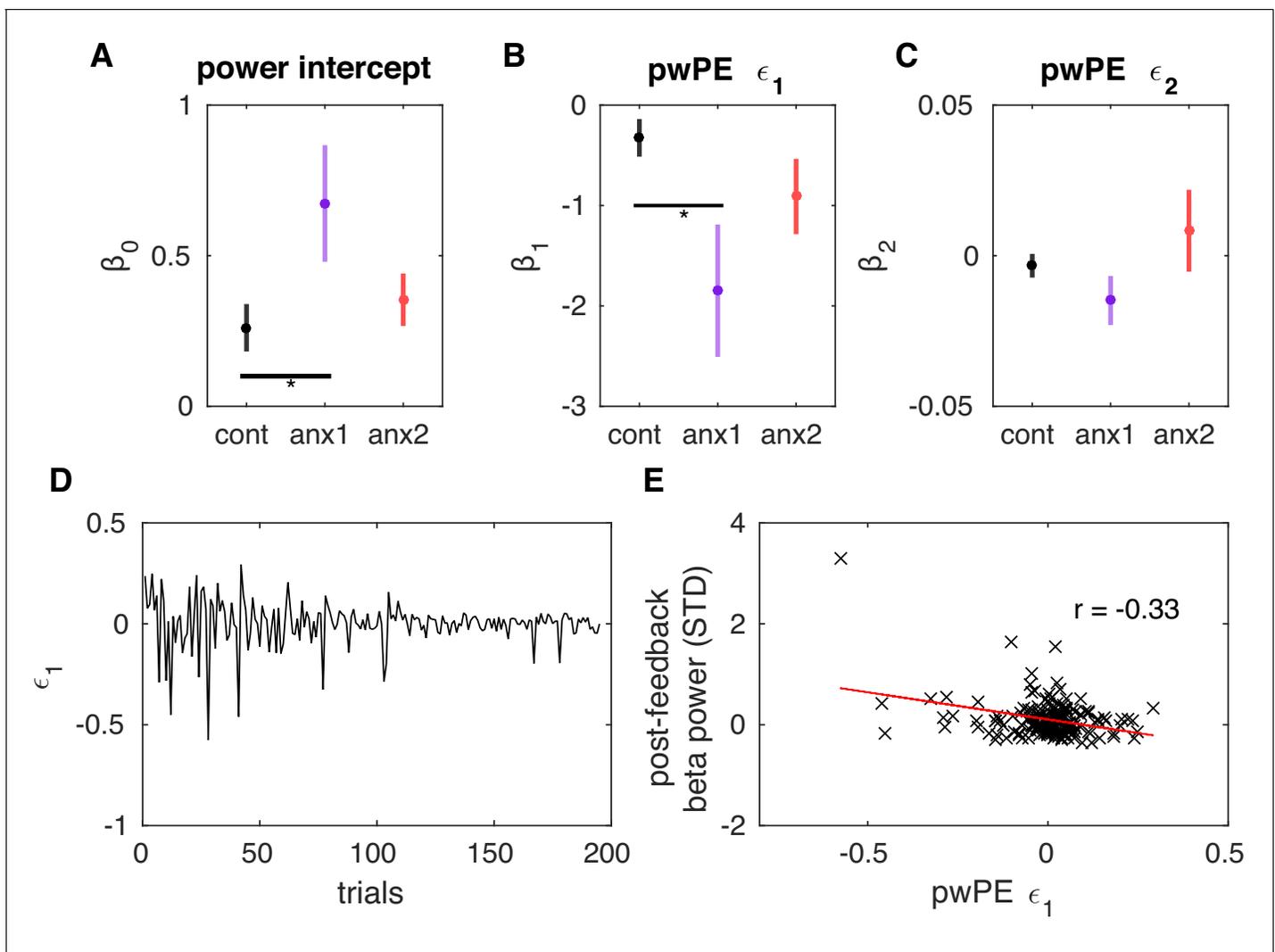
within 400–1600 ms following feedback presentation and across the sensorimotor and prefrontal electrodes where the post-feedback group effects were found (Figure 9). The results indicate that  $\epsilon_1$  had a significant negative effect on the measure of beta power (Figure 10; similarly for the rate of long bursts, see Figure 10—figure supplement 1), as  $\beta_1$  was significantly smaller than zero in each group ( $P_{FDR} < 0.05$ ). In addition, the  $\beta_1$  coefficient was decreased in anx1 relative to the control group ( $P_{FDR} < 0.05$ ,  $\Delta = 0.72$ ,  $CI = [0.57, 0.81]$ ; there were no differences between anx2 and control group). Thus, a reduction in  $\epsilon_1$  contributed to an increase in post-feedback beta power and the rate of long beta bursts. The intercept also significantly differed between anx1 and control groups, with a larger coefficient representing a larger level of post-feedback beta power as found in anx1 ( $P_{FDR} < 0.05$ ,  $\Delta = 0.69$ ,  $CI = [0.55, 0.75]$ ; no differences were obtained in anx2 relative to control participants). The  $\beta_2$  coefficient modulating the contribution of  $\epsilon_2$  to beta activity was not different than 0 in any group ( $P > 0.05$ ). Accordingly, these results provide evidence for a pattern of neural oscillatory modulation that is associated with the updating of beliefs about reward. Furthermore, they link enhanced post-feedback beta activity—as found in anx1—to reduced pwPE about reward.

**Discussion** The results revealed several interrelated mechanisms through which state anxiety

impairs reward-based motor learning. First, state anxiety reduced motor variability during an initial exploration phase. This was associated with limited improvement in scores during subsequent learning. Second, the smaller change in the expectation of reward throughout time led to a decrease in the expectation of volatility. Along with those results, we observed an overestimation of uncertainty about volatility due to state anxiety, which promoted the drop in the volatility estimate. Additional computational results demonstrated that larger precision-weighted prediction errors relating to reward and volatility had the effect of constraining the trial-to-trial behavioral adaptations in state anxiety. This contrasted with the findings for volatility in control participants, where larger pwPE relating to this quantity promoted behavioral exploration.

On the neural level, anxiety during initial exploration was associated with elevated sensorimotor beta power and a distribution of bursts of sensorimotor beta oscillations with a longer tail (smaller scaling exponent). The latter result indicated a more frequent presence of longer bursts, resembling recent findings of abnormal burst duration in movement disorders (Tinkhauser et al., 2017). The anxiety-induced higher rate of long burst events and higher beta power during initial exploration also manifested in prefrontal electrodes and extended to the following learning phase, where phasic trial-by-trial feedback-locked increases in these measures accounted for the attenuated updating of expectation on reward. These results provide the first evidence that state anxiety induces changes in the distribution of sensorimotor and prefrontal beta bursts, as well as in beta power, which may account for the observed deficits in the update of predictions during reward-based motor learning.

Evidence from our main experiment suggested that the finding of anxiety-related reduced motor variability during exploration was associated with the outcome of subsequently impaired learning from reward. These results validate previous accounts on the relationship between motor variability and Bayesian inference (Wu et al., 2014). In addition, the association between larger initial task-related variability and higher scores during the following learning phase extends results on the facilitatory effect of exploration on motor learning, at least in tasks that require learning from reinforcement (Wu et al., 2014; Pekny et al., 2015; Dhawale et al., 2017; see also critical view in He et al., 2016).



**Figure 10.** Post-feedback increases in beta power represent attenuated precision-weighted prediction errors about reward estimates. (A–C) Mean (and SEM) values of the  $\beta$  coefficients that explain the post-feedback beta power as a linear function of a constant value (beta power) (A), the precision-weighted prediction errors driving updates in the expectation of reward (pwPE,  $\epsilon_1$ ) (B), and pwPE driving updates in the expectation of volatility ( $\epsilon_2$ ) (C). The measure of beta power used here was the average within 400–1600 ms following feedback presentation and across sensorimotor and prefrontal electrodes, as shown in **Figure 9**. The  $\beta$  values are plotted separately for each control and experimental group. The  $\beta_0$  and  $\beta_1$  regression coefficients were significantly different from 0 in all groups ( $P_{FDR} < 0.05$ ). In addition,  $\beta_0$  was larger in the anx1 group relative to the control group ( $P_{FDR} < 0.05$ , denoted by the horizontal black line and the asterisk). In anx1 relative to control participants, we found that  $\beta_1$  was negative and significantly smaller in anx1 participants ( $P_{FDR} < 0.05$ ). Thus, a reduction in  $\epsilon_1$  contributed to an increase in post-feedback beta power. The multiple regression analysis did not support a significant contribution of the second regressor, pwPE relating to volatility, to explaining the changes in beta power (see main text, also  $\beta_2$  on average did not differ from 0 in any group of participants,  $P > 0.05$ ). (D) Illustration of the trajectories of pwPE  $\epsilon_1$  in one representative anx1 subject. (E) The linear regression between the trial-wise beta power and pwPE  $\epsilon_1$  for the same representative subject.

The online version of this article includes the following figure supplement(s) for figure 10:

**Figure supplement 1.** The rate of long beta bursts following feedback is modulated by the magnitude of precision-weighted prediction errors relating to reward.

**Figure supplement 2.** Topographic map illustrating the EEG channels used for the feedback-locked oscillatory analysis.

Crucially, state anxiety constrained the total amount of task-related variability only when induced during the initial exploration phase. The lack of between-group differences in cvIKI during learning in both experiments suggests that this measure could not account for the anxiety-related deficits in reward-based learning. Our Bayesian learning model provided additional insight on this aspect. The

modelling results suggested that state anxiety can impair learning from reward not only by influencing the posterior distributions of beliefs (expectations and uncertainty) but also by altering how pwPE relating to those beliefs affect behavioral variability. The response model consistently demonstrated in experimental and control groups that smaller pwPEs driving reward updates on the previous trial (leading to decreased expectation of reward) were followed by an increase in task-related motor variability (higher exploration). On the other side, trials of larger pwPE relating to reward were followed by reduced task-related behavioral changes. By contrast, the effect of pwPE on volatility differed substantially in control and anxiety groups. Although large pwPEs on volatility promoted subsequent larger task-related behavioral changes in control participants, they constrained behavioral exploration in the anx1 and anx2 groups.

Accordingly, state anxiety facilitated the use of task-related variability during reward-based learning only in trials following smaller pwPE reducing volatility estimates. This led participants who were affected by the prior or concurrent state anxiety manipulation to underestimate environmental volatility. Thus, they had the expectation that reward estimates are more stable throughout time. Anx1 and anx2 participants also had larger uncertainty about volatility. This implies that they were less confident about their volatility estimate, and allowed for a greater influence of new information in updating this quantity. This finding is additionally reinforced in anx1 by the result of a larger  $\omega_2$ , reflecting a different learning style (Weber et al., 2019) that is characterized by sharper and more pronounced steps of update in  $\mu_2$ . The results align well with recent computational work in decision-making tasks, showing that high trait anxiety leads to alterations in uncertainty estimates and adaptation to the changing statistical properties of the environment (Browning et al., 2015; Huang et al., 2017; Pulcu and Browning, 2019).

Notwithstanding the similarities in the anx1 and anx2 groups concerning the expectation of volatility and associated uncertainty, the fact that anx2 participants achieved high scores in the task and were not impaired in learning requires further clarification. Our post-hoc analyses revealed that the drop in  $\mu_2$  in anx2 could be accounted for by the narrower distribution of scores encountered by this group. In addition, these participants introduced smaller trial-to-trial changes in temporal variability when compared to control participants. Thus, anx2 participants had a tendency to exploit the current motor program more than control participants, suggesting a more conservative approach to success. Anx1 participants also introduced smaller trial-to-trial changes in temporal variability, yet their behavioral changes had a slower benefit on reward. In both groups, however, the more pronounced tendency to exploit the current motor program was associated with alterations in how pwPE relating to volatility influenced behavioral changes. Overall, our findings provide the first evidence that computational mechanisms similar to those described for trait anxiety and decision-making underlie the effect of temporary anxious states on motor learning. This might be the case particularly in the context of learning from rewards, such as feedback about success or failure, which is considered one of the fundamental processes through which motor learning is accomplished (Wolpert et al., 2011).

Previous studies manipulating psychological stress and anxiety to assess motor learning showed both a deleterious and a facilitatory effect (Hordacre et al., 2016; Vine et al., 2013; Bellomo et al., 2018). Differences in experimental tasks, which often assess motor learning during or after high-stress situations but not during anxiety induction in anticipation of a stressor, could account for the previous mixed results. Here, we adhered to the neurobiological definition of anxiety as a psychological and physiological response to an upcoming diffuse and unpredictable threat (Grupe and Nitschke, 2013; Bishop, 2007). Accordingly, anxiety was induced using the threat of an upcoming public speaking task (Feldman et al., 2004; Lang et al., 2015), and was associated with a drop in the HRV and an increase in state anxiety scores during the targeted blocks. Although the average state anxiety scores were not particularly high, they were significantly higher during the targeted phases than during the initial resting state phase. Future studies should use more impactful stressors to study the effect of the full spectrum of state (and trait) anxiety on motor learning (Bellomo et al., 2018).

What is the relationship between the expression of motor variability and state anxiety? As hypothesized, state anxiety during initial exploration reduced the use of variability across trials. This converges with recent evidence demonstrating that anxiety leads to ritualistic behavior (repetition, redundancy, and rigidity of movements) that allow the subject to regain a sense of control (Lang et al., 2015). The outcome also aligns well with animal studies in which evidence shows a

reduction in motor exploration when the stakes are high (high-reward situations, social context; *Kao et al., 2008; Dhawale et al., 2017; Woolley et al., 2014*). These interpretations, however, seem to stand in contrast with our findings in anx2 participants, which were affected by the anxiety manipulation during learning but with no significant effect on the total degree of motor variability expressed during this phase. Similar results were obtained in the second experiment, as anx3 and control participants did not differ in the amount of across-trials variability expressed during learning. Bayesian computational modelling clarified these findings by demonstrating that anx2 participants had a tendency to increase the exploitation of their current motor program. And their trial-to-trial changes in temporal variability were smaller than those in the control group, particularly following large pwPE that increase the expectation on volatility. This outcome was also found in anx1 participants, as well as in anx3 participants in the second experiment. Thus, anxiety had the consistent effect to constrain dynamic trial-to-trial changes in temporal variability, and these changes were negatively influenced by pwPE on volatility. Notably, however, the strategy in anx2 participants of more extensively exploiting the inferred rewarded solution than control participants was successful, and therefore differs from the learning impairment exhibited by anx1 participants. In the second experiment, removal of the initial exploration phase led to the impairment of reward-based learning in anx3 participants. This group also demonstrated a tendency to explore less than control participants on a trial to trial level as a function of increases in volatility pwPEs. Thus, the combined evidence suggests that the normal use of initial variability in anx2 participants might have protected them from the effects of the following anxiety manipulation. Initial use of variability in anx2 might have promoted faster learning of the mapping between actions and their associated outcome, contributing to successful goal-directed exploitation. The results further favor the interpretation that initial unconstrained exploration is important for *subsequent* successful motor learning.

Some considerations should be taken into account. Task-related motor variability might be pivotal for learning from reinforcement or reward signals (*Sutton and Barto, 1998; Dhawale et al., 2017; Wu et al., 2014*), whereas in other contexts, such as during motor adaptation, the evidence is conflicting (*He et al., 2016; Singh et al., 2016*). An additional consideration is that greater levels of motor variability could reflect both an intentional pursuit of an explorative regime and an unintentional higher level of motor noise, in the latter case similar to that observed in previous work (*Wu et al., 2014; Pekny et al., 2015*). A recent study established that motor learning is improved by the use of intended exploration, not motor noise (*Chen et al., 2017*). Our paradigm cannot dissociate intended and unintended exploration. This limitation will be addressed in future work by using a separate initial phase with regular performance to assess motor noise as a measure of unintended exploration.

Another consideration is that our use of an initial exploration phase that did not provide reinforcement or feedback signals was motivated by the work of *Wu et al. (2014)*, which demonstrated a correlation between initial variability (no feedback) and learning curve steepness in a subsequent reward-based learning phase—a relationship previously observed in the zebra finch (*Kao et al., 2005; Olveczky et al., 2005; Ölveczky et al., 2011*). This suggests that higher levels of motor variability do not solely amount to increased noise in the system. Instead, this variability represents a broader action space that can be capitalized upon during subsequent reinforcement learning by searching through previously explored actions (Herzfeld and Shadmehr, 2014). Accordingly, an implication of our results is that state anxiety could impair the potential benefits of an initial exploratory phase for subsequent learning.

Last, we used a reward-based motor learning paradigm in which different performances could provide the same feedback score. The rationale for using this task was to explore the effect of state anxiety on volatility estimates, as recent work demonstrates that anxiety primarily affects learning in volatile conditions (*Browning et al., 2015; Huang et al., 2017*). This scenario, however, implied that a high expression of task-related motor variability during learning would be associated with a more volatile perception of the task, which is indeed supported by our correlation results. This could be a confounding factor when explaining the group effects. Importantly, however, further analyses revealed that the total degree of motor variability during learning and the mean learned performance did not differ between groups, suggesting that these are not confounding factors that could explain the reward-based-learning group results. Instead, our findings underscore that computational mechanisms related to how pwPE on reward and volatility influence behavioral changes are

the main factors driving the effects of concurrent or prior state anxiety on reward-based motor learning.

At the neural level, an important finding was that anxiety during initial exploration increased the power of beta oscillations and the rate of long beta bursts (long-tailed distribution). The increases in power and the rate of long-lived bursts manifested after completion of the sequence, reflecting an anxiety-related enhancement of the post-movement beta rebound (Kilavik *et al.*, 2012; Kilavik *et al.*, 2013). This effect was observed in a region of contralateral sensorimotor and right prefrontal channels, and could be explained by anxiety alone, despite a small effect of motor variability on the modulation of these neural changes across sensorimotor electrodes. Further, larger sensorimotor beta power at the termination of the sequence performance was associated with a more constrained use of task-related variability. Our analyses did not provide a detailed anatomical localization of the effect, but the findings in sensorimotor regions that partially contribute to changes in motor variability are consistent with the involvement of premotor and motor cortex in driving motor variability and learning, as previously reported in animal studies (Churchland *et al.*, 2006; Mandelblat-Cerf *et al.*, 2009; Santos *et al.*, 2015). The results also converge with the representation in the premotor cortex of temporal and sequential aspects of rhythmic performance (Crowe *et al.*, 2014; Kornysheva and Diedrichsen, 2014).

During learning, an unexpected result was that, in anx2 participants, there was an increase in beta power at the end of the sequence performance but not during feedback processing—and despite the anxiety manipulation successfully affecting the HRV. This outcome, as well as the lack of beta burst effects in this group, seems to be in agreement with the lack of learning impairments when compared with control participants. An additional unexpected result during learning blocks was the presence in anx1 participants of higher rates of long bursts and greater beta power at the end of the trial and during feedback processing, across both sensorimotor and prefrontal electrodes. These phasic changes in beta activity in anx1 participants extended from the previous phase, and the outcome aligns with the finding of prefrontal involvement in the emergence and maintenance of anxiety states (Davidson, 2002; Grupe and Nitschke, 2013; Bishop, 2007). Thus, our results revealed that, in the context of motor learning, anxious states induce changes in sensorimotor and prefrontal beta power and burst distribution. These changes are maintained after physiological measures of anxiety return to baseline, and thus continue to affect relevant behavioral parameters. Anxiety has been shown to modulate different oscillatory bands depending on the context, such as gamma activity in visual areas and amygdala when processing fearful faces (Schneider *et al.*, 2018), alpha activity in response to processing emotional faces (Knyazev *et al.*, 2008) or theta activity during rumination (Andersen *et al.*, 2009). Beta-band oscillations could be particularly relevant to flesh out the effects of anxiety on performance during motor tasks.

Mechanistically, phasic trial-by-trial feedback-locked changes in the sensorimotor beta power and burst distribution were related to the computational alterations in updating expectations on reward found in anx1 participants, and thus explained their poorer performance during reward-based learning. Specifically, a higher rate of long beta bursts and increased power following feedback were associated with a reduced update in the expectation of reward.

The computational quantity that determines the update of expectations in our Bayesian model is the precision-weighted PEs. Here, pwPE relating to reward were inversely related to the rate of long beta bursts and beta power, and were therefore attenuated in anx1 participants because of their enhanced feedback-related beta activity. We found no significant contribution of pwPE relating to volatility to explaining changes in beta activity, suggesting that additional frequency ranges should be considered when linking hierarchical pwPEs to neural oscillations during learning. In the context of the predictive coding hypothesis, PEs (or pwPEs) are hypothesized to be mediated by gamma oscillations, whereas the neuronal signaling of predictions is mediated by lower frequencies (e.g., alpha 8–12 Hz, Friston *et al.*, 2015). Further studies point to beta oscillations as the cortical oscillatory rhythm associated with encoding predictions, although the evidence to date is scarce (Arnal and Giraud, 2012). More recently, beta oscillations have been associated with the change to predictions rather than with predictions themselves (Sedley *et al.*, 2016), which is consistent with our findings as pwPEs were the quantities determining the change to predictions. In line with these results, a post-performance increase in beta power during motor adaptation is considered to index confidence in priors, and thus a reduced tendency to change the ongoing motor command (Tan *et al.*, 2016). More generally, beta oscillations along cortico-basal ganglia networks have been

proposed to gate incoming information to modulate behavior (*Leventhal et al., 2012*) and to maintain the current motor state (*Engel and Fries, 2010*). Consequently, the phasic increase in beta power and the rate of beta bursts following feedback presentation could represent neural states that impair the encoding of pwPEs and the update of predictions about lower level quantities—reward here—induced by anxiety states. Notably, the modulation of feedback-locked beta activity was not explained by changes in pwPE relating to volatility. We speculate that the effect of reduced reward estimates on the expectation of volatility in the HGF suggests that abnormal increases in beta activity following feedback presentation indirectly influenced volatility estimates, while it had a direct effect on reward expectation.

Our findings show that the assessment of neural activity in sensorimotor regions is crucial to understanding the effects of anxiety on motor learning and to determining mechanisms, above and beyond the role of prefrontal control of attention, in mediating the effects of anxiety on cognitive and perceptual tasks (*Bishop, 2007; Bishop, 2009; Eysenck and Calvo, 1992*). Our data imply that the combination of Bayesian learning models and analysis of oscillation properties can help us to better understand the mechanisms through which anxiety modulates motor learning. Future studies should investigate how the brain circuits that are involved in anxiety interact with motor regions to affect motor learning. In addition, assessing burst properties across both beta and gamma frequency ranges would further allow us to delineate and dissociate the neural mechanisms responsible for anxiety biasing decision-making and motor learning.

**Materials and methods**  
**Participants and sample-size estimation** Sixty right-handed healthy volunteers (37 females) aged 18 to 44 (mean 27 years, SEM, 1 year) participated in the main study. In a second, control experiment, 26 right-handed healthy participants (16 females, mean age 25.8, SEM 1, range 19–40) took part in the study. Participants gave written informed consent prior to the start of the experiment, which had been approved by the local Ethics Committee at Goldsmiths University. Participants received a base rate of either course credits or money (15 GBP; equally distributed across groups) and were able to earn an additional sum up to 20 GBP during the task depending on their performance.

We used pilot data from a behavioral study using the same motor task to estimate the minimum sample sizes for a statistical power of 0.95, with an  $\alpha$  of 0.05, using the MATLAB (The MathWorks, Inc., MA, USA) function `sampsizepwr`. In the pilot study, we had one control and one experimental group of 20 participants each. In the experimental group, we manipulated the reward structure during the first reward-based learning block (in this block, feedback scores did not count towards the final average monetary reward). For each behavioral measure (motor variability and mean score), we extracted the standard deviation (sd) of the joint distribution from both groups and the mean value of each separate distribution (e.g.,  $m_1$ , control;  $m_2$ , experimental), which provided the following minimum sample sizes:

Between-group comparison of behavioral parameters (using a two-tailed t-test):  
 $\text{MinSampleSizeA} = \text{sampsizepwr}('t', [m_1 \text{ sd}], m_2, 0.95) = 18\text{--}20$  participants.

Accordingly, we recruited 20 participants for each group in the main experiment. Next, using the behavioral data from the anxiety and control groups in the current main experiment, we estimated the minimum sample size for the second, behavioral control experiment:

Between-group comparison of behavioral parameters (using a two-tailed t-test):  
 $\text{MinSampleSizeA} = \text{sampsizepwr}('t', [m_1 \text{ sd}], m_2, 0.95) = 13$  participants.

Therefore, for the second control experiment, we recruited 13 participants for each group.

**Apparatus** Participants were seated at a digital piano (Yamaha Digital Piano P-255, London, UK) and in front of a PC monitor in a light-dimmed room. They sat comfortably in an arm-chair with their forearms resting on the armrests of the chair. The screen displayed the instructions, feedback and visual cues for the start and end of a trial (*Figure 1A*). Participants were asked to place four fingers of their right hand (excluding the thumb) comfortably on four pre-defined keys on the keyboard. Performance information was transmitted and saved as Musical Instrument Digital Interface (MIDI) data, which provided time onsets of keystrokes relative to the previous one (inter-keystroke-interval—IKI in ms), MIDI velocities (related to the loudness, in arbitrary units, a.u.), and MIDI note numbers that corresponded to the pitch. The experiment was run using Visual Basic, an additional parallel port and MIDI libraries.

**Materials and experimental design** In all blocks, participants initiated the trial by pressing a pre-defined key with their left index finger. After a jittered interval of 1–2 s, a green ellipse appeared

in the center of the screen representing the GO signal for task execution (**Figure 1A**). Participants had 7 s to perform the sequence, which was ample time to complete it before the green circle turned red indicating the end of the execution time. If participants failed to perform the sequence in the correct order or initiated the sequence before the GO signal, the screen turned yellow. In blocks 2 and 3 during learning, performance-based feedback in the form of a score between 0 and 100 was displayed on the screen 2 s after the red ellipse, that is, 9 s from the beginning of the trial. The scores provided participants with information regarding the target performance.

The performance measure that was rewarded during learning was the Euclidean norm of the vector corresponding to the pattern of temporal differences between adjacent IKIs for a trial-specific performance. Here, we denote the vector norm by  $\|\Delta z\|$ , with  $\Delta z$  being the vector of differences,  $\Delta z = (z_2 - z_1, z_3 - z_2, \dots, z_n - z_{n-1})$ , and  $z_i$  representing the IKI at each keystroke ( $i = 1, 2, \dots, n$ ). Note that IKI values themselves represent the difference between the onset of consecutive keystrokes, and therefore  $\Delta z$  indicates a vector of differences of differences. Specifically, the target value of the performance measure was a vector norm of 1.9596 (e.g., one of the maximally rewarded performances leading to this vector norm of IKI-differences would consist of IKI values: [0.2, 1, 0.2, 1, 0.2, 1, 0.2] s; that is a combination of short and long intervals). The score was computed in each trial using a measure of proximity between the target vector norm  $\|\Delta z^*\|$  and the norm of the performed pattern of IKI differences  $\|\Delta z^p\|$ , using the following expression:

$$\text{score} = 100\exp(-\|\Delta z^*\| - \|\Delta z^p\|) \quad (3)$$

In practice, different temporal patterns leading to the same vector norm  $\|\Delta z^p\|$  could achieve the same score. Participants were unaware of the existence of various solutions. Higher exploration across trials during learning could thus reveal that several IKI patterns were similarly rewarded. To account for this possibility, the perceived rate of change of the hidden goal (environmental volatility) during learning was estimated and incorporated into our mathematical description of the relationship between performance and reward (see below).

**Anxiety manipulation** Anxiety was induced during block1 performance in group anx1, and during block2 performance in the anx2 group by informing participants about the need to give a 2 min speech to a panel of experts about an unknown art object at the end of that block (**Lang et al., 2015**). We specified that they would first see the object at the end of the block (it was a copy of Wassily Kandinsky' Reciprocal Accords [1942]) and would have 2 min to prepare for the presentation. Participants were told that the panel of experts would take notes during their speech and would be standing in front of the testing room (due to the EEG setup participants had to remain seated in front of the piano). Following the 2 min preparation period, participants were informed that due to the momentary absence of panel members, they instead had to present in front of the lab members. Participants in the control group had the task of describing the artistic object to themselves, and not in front of a panel of experts. They were informed about this secondary task before the beginning of block1.

**Assessment of state anxiety** To assess state anxiety, we acquired two types of data: (1) the short version of the Spielberger State-Trait Anxiety Inventory (STAI, state scale X1, 20 items; **Spielberger, 1970**) and (2) a continuous electrocardiogram (ECG, see EEG, ECG and MIDI recording session). The STAI X1 subscale was presented four times throughout the experiment. A baseline assessment at the start of the experiment before the resting state recording was followed by an assessment immediately before each experimental block to determine changes in anxiety levels. In addition, a continuous ECG recording was obtained during the resting state and three experimental blocks were used to assess changes in autonomic nervous system responses. The indexes of heart rate variability (HRV, coefficient of variation of the inter-beat-interval) and mean heart rate (HR) were evaluated, as their reduction has been linked to changes in anxiety state due to a stressor (**Feldman et al., 2004**).

**Computational model** Here, we provide details on the computational Bayesian model that we adopted to estimate participant-specific belief trajectories, determined by the mean (expectation) and variance (uncertainty) of the posterior distribution. The model was implemented using the HGF toolbox for MATLAB (<http://www.translationalneuromodeling.org/tapas/>). The model consists of a perceptual and a response model, representing an agent (a Bayesian observer) who generates behavioral responses on the basis of a sequence of sensory inputs that it receives. In many

implementations of the HGF, the sensory input is replaced with a series of outcomes (e.g. feedback, reward) associated with participants' responses (*de Berker et al., 2016; Diaconescu et al., 2017*). As general notation, we let lowercase italics denote scalars ( $x$ ), which can be further characterized by a trial superscript  $x^k$  and a subscript  $i$  denoting the level in the hierarchy  $x_i^k$  ( $i = 1, 2$ ). We use lower-case bold font for vectors with  $n$  components,  $\mathbf{x}$ .

The HGF corresponds to the perceptual model, representing a hierarchical belief-updating process, that is a process that infers hierarchically related environmental states that give rise to sensory inputs (*Stefanics et al., 2018; Mathys et al., 2014*). In the version for continuous inputs (see *Mathys et al. (2014)*; function `tapas_hgf.m`), we used the series of feedback scores as input:  $u^k = \text{score}$ ; normalized to range 0–1. From the series of inputs, the HGF then generates belief trajectories about external states, such as the reward value of an action or a stimulus. Learning occurs in two hierarchically coupled levels ( $x_1, x_2$ ), one for 'perceptual' beliefs ( $x_1$ : the reward associated with the current performance), and the phasic volatility of those beliefs ( $x_2$ ). These two levels evolve as coupled Gaussian random walks, with the lower level coupled to the higher level through its variance (inverse precision). The Gaussian random walk at each level  $x_i$  is determined by its posterior mean ( $\mu_i$ ) and its variance ( $\sigma_i$ ). Further, the variance of the lower level,  $x_1$ , depends on  $x_2$  through an exponential function:

$$f(x_2) = \exp(\kappa x_2 + \omega_1) \tag{4}$$

where  $\kappa$  was fixed to 1 and  $\omega_1$  is a model parameter that was estimated for each participant by fitting the HGF model to the experimental data (scores and responses) using Variational Bayes.

At the top level, the variance is typically fixed to a constant parameter,  $\vartheta = \exp(\omega_2)$ , where  $\omega_2$  is also a free parameter to be estimated in each individual. The specific coupling between levels indicated above has the advantage of allowing simple variational inversion of the model and the derivation of one-step update equations under a mean-field approximation. This is achieved by iteratively integrating out all previous states up to the current trial  $k$  (see appendices in *Mathys et al. [2014]*). Importantly, the update equations for the posterior mean at level  $i$  and for trial  $k$  depend on the prediction errors weighted by uncertainty  $\sigma_i$  (or its inverse, precision  $\pi_i = 1/\sigma_i$ ) according to the following expression:

$$\Delta\mu_i^k = \mu_i^k - \mu_i^{k-1} \propto \frac{\hat{\pi}_{i-1}^k}{\pi_i^k} \delta_{i-1}^k \tag{5}$$

The first term in the above expression is the change in the expectation for state  $x_i$  on trial  $k$ ,  $\mu_i^k$ , relative to the prediction on trial  $k-1$ ,  $\mu_i^{k-1}$ . The prediction on trial  $k-1$  is denoted by the 'hat' or diacritical mark  $\hat{\cdot}$ ,  $\mu_i^{k-1} = \hat{\mu}_i^k$ . The term prediction thus refers to the expectation of  $x_i$  before seeing the feedback score from the current trial: it corresponds with the mean of the posterior distribution of  $x_i$  up to trial  $k-1$ . By contrast, the term expectation refers to the mean of the posterior distribution of  $x_i$  up to trial  $k$ . The difference term  $\Delta\mu_i^k$  is proportional to the prediction error of the level below,  $\delta_{i-1}^k$ , representing the difference between the expectation  $\mu_{i-1}^k$  and the prediction  $\hat{\mu}_{i-1}^k$  of the level below. The prediction error is weighted by the ratio between the prediction of the precision of the level below,  $\hat{\pi}_{i-1}^k$ , and the precision of the current belief,  $\pi_i^k$ . Thus the product of the precision weights and the prediction error constitute the precision-weighted prediction error (pwPE), which therefore regulates the update of expectations on trial  $k$ :  $\Delta\mu_i^k = e_i^k$ . The pwPE expressions for level 1 and 2 are defined below in **Equation 14** and **Equation 15**. **Equation 5** illustrates that higher uncertainty in the current level ( $\sigma_i^k$ , lower  $\pi_i^k$  in the denominator) leads to faster update of beliefs; moreover, smaller uncertainty (higher precision) of the prediction of the level below also increases the update of beliefs. For the two-level HGF model for continuous inputs, the generic equation **Equation 5** takes the explicit forms shown below (**Equation 6** and **Equation 10**; equations taken directly from the TAPAS toolbox; see also *Mathys et al. [2011]; Mathys et al. [2014]*).

Updates of expectations for level 1:

$$\mu_1^k = \hat{\mu}_1^k + \frac{\hat{\pi}_u^k}{\pi_1^k} \delta_u^k, \tag{6}$$

with  $\hat{\pi}_u^k$  representing the prediction of the precision of the input (feedback scores; see **Table 1**) and  $\delta_u^k$  the prediction error about the input:

$$\delta_u^k = u^k - \hat{\mu}_1^k, \tag{7}$$

Precision updates for level 1:

$$\pi_1^k = \hat{\pi}_1^k + \pi_u^k, \tag{8}$$

where  $\hat{\pi}_1^k$  is defined as (using  $\rho = 0, \kappa = 1, t^k$ ):

$$\hat{\pi}_1^k = \frac{1}{\left(\frac{1}{\pi_1^{k-1}} + \exp(\mu_2^{k-1} + \omega_1)\right)}, \tag{9}$$

Update of expectations for level 2:

$$\mu_2^k = \hat{\mu}_2^k + \frac{1}{2} \frac{1}{\pi_2^k} w_1^k \delta_1^k, \tag{10}$$

with

$$w_1^k = \exp(\mu_2^{k-1} + \omega_1) \hat{\pi}_1^k \tag{11}$$

Precision updates for level 2:

$$\pi_2^k = \hat{\pi}_2^k + \frac{1}{2} w_1^k (w_1^k + (2w_1^k - 1)\delta_1^k), \tag{12}$$

and

$$\hat{\pi}_2^k = \frac{1}{\frac{1}{\pi_2^{k-1}} + \exp(\omega_2)}. \tag{13}$$

From **Equation 6** and **Equation 10**, it follows that the pwPEs for level 1 and 2,  $\epsilon_1$  and  $\epsilon_2$ , respectively, are:

$$\epsilon_1^k = \mu_1^k - \hat{\mu}_1^k = \frac{\hat{\pi}_u^k}{\pi_1^k} \delta_u^k, \tag{14}$$

$$\epsilon_2^k = \mu_2^k - \hat{\mu}_2^k = \frac{1}{2} \frac{1}{\pi_2^k} w_1^k \delta_1^k. \tag{15}$$

Next, we mapped the expectation on the inferred perceptual beliefs, reward  $\mu_1$  and volatility  $\mu_2$ , and the corresponding pwPEs to the performance output that the participant generates during every trial using a separate response model. We adapted the family of response models used by **Marshall et al. (2016)** to our task. In that work, the authors explained participant’s observed log (RT) responses on a trial-by-trial basis as a linear function of various HGF quantities using a multiple regression. We implemented similar models, but adapted them to our task (new scripts are available in the Open Science Framework Data Repository: <https://osf.io/sg3u7/>). The models we tested used two different performance parameters:

The coefficient of variation of inter-keystroke intervals,  $cvIKI_{\text{trial}}$ , as a measure of the extent of timing variability within the trial.

The logarithm of the mean performance tempo in a trial,  $\log(mIKI_{\text{trial}})$ , with IKI in milliseconds.

We were interested in how HGF quantities on the previous trial explained changes in the performance parameters in the subsequent trial and therefore used these dependent variables:

$$\Delta cvIKI_{\text{trial}}^k = cvIKI_{\text{trial}}^k - cvIKI_{\text{trial}}^{k-1}$$

$$\Delta \log(mIKI_{\text{trial}})^k = \log(mIKI_{\text{trial}}^k) - \log(mIKI_{\text{trial}}^{k-1})$$

For each of those two performance measures, the corresponding response model was a function of a constant component of the performance measure (intercept) and HGF quantities on the previous trial, such as: the expectation on reward ( $\mu_1$ ), the expectation on volatility ( $\mu_2$ ), the precision-weighted PE relating to reward ( $\epsilon_1$ ), or the precision-weighted PE relating to volatility ( $\epsilon_2$ ). In total, we assessed the following two families of four alternative response models HGF11-14 and HGF21-24.

Model HGF11:

$$\Delta cvIKI_{\text{trial}}^k = \beta_0 + \beta_1 \mu_1^{k-1} + \beta_2 \epsilon_1^{k-1} + \zeta$$

Model HGF12:

$$\Delta cvIKI_{\text{trial}}^k = \beta_0 + \beta_1 \mu_1^{k-1} + \beta_2 \mu_2^{k-1} + \zeta \tag{16}$$

Model HGF13:

$$\Delta cvIKI_{\text{trial}}^k = \beta_0 + \beta_1 \mu_2^{k-1} + \beta_2 \epsilon_2^{k-1} + \zeta \tag{17}$$

Model HGF14:

$$\Delta cvIKI_{\text{trial}}^k = \beta_0 + \beta_1 \epsilon_1^{k-1} + \beta_2 \epsilon_2^{k-1} + \zeta$$

Model HGF21:

$$\Delta \log(mIKI_{\text{trial}})^k = \beta_0 + \beta_1 \mu_1^{k-1} + \beta_2 \epsilon_1^{k-1} + \zeta \tag{18}$$

Model HGF22:

$$\Delta \log(mIKI_{\text{trial}})^k = \beta_0 + \beta_1 \mu_1^{k-1} + \beta_2 \mu_2^{k-1} + \zeta \tag{19}$$

Model HGF23:

**Table 1.** Means and variances of the priors on perceptual parameters and initial values.

Priors on the parameters and initial values of the HGF perceptual model for continuous inputs. The continuous inputs here were the trial-by-trial scores that the participants received, normalized to the 0–1 range. Quantities estimated in the logarithmic space are denoted by log(). Prior mean and variance for  $\mu_1^0$ , as well as the prior mean for  $\sigma_1^0$ ,  $\omega_1$  and the precision of the input,  $\pi_u^0$ , were defined by the initial 20 input values. When providing prior values that depend on the first 20 input scores, we indicate the median across the total population of 60 participants. For the remaining quantities, the prior mean and variance were pre-defined according to the values indicated in the table.

	Prior mean	Prior variance
$\log(\kappa)$	$\log(1)$	0
$\omega_1$	log-variance of 1:20 input scores: -3.04	16
$\omega_2$	-4	16
$\log(\pi_u^0)$	negative log-variance of 1:20 input scores: 3.04	4
$\mu_1^0$	value of the first input score: 0.21	variance of 1:20 input scores: 0.05
$\log(\sigma_1^0)$	log-variance of 1:20 input scores: -3.04	1
$\mu_2^0$	1	0
$\log(\sigma_2^0)$	$\log(0.01)$	1
$\beta_0$	individual mean of behavioral parameter	4
$\beta_1$	0	4
$\beta_2$	0	4

$$\Delta \log(\text{mIKI}_{\text{trial}})^k = \beta_0 + \beta_1 \mu_2^{k-1} + \beta_2 \epsilon_2^{k-1} + \zeta \quad (20)$$

Model HGF24:

$$\Delta \log(\text{mIKI}_{\text{trial}})^k = \beta_0 + \beta_1 \epsilon_1^{k-1} + \beta_2 \epsilon_2^{k-1} + \zeta \quad (21)$$

The priors on the model parameters ( $\omega_1, \omega_2$ ), the response model parameters ( $\beta_0, \beta_1, \beta_2, \zeta$ ), the initial expected states ( $\mu_1^0, \mu_2^0$ ) and the precision of the input ( $\pi_u$ ) are provided in **Table 1**. All priors are Gaussian distributions in the space in which they are estimated and are therefore determined by their mean and variance. The variance is relatively broad to let the priors be modified by the series of inputs (feedback scores). Quantities that need to be positive (e.g., the variance or uncertainty of belief trajectories) are estimated in the log-space, whereas general unbounded quantities are estimated in their original space.

We used Random Effects Bayesian Model Selection (BMS) to assess the different models of learning at the group level (**Stephan et al., 2009**; code freely available from the MACS toolbox, **Soch and Allefeld, 2018**). First, the log-model evidence (LME) values for models HGF11-14 were combined to get the log-family evidence (LFE), and similarly for models HGF21-24. The LFE values were subsequently compared using BMS to assess which family of models provided more evidence. BMS generated (i) the estimated model-family frequencies, that is, how frequently each family of models is optimal in the sample of participants; and (ii) the exceedance probabilities, reflecting the posterior probability that one family is more frequent than the others (**Soch et al., 2016**). In the winner family, additional BMS determined the final optimal model.

**EEG, ECG and MIDI recording** EEG and ECG signals were recorded using a 64-channel (extended international 10–20 system) EEG system (ActiveTwo, BioSemi Inc.) placed in an electromagnetically shielded room. During the recording, the data were high-pass filtered at 0.16 Hz. The vertical and horizontal eye-movements (EOG) were monitored by electrodes above and below the right eye and from the outer canthi of both eyes, respectively. Additional external electrodes were placed on both left and right earlobes as reference. The ECG was recorded using two external channels with a bipolar ECG lead II configuration. The sampling frequency was 512 Hz. Onsets of visual stimuli, key presses and metronome beats were automatically documented with markers in the EEG file. The performance was additionally recorded as MIDI files using the software Visual Basic and a standard MIDI sequencer program on a Windows Computer.

**EEG and ECG pre-processing** We used MATLAB and the FieldTrip toolbox (**Oostenveld et al., 2011**) for visualization, filtering and independent component analysis (ICA; runica). The EEG data were highpass-filtered at 0.5 Hz (Hamming windowed sinc finite impulse response [FIR] filter, 3380 points) and notch-filtered at 50 Hz (847 points). Artifact components in the EEG data related to eye blinks, eye movements and the cardiac-field artifact were identified using ICA. Following IC inspection, we used the EEGLAB toolbox (**Delorme and Makeig, 2004**) to interpolate missing or noisy channels using spherical interpolation. Finally, we transformed the data into common average reference.

Analysis of the ECG data with FieldTrip focused on detection of the QRS-complex to extract the R-peak latencies of each heartbeat and use them to evaluate the HRV and HR measures in each experimental block.

**Analysis of power spectral density** We first assessed the standard power spectral density (PSD, in  $\text{mV}^2/\text{Hz}$ ) of the continuous raw data in each performance block and separately for each group. The PSD was computed with the standard fast Fourier Transform (Welch method, Hanning window of 1 s with 50% overlap). The raw PSD estimation was normalized into decibels (dB) with the average PSD from the initial rest recordings (3 min). Specifically, the normalized PSD during the performance blocks was calculated as ten times the base-10 logarithm of the quotient between the performance-block PSD and the resting state power.

In addition, we assessed the time course of the spectral power over time during performance. Trials during sequence performance were extracted from  $-1$  to  $11$  s locked to the GO signal. This interval included the STOP signal (red ellipse), which was displayed at  $7$  s, and—exclusively in learning blocks—the score feedback, which was presented at  $9$  s. Thus, epochs were effectively also locked to the STOP and Score signals. Artifact-free EEG epochs were decomposed into their time-frequency representations using a 7-cycle Morlet wavelet in successive overlapping windows of  $50$

ms within the total 12s-epoch. The frequency domain was sampled within the beta range from 13 to 30 Hz at 1 Hz intervals. For each trial, we thus obtained the complex wavelet transform, and computed its squared norm to extract the wavelet energy (Ruiz et al., 2009). The time-varying spectral power was then simply estimated by averaging the wavelet energy across trials. This measure of spectral power was further averaged within the beta-band frequency bins and normalized by subtracting the mean and dividing by the standard deviation of the power estimate in the pre-movement baseline period ( $[-1, 0]$  s prior to the GO signal).

**Extraction of beta-band oscillation bursts** We estimated the distribution, onset and duration of oscillation bursts in the time series of beta-band amplitude envelope. We followed a procedure adapted from previous work to identify oscillation bursts (Poil et al., 2008; Tinkhauser et al., 2017). In brief, we used as threshold the 75% percentile of the amplitude envelope of beta oscillations. Amplitude values above this threshold were considered to be part of an oscillation burst if they extended for at least one cycle (50 ms, as a compromise between the duration of one 13 Hz-cycle [76 ms] and 30 Hz-cycle [33 ms]). Threshold-crossings that were separated by less than 50 ms were considered to be part of the same oscillation burst. As an additional threshold, the median amplitude was used in a control analysis, which revealed qualitatively similar results, as expected from previous work (Poil et al., 2008). Importantly, because threshold crossings are affected by the signal-to-noise ratio in the recording, which could vary between the different performance blocks, we selected a common threshold from the initial rest recordings separately for each participant (Tinkhauser et al., 2017). Distributions of the rate of oscillation bursts per duration were estimated using equidistant binning on a logarithmic axis with 20 bins between 50 ms and 2000 ms.

General burst properties were assessed during exploration and learning blocks separately, first as averaged values within the full block-related recording, and next as phasic changes over time during trial performance. Trial-based analysis focused on the interval 0–11000 ms following the GO signal, which included the time window following the STOP signal (at 7000 ms: exploration and learning blocks) and the score feedback (at 9000 ms: learning blocks).

**Statistical analysis** Statistical analysis of behavioral and neural measures focused on the separate comparison between each experimental group and the control group (contrasts: anx1 – controls, anx2 –controls). Differences between experimental groups, anx1 – anx2, were evaluated exclusively concerning the overall achieved monetary reward. We used non-parametric pair-wise permutation tests to assess differences between conditions or between groups in the statistical analysis of behavioral or computational measures. When multiple testing was performed, we implemented a control of the false discovery rate (FDR) at level  $q = 0.05$  using an adaptive linear step-up procedure (Benjamini et al., 2006). This control provided an adapted threshold p-value (termed  $P_{FDR}$ ). Further, to evaluate differences between sets of multi-channel EEG signals corresponding to two conditions or groups, we used two approaches:

1. Tonic changes in average beta PSD or the scaling exponent of the burst distribution were assessed using two-sided cluster-based permutation tests (Maris and Oostenveld, 2007) and an alpha level of 0.025. Here, we used all 64 channels and let the statistical method extract the significant clusters. Control of the family-wise error (FWE) rate was implemented in these tests to account for the problem of multiple comparison (Maris and Oostenveld, 2007).
2. Phasic or event-related changes in beta power or burst rate across time were assessed using pair-wise permutation tests at each time point and exclusively in a subset of channels across sensorimotor and anterior (prefrontal) electrode regions (Figure 10—figure supplement 1). The relevant subset was chosen to ameliorate the number of multiple comparisons arising from time and space—channels). When using these tests, we implemented a control of the FDR at level  $q = 0.05$  to correct for multiple comparisons.

Non-parametric effect size estimators were used in association with our pair-wise nonparametric statistics, following Grissom and Kim (2012). In the case of between-subject comparisons, the standard probability of superiority,  $\Delta$ , was used.  $\Delta$  is defined as the proportion of greater values in sample B relative to A, when values in samples A and B are not paired:  $\Delta = P(A > B)$  ranges from 0 to 1. The total number of comparisons is the product of the size of sample A and sample B ( $N_{tot} = sizeA * sizeB$ ), and therefore,  $\Delta = N(A > B) / N_{tot}$ . In the case of ties,  $\Delta$  is corrected by subtracting in the denominator the number of ties from the total number of comparisons ( $N_{tot} - N_{ties}$ ). For within-subject comparisons, we used the probability of superiority for dependent samples,  $\Delta_{dep}$ , which is the proportion of all within-subject (paired) comparisons in which the values for condition B are larger than for condition A. Confidence intervals (CI) for  $\Delta$  and  $\Delta_{dep}$  were estimated with

bootstrap methods (*Ruscio and Mullen, 2012*). Last, associations between parameters were quantified using non-parametric rank correlations (Spearman  $\rho$ ), which are robust against outliers. However, we used linear correlations in the case of multiple linear regressions for the HGF response model, following *Marshall et al. (2016)*.

[transparent\\_reporting.pdf](#)

**Acknowledgements** This research is supported by the British Academy through grant R134610 to MHR and by the Economic and Social Research Council through PhD grant ES/P00072X/1 to TPH. MHR was partially supported by the HSE Basic Research Program and the Russian Academic Excellence Project '5–100'. We thank Marta García Huesca and Silvia Aguirre for carrying out some of the EEG experiments.

**References** Andersen SB, Moore RA, Venables L, Corr PJ. 2009. Electrophysiological correlates of anxious rumination. *International Journal of Psychophysiology* 71:156–169.

Arnal LH, Giraud AL. 2012. Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences* 16:390–398.

Bartolo R, Merchant H. 2015. Oscillations Are Linked to the Initiation of Sensory-Cued Movement Sequences and the Internal Guidance of Regular Tapping in the Monkey. *Journal of Neuroscience* 35:4635–4640.

Baumeister RF. 1984. Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology* 46:610–620.

Beilock SL, Carr TH. 2001. On the fragility of skilled performance: what governs choking under pressure? *Journal of Experimental Psychology: General* 130:701–725.

Bellomo E, Cooke A, Hardy J. 2018. Chunking, conscious processing, and EEG during sequence acquisition and performance pressure: a comprehensive test of reinvestment theory. *Journal of Sport and Exercise Psychology* 40:135–145.

Benjamini Y, Krieger AM, Yekutieli D. 2006. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93:491–507.

Bishop SJ. 2007. Neurocognitive mechanisms of anxiety: an integrative account. *Trends in Cognitive Sciences* 11:307–316.

Bishop SJ. 2009. Trait anxiety and impoverished prefrontal control of attention. *Nature Neuroscience* 12:92–98.

Browning M, Behrens TE, Jocham G, O'Reilly JX, Bishop SJ. 2015. Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience* 18:590–596.

Chen X, Mohr K, Galea JM. 2017. Predicting explorative motor learning using decision-making and motor noise. *PLOS Computational Biology* 13:e1005503.

Chialvo DR. 2010. Emergent complex neural dynamics. *Nature Physics* 6:744–750.

Churchland MM, Afshar A, Shenoy KV. 2006. A central source of movement variability. *Neuron* 52:1085–1096.

Crowe DA, Zarco W, Bartolo R, Merchant H. 2014. Dynamic representation of the temporal and sequential structure of rhythmic movements in the primate medial premotor cortex. *Journal of Neuroscience* 34:11972–11983.

Davidson RJ. 2002. Anxiety and affective style: role of prefrontal cortex and amygdala. *Biological Psychiatry* 51:68–80.

de Berker AO, Rutledge RB, Mathys C, Marshall L, Cross GF, Dolan RJ, Bestmann S. 2016. Computations of uncertainty mediate acute stress responses in humans. *Nature Communications* 7:10996.

Delorme A, Makeig S. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* 134:9–21.

Dhawale AK, Smith MA, Ölveczky BP. 2017. The role of variability in motor learning. *Annual Review of Neuroscience* 40:479–498.

Diaconescu AO, Litvak V, Mathys C, Kasper L, Friston KJ, Stephan KE. 2017. A computational hierarchy in human cortex. *arXiv*. <https://arxiv.org/pdf/1709.02323.pdf>.

Engel AK, Fries P. 2010. Beta-band oscillations—signalling the status quo? *Current Opinion in Neurobiology* 20:156–165.

Eysenck MW, Calvo MG. 1992. Anxiety and performance: the processing efficiency theory. *Cognition & Emotion* 6:409–434.

- Feingold J, Gibson DJ, DePasquale B, Graybiel AM. 2015. Bursts of beta oscillation differentiate postperformance activity in the striatum and motor cortex of monkeys performing movement tasks. *PNAS* 112:13687–13692.
- Feldman PJ, Cohen S, Hamrick N, Lepore SJ. 2004. Psychological stress, appraisal, emotion and cardiovascular response in a public speaking task. *Psychology & Health* 19:353–368.
- Friston KJ, Bastos AM, Pinotsis D, Litvak V. 2015. LFP and oscillations—what do they tell us? *Current Opinion in Neurobiology* 31:1–6.
- Friston K, Kiebel S. 2009. Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364:1211–1221.
- Grupe DW, Nitschke JB. 2013. Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience* 14:488–501.
- HajiHosseini A, Rodríguez-Fornells A, Marco-Pallarés J. 2012. The role of beta-gamma oscillations in unexpected rewards processing. *NeuroImage* 60:1678–1685.
- He K, Liang Y, Abdollahi F, Fisher Bittmann M, Kording K, Wei K. 2016. The statistical determinants of the speed of motor learning. *PLOS Computational Biology* 12:e1005023.
- Herrojo Ruiz M, Brücke C, Nikulin VV, Schneider GH, Kühn AA. 2014. Beta-band amplitude oscillations in the human internal globus pallidus support the encoding of sequence boundaries during initial sensorimotor sequence learning. *NeuroImage* 85:779–793.
- Hordacre B, Immink MA, Ridding MC, Hillier S. 2016. Perceptual-motor learning benefits from increased stress and anxiety. *Human Movement Science* 49:36–46.
- Huang H, Thompson W, Paulus MP. 2017. Computational dysfunctions in anxiety: failure to Differentiate Signal From Noise. *Biological Psychiatry* 82:440–446.
- Kao MH, Doupe AJ, Brainard MS. 2005. Contributions of an avian basal ganglia–forebrain circuit to real-time modulation of song. *Nature* 433:638–643.
- Kao MH, Wright BD, Doupe AJ. 2008. Neurons in a forebrain nucleus required for vocal plasticity rapidly switch between precise firing and variable bursting depending on social context. *Journal of Neuroscience* 28:13232–13247.
- Kilavik BE, Ponce-Alvarez A, Trachel R, Confais J, Takerkart S, Riehle A. 2012. Context-related frequency modulations of macaque motor cortical LFP beta oscillations. *Cerebral Cortex* 22:2148–2159.
- Kilavik BE, Zaepffel M, Brovelli A, MacKay WA, Riehle A. 2013. The ups and downs of  $\beta$  oscillations in sensorimotor cortex. *Experimental Neurology* 245:15–26.
- Knyazev GG, Bocharov AV, Levin EA, Savostyanov AN, Slobodskoj-Plusnin JY. 2008. Anxiety and oscillatory responses to emotional facial expressions. *Brain Research* 1227:174–188.
- Kornysheva K, Diedrichsen J. 2014. Human premotor Areas parse sequences into their spatial and temporal features. *eLife* 3:e03043.
- Lang M, Krátký J, Shaver JH, Jerotijević D, Xygalatas D. 2015. Effects of anxiety on spontaneous ritualized behavior. *Current Biology* 25:1892–1897.
- Leventhal DK, Gage GJ, Schmidt R, Pettibone JR, Case AC, Berke JD. 2012. Basal ganglia beta oscillations accompany cue utilization. *Neuron* 73:523–536.
- Little S, Bonaiuto J, Barnes G, Bestmann S. 2018. Motor cortical beta transients delay movement initiation and track errors. *bioRxiv*.
- Mandelblat-Cerf Y, Paz R, Vaadia E. 2009. Trial-to-trial variability of single cells in motor cortices is dynamically modified during visuomotor adaptation. *Journal of Neuroscience* 29:15053–15062.
- Maris E, Oostenveld R. 2007. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods* 164:177–190.
- Marshall L, Mathys C, Ruge D, de Berker AO, Dayan P, Stephan KE, Bestmann S. 2016. Pharmacological fingerprints of contextual uncertainty. *PLOS Biology* 14:e1002575.
- Mathys C, Daunizeau J, Friston KJ, Stephan KE. 2011. A bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience* 5:39.
- Mathys CD, Lomakina EI, Daunizeau J, Iglesias S, Brodersen KH, Friston KJ, Stephan KE. 2014. Uncertainty in perception and the hierarchical gaussian filter. *Frontiers in Human Neuroscience* 8:825.
- Morgan KN, Tromborg CT. 2007. Sources of stress in Captivity. *Applied Animal Behaviour Science* 102:262–302.

- Muthukumaraswamy SD. 2013. High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. *Frontiers in Human Neuroscience* 7:138.
- Olveczky BP, Andalman AS, Fee MS. 2005. Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLOS Biology* 3:e153.
- Ölveczky BP, Otchy TM, Goldberg JH, Aronov D, Fee MS. 2011. Changes in the neural control of a complex motor sequence during learning. *Journal of Neurophysiology* 106:386–397.
- Oostenveld R, Fries P, Maris E, Schoffelen JM. 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience* 2011:1–9.
- Ouellet C, Langlois F, Provencher MD, Gosselin P. 2019. Intolerance of uncertainty and difficulties in emotion regulation: proposal for an integrative model of generalized anxiety disorder. *European Review of Applied Psychology* 69:9–18.
- Pekny SE, Izawa J, Shadmehr R. 2015. Reward-dependent modulation of movement variability. *Journal of Neuroscience* 35:4015–4024.
- Pijpers JR, Oudejans RRD, Bakker FC. 2005. Anxiety-Induced changes in movement behaviour during the execution of a complex whole-body task. *The Quarterly Journal of Experimental Psychology Section A* 58:421–445.
- Poil SS, van Ooyen A, Linkenkaer-Hansen K. 2008. Avalanche dynamics of human brain oscillations: relation to critical branching processes and temporal correlations. *Human Brain Mapping* 29:770–777.
- Robinson OJ, Pike AC, Cornwell B, Grillon C. 2019. The translational neural circuitry of anxiety. *Journal of Neurology, Neurosurgery & Psychiatry* 90:1360.
- Ruiz MH, Koelsch S, Bhattacharya J. 2009. Decrease in early right alpha band phase synchronization and late gamma band oscillations in processing syntax in music. *Human Brain Mapping* 30:1207–1225.
- Ruscio J, Mullen T. 2012. Confidence intervals for the probability of superiority effect size measure and the area under a receiver operating characteristic curve. *Multivariate Behavioral Research* 47:201–223.
- Santos FJ, Oliveira RF, Jin X, Costa RM. 2015. Corticostriatal dynamics encode the refinement of specific behavioral variability during skill learning. *eLife* 4:e09423.
- Schneider TR, Hipp JF, Domnick C, Carl C, Büchel C, Engel AK. 2018. Modulation of neuronal oscillatory activity in the beta- and gamma-band is associated with current individual anxiety levels. *NeuroImage* 178:423–434.
- Sedley W, Gander PE, Kumar S, Kovach CK, Oya H, Kawasaki H, Howard MA, Griffiths TD. 2016. Neural signatures of perceptual inference. *eLife* 5:e11476.
- Singh P, Jana S, Ghosal A, Murthy A. 2016. Exploration of joint redundancy but not task space variability facilitates supervised motor learning. *PNAS* 113:14414–14419.
- Soch J, Haynes J-D, Allefeld C. 2016. How to avoid mismodelling in GLM-based fMRI data analysis: cross-validated Bayesian model selection. *NeuroImage* 141:469–489.
- Soch J, Allefeld C. 2018. MACS - a new SPM toolbox for model assessment, comparison and selection. *Journal of Neuroscience Methods* 306:19–31.
- Spielberger C. 1970. *Manual for the State-Trait Anxiety Inventory (Self-Evaluation Questionnaire)*. Consulting Psychologists Press.
- Stefanics G, Heinzle J, Horváth AA, Stephan KE. 2018. Visual mismatch and predictive coding: a computational Single-Trial ERP study. *The Journal of Neuroscience* 38:4020–4030.
- Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. 2009. Bayesian model selection for group studies. *NeuroImage* 46:1004–1017.
- Sutton RS, Barto AG. 1998. *Introduction to Reinforcement Learning*. MIT press.
- Tan H, Jenkinson N, Brown P. 2014. Dynamic neural correlates of motor error monitoring and adaptation during trial-to-trial learning. *Journal of Neuroscience* 34:5678–5688.
- Tan H, Wade C, Brown P. 2016. Post-Movement beta activity in sensorimotor cortex indexes confidence in the estimations from internal models. *The Journal of Neuroscience* 36:1516–1528.
- Tinkhauser G, Pogosyan A, Tan H, Herz DM, Kühn AA, Brown P. 2017. Beta burst dynamics in Parkinson's disease OFF and ON dopaminergic medication. *Brain* 140:2968–2981.
- Todorov E, Jordan MI. 2002. Optimal feedback control as a theory of motor coordination. *Nature Neuroscience* 5:1226–1235.

Torrecillos F, Tinkhauser G, Fischer P, Green AL, Aziz TZ, Foltynie T, Limousin P, Zrinzo L, Ashkan K, Brown P, Tan H. 2018. Modulation of beta bursts in the subthalamic nucleus predicts motor performance. *The Journal of Neuroscience* 38:8905–8917.

van Beers RJ, Haggard P, Wolpert DM. 2004. The role of execution noise in movement variability. *Journal of Neurophysiology* 91:1050–1063.

Vine SJ, Freeman P, Moore LJ, Chandra-Ramanan R, Wilson MR. 2013. Evaluating stress as a challenge is associated with superior attentional control and motor skill performance: testing the predictions of the biopsychosocial model of challenge and threat. *Journal of Experimental Psychology: Applied* 19:185–194.

Weber LA, Diaconescu AO, Mathys C, Schmidt A, Komater M, Vollenweider F, Stephan KE. 2019. Ketamine affects prediction errors about statistical regularities: a computational Single-Trial analysis of the mismatch negativity. *bioRxiv*.

Wolpert DM, Diedrichsen J, Flanagan JR. 2011. Principles of sensorimotor learning. *Nature Reviews Neuroscience* 12:739–751.

Woolley SC, Rajan R, Joshua M, Doupe AJ. 2014. Emergence of Context-Dependent variability across a basal ganglia network. *Neuron* 82:208–223.

Wu HG, Miyamoto YR, Gonzalez Castro LN, Ölveczky BP, Smith MA. 2014. Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nature Neuroscience* 17:312.

Pulcu E, Browning M. 2019. The misestimation of uncertainty in affective disorders. *Trends in Cognitive Sciences* 23:865–875.

**Decision letter** In the interests of transparency, eLife publishes the most substantive revision requests and the accompanying author responses.

**Acceptance summary:**

In this article, the authors manipulate state anxiety and examine the relationship between anxiety and motor learning. Using electrophysiology and modeling approaches, they show that anxiety constrains flexible behavioral updating.

**Decision letter after peer review:**

Thank you for submitting your article "Alterations in the amplitude and burst rate of beta oscillations impair reward-dependent motor learning in anxiety" for consideration by eLife. Your article has been reviewed by three peer reviewers, including Nicole C Swann as the Reviewing Editor and Reviewer #3, and the evaluation has been overseen by a Reviewing Editor and Laura Colgin as the Senior Editor. The following individuals involved in review of your submission have agreed to reveal their identity: Preeya Khanna (Reviewer #1).

The reviewers have discussed the reviews with one another and the Reviewing Editor has drafted this decision to help you prepare a revised submission.

**Summary:**

This article addresses the relationship between anxiety and motor learning. Specifically, the authors show that anxiety during a baseline exploration phase caused subsequent impairments in motor learning. They go on to use a Bayesian modeling approach to show that this impairment was due to biased estimates of volatility and performance goal estimates. Finally, they couple their behavioral analyses to electrophysiology recordings with a particular focus on sensorimotor (and to a lesser extent prefrontal) beta. They show that post-movement beta rebound is elevated in the anxiety condition. The authors also utilized a novel "beta bursting approach", which in some ways recapitulated the beta power findings, but using a contemporary and exciting method which likely more accurately captures brain activity. Using this approach they show power difference may be driven by increases in burst duration in the anxiety condition – which parallels recent findings in Parkinson's disease populations.

Overall, the reviewers were impressed with many aspects of this manuscript. We appreciated the multi-modal approach (incorporating heart rate measures, clinical rating scales, modeling, and electrophysiology). We also found the behavioral results related to anxiety and motor learning particularly interesting given that they contribute to the existing literature on reward-based learning and volatility, but extends these findings to the motor domain. We also appreciated that the author's actually manipulated state anxiety (rather than relying on individual differences) since this approach allows stronger inferences to be made about causality. Finally, the reviewers noted that the extension of sensorimotor beta outside the motor domain is a novel contribution.

While we were overall enthusiastic, the reviewers did note difficulty in reading the paper. Although the writing was generally clear, the rationale and flow of the presentation of results, particularly for the modeling and EEG findings, were often difficult to understand. For instance, we felt that overall the presentation of the EEG results did not follow a logical flow, and it was sometimes difficult to understand why certain analytic methods were chosen. We also noted that the link between the different modalities could be made more clearly and that additional controls could be added to rule out a motoric contribution to the beta effects. Finally, additional information is needed for the model results. We elaborate on each of these points further below.

Essential revisions:

1) We suggest the authors carefully consider the nomenclature of the conditions and how each relates to motor learning.

For instance, referring to the "exploration" phase as "baseline" caused some confusion since "baseline" typically implies some "pre-manipulation" phase of task.

Related to above, further consideration of how the conditions map onto motor learning would be helpful. In this study, subjects were already instructed to explore task-related dimensions during the baseline period, but were not given feedback during this period. It is unclear how this maps to typical motor "exploration" in the reinforcement learning sense since there is no reinforcement during this period. Additionally, it isn't just a passive baseline measurement since subjects are actively doing something. Further interpretation of how this exploration/baseline phase maps onto other motor learning paradigms, either in the Introduction or Discussion section, would be helpful.

2) Similarly, the use of the terms "learning" and "training" for the second phase of the experiment caused us some confusion. A consistent terminology would have made the manuscript easier to follow.

3) Overall, a strength of the study is the use of many different modalities; however, at present, findings from these modalities are often not linked together. It would be helpful to tie the disparate methods together if some analyses were done to link the different measures. For instance, additional plots like those in Figure 3C-D could be included which correlate different measures to one another across participants. (For example, (a) correlating the model predictions (i.e. belief of environment volatility) and higher variability in cvKI on a subject-to-subject basis to help link the more abstract model parameters to behavioral findings and (b) correlating post-feedback beta power with both volatility estimates and cvKI variability.)

4) In general, the figures could benefit from more labeling and clarification. Some specific examples are mentioned below, but in general, it was not always clear which electrodes data were from, what time periods were shown, which groups, etc.

5) Please include model fits with the results (i.e. how well do they estimate subjects' behavior on a trial-by-trial basis and are there any systemic differences in the model fits across groups?).

6) Please provide a summary figure showing what data is included in the model and perhaps a schematic that illustrates what the model variables are and example trajectories that the model generates.

7) It would be helpful to provide examples to give some intuition about what types of behavior would drive a change in "volatility". For example, can more information be provided to help the reader understand if the results (presented in Figure 10 for instance) enable predictions about subjects' behavior? If beta is high on one trial during the feedback period, does that mean that the model makes a small change in the volatility estimate? How does this influence what the participants are likely to do on the next trial?

8) Generally, the EEG analysis opens up a massive search space (all electrodes, several seconds of data, block-wise analyses, trial-wise analyses, sample-wise analyses, power quantifications, burst-quantifications, long bursts, short bursts, etc.), and the presentation of the findings often jump around frequently between power quantification, burst-quantifications, block-wise, and trial-wise analysis etc. It would be much easier to follow if a few measurements were focused on that were a priori justified. These could be clearly laid out in the introduction with some explanation as to why they were investigated and what each measure might tell the reader. Then, if additional analyses were conducted, these should be explained as post hoc with appropriate justifications and statistical corrections.

9) The EEG results could be better connected to the other findings: for instance, by correlating beta results to model volatility estimates or cvKI variability, as described above.

10) The reviewers felt that an important contribution of this paper was the potential non-motor findings related to sensorimotor beta. However, because there were also motoric differences between conditions, it seems very important to verify whether the beta differences were driven by motoric differences or anxiety-related manipulations. We appreciate the analyses in Figure 8—figure supplement 1 to try to rule out the motoric contribution to the sensorimotor beta differences, but note that this only controlled for certain kinds of movement variability. We would like to see controls for other possible differences in movement between the conditions, for instance differences in movement length, or movement length variability. Finally, is there a way to verify if the participants moved at all after they performed the task?

11) We would like to see what the between group differences for beta power and beta bursts look like during the rest period before the baseline? (For instance, if Figure 7 were generated for rest data?)

[Editors' note: further revisions were suggested prior to acceptance, as described below.]

Thank you for re-submitting your article "Alterations in the amplitude and burst rate of beta oscillations impair reward-dependent motor learning in anxiety" for consideration by *eLife*. Your article has been reviewed by three peer reviewers including Nicole C Swann as the Reviewing Editor and Reviewer #3, and the evaluation has been overseen by a Reviewing Editor and Laura Colgin as the Senior Editor. The following individuals involved in review of your submission have agreed to reveal their identity: Preeya Khanna (Reviewer #1); Jan R Wessel (Reviewer #2).

The reviewers have discussed the reviews with one another and the Reviewing Editor has drafted this decision to help you prepare a revised submission.

Summary:

In general, the authors did a good job addressing our comments. We were especially happy with clarified EEG analysis and, in particular, the care the authors took to avoid potential motoric drivers of beta differences. Given that the authors made significant changes to the manuscript in response to our previous comments, we send the paper out for re-review, and identified a few remaining items in need of clarification. The majority of these are related to the updated model, but we also had a few questions about the EEG analysis, code sharing, and minor points (typos, etc.) We elaborate on these below.

Essential revisions:

Related to the updated model: We have summarized some aspects of the modeling that we believe would benefit from additional explanation (to make the manuscript more broadly accessible). We apologize that we did not bring some of these up in the first submission, but these questions arose either due to the use of the new model or because of clarifications in the revision that provided new insight to us about the model.

1) Explanation/interpretation of the Bayesian modeling – Definitions:

Thank you for Figure 5 – this added clarity to the modeling work but we still are having trouble understanding the general structure of the model. It would be helpful to clearly define the following quantities that are used in the text (in the Materials and methods section before any equations are listed), and ideally also in a figure of example data.

"input" – does this mean the score for a specific trial  $k$ ? We found this a little misleading since an "input" would usually mean some sort of sensory or perceptual input (as in Mathys 2014), but in this case it actually means feedback score (if we understood correctly). Also, please define  $u^k$  before Equation 3 and Equation 4, and ideally somewhere in Figure 5;

- Please clarify how the precision of the input is measured? (Used in Equation 3).

"predicted reward" – from the Mathys et al., 2014 paper we gathered that this is the mean of  $x_1$  obtained on the previous trial? Is this correct? If so, please clarify/emphasize. To increase broad accessibility of the manuscript, it would be helpful to summarize in words somewhere what the model is doing to make predictions. For example, in the paragraph after Equation 2, we weren't sure what the difference is between prediction of  $x_1$  and expectation of  $x_1$ . Typically this terminology would correspond to: prediction =  $E(x_1 \text{ on trial } k \mid \text{information up to trial } k-1)$ , and expectation =  $E(x_1 \text{ on trial } k \mid \text{information up to trial } k)$  but it wasn't clear to us.

"variance vs. precision vs. uncertainty" – These are well defined words but it would also help immensely to only use "variance" or "precision" or "uncertainty" in the explanation/equations. Mentally jumping back and forth gets confusing.

- "belief vs. expectation" – Are these the same? What is the mathematical definition (is it Equations 3 and 7)?

- "pwPE" – please list the equation for this somewhere in the methods, ideally before use of the epsilons in the response variable models.

2) Inputs/outputs of the model:

Inputs – we gather that the input to the model is the score that the participant receives. Then  $x_1$  gets updated according to Equation 3. So  $x_1$  is tracking the expected reward on this trial assuming that the reward on the previous trial must be updated by a prediction error from the current trial? Is this reasonable assumption for this task? What if the participants are exploring new strategies trial to trial? Why would they assume that the reward on the next trial is the same as the current trial (i.e. why is the predicted reward =  $u_{ik-1}$ )? Or is this the point (i.e. if trial to trial the subjects change their strategy a lot that this will end up being reflected as a higher "volatility")? It would be helpful to outline how the model reflects different regimes of behavior (i.e. what does more exploratory behavior look like vs. what is learning expected to look like).

Outputs – response models; please clarify why  $cvKI_{\text{trial}}$  and  $\log(mKI)$  are the chosen responses since these are not variables that are directly responsible for the reward? We thought that the objective of this response modeling was to determine how a large prediction error on the previous trial would influence action on the next trial? Perhaps an output metric could be [similarity between trial  $k$ , trial  $k-1$ ] =  $B_0 + B_1(u_{ik-1}) + B_2(pwPE_{k-1})$ ? So, depending on the reward and previous prediction error, you get a prediction of how similar the next trials' response is to the current trials' response? Right now, we don't understand what is learned from seeing that  $cvKI_{\text{trial}}$  is higher with higher reward expectation (this is almost by necessity right? because the rewarded pattern needs high  $cvKI$ ) or higher prediction error.

3) Interpretation of the model:

Is the message from Figure 6A that the expected reward is lower for  $anx1$  than  $anx2$  and control? Since the model is trying to predict scores from actual score data, isn't this result expected given Figure 4A. Can the authors please clarify this?

We noticed that  $\log(\mu_2)$  is lower for  $anx1$  and  $anx2$  than control. Should this correspond to a shallower slope (or a plateau in score that is reached more quickly) in Figure 4A over learning for  $anx1$ ? If so, why don't we see that for  $anx2$ ? If this is true, and given that  $cvKI$  is no different for  $anx1$ ,  $anx2$ , and control, wouldn't that mean that the reward rate is plateauing faster for  $anx1$  and  $anx2$  while they are still producing actions that are equally variable to control? So, are participants somehow producing actions that are variable yet getting the same reward – so they're getting "stuck" earlier on in the learning process? Can the authors provide some insight into what type of behavior trends to expect given the finding of Figure 6B-C? Right now all the reader gets as far as interpretation goes is that the  $anx1$  group underestimates "environmental volatility" and that the mean behavior and  $cvKI$  is the same across all groups.

Does underestimating volatility mean that subjects just keep repeating the same sequence over and over? If so, can that be shown? Or does it mean that they keep trying new sequences but fail to properly figure out what drives a higher reward? Since the model is fit on the behavior of the participants, it should be possible to explain more clearly what drives the different model fits.

Related EEG Analysis: We greatly appreciated the clarified EEG analysis. Re-reading this section, we were able to understand what was done much better, but had two queries related to the analysis.

1) We noted that the beta envelope in Figure 7A looks unusual. It looks almost like the absolute value of the beta – filtered signal rather than the envelope, which is typically smoother and does not follow peaks and troughs of the oscillation. Can the authors please clarify how this was calculated?

2) In subsection "Analysis of power spectral density", the authors write: "The time-varying spectral power was computed as the squared norm of the complex wavelet transform, after averaging across trials within the beta range." This sounds like the authors may have calculated power after averaging across trials? Is this correct (i.e. was the signal averaged before the wavelet transform, such that trial to trial phase differences may cancel out power changes?), or do the authors mean that they averaged across trials after extracting beta power for each trial? If the former the author should emphasize that this is what they did, since it is unconventional.

3) To try to understand point 2 above, we checked if the authors had shared their code, and found that, although data was shared, code was not, as far as we could tell. *eLife* does require code

sharing as part of their policies (<https://reviewer.elifesciences.org/author-guide/journal-policies>) so please include that.

#### Author response Essential revisions:

1) We suggest the authors carefully consider the nomenclature of the conditions and how each relates to motor learning.

For instance, referring to the "exploration" phase as "baseline" caused some confusion since "baseline" typically implies some "pre-manipulation" phase of task.

Related to above, further consideration of how the conditions map onto motor learning would be helpful. In this study, subjects were already instructed to explore task-related dimensions during the baseline period, but were not given feedback during this period. It is unclear how this maps to typical motor "exploration" in the reinforcement learning sense since there is no reinforcement during this period. Additionally, it isn't just a passive baseline measurement since subjects are actively doing something.

Agreed; the first experimental phase (termed baseline before) has been relabeled as "initial exploration" or, in some instances, "exploration" phase.

We prefer the term "initial exploration" as it should be understood as the first experimental phase (block 1). This does not imply that participants did not to use some degree of exploration in learning phase. The learning phase was indeed expected to require some degree of exploration during the first trials, followed by exploitation of the inferred performance goal (see below, and Figure 4—figure supplement 1). This transition from exploration to exploitation during the learning blocks directly relates to earlier investigations of reinforcement learning (see below).

In the revised manuscript, we have clarified why we used the initial motor exploration phase: "The rationale for including a motor exploration phase in which participants did not receive trial-based feedback or reinforcement was based on the findings that initial motor variability (in the absence of reinforcement) can influence the rate at which participants learn in a subsequent motor task (Wu et al., 2014)."

The findings of Wu et al., 2014 are significant in demonstrating that initial motor variability measured when participants perform ballistic arm movements in the absence of reinforcement or visual feedback can predict the rate of reward-based learning in a subsequent phase.

Similarly, in our study the initial motor exploration phase aimed to assess an individual's use of motor variability in the absence of feedback and when there was no hidden goal to infer. Motor variability here would be driven by internal motivation (and/or motor noise) and would not be guided by explicit external reward.

The fundamental question for us was to determine whether larger task-related variability during block 1 would improve subsequent reward-based learning, even if during the learning blocks a successful performance required participants to exploit the inferred goal. We have created Figure 4—figure supplement 1, which illustrates the result of progressive reduction in temporal variability in the learning blocks (increased exploitation) as participants approached and aimed to maintain the solution. This drop in temporal variability is one of the hallmarks of learning (Wolpert et al., 2010).

Based on our results, we suggest that initial exploration may facilitate learning of the mapping between the actions and their sensory consequences (even without external feedback)", which had a positive influence on subsequent learning "from performance-related feedback".

Further interpretation of how this exploration/baseline phase maps onto other motor learning paradigms, either in the Introduction or Discussion section, would be helpful.

Thanks. By assessing motor variability during an initial exploration period before a reward-based learning period, Wu et al., (2014) positively correlated initial variability with learning curve steepness during training – a relationship previously observed in the zebra finch (Kao et al. 2005, Olveczky et al., 2005, 2011). This suggests that higher levels of motor variability do not solely amount to increased noise in the system. Instead, this variability represents a broader action space that can be capitalised upon during subsequent reinforcement learning by searching through previously explored actions (Herzfeld and Shadmehr, 2014). However, two recent studies using visuomotor adaptation paradigms could not find a similar correlation between motor variability and the rate of motor adaptation (He et al., 2016, Singh et al., 2016). Aiming to align this discrepancy in results, Dhawale et al., (2017) identified that in contrast to Wu et al., (2014), the aforementioned studies gave task-relevant feedback during baseline, which in turn updates the internal model of the action, accentuating execution noise over planning noise. They hypothesise that variability driven by

planning noise underlies learning-related motor exploration (Dhawale et al., 2017). In this study, we aimed to investigate the effect of state anxiety on initial variability prior to a reward-based learning period.

We had summarised those arguments in the previous Discussion. Now, we have also added:

Discussion section: "Another consideration is that our use of an initial exploration phase that did not provide reinforcement or feedback signals was motivated by the work of Wu and colleagues (2014), which demonstrated a correlation between initial variability (no feedback) and learning curve steepness in a subsequent reward-based learning phase— a relationship previously observed in the zebra finch (Kao et al., 2005; Olveczky et al., 2005, 2011). This suggests that higher levels of motor variability do not solely amount to increased noise in the system. Instead, this variability represents a broader action space that can be capitalised upon during subsequent reinforcement learning by searching through previously explored actions (Herzfeld and Shadmehr, 2014). Accordingly, an implication of our results is that state anxiety could impair the potential benefits of an initial exploratory phase for subsequent learning."

2) Similarly, the use of the terms "learning" and "training" for the second phase of the experiment caused us some confusion. A consistent terminology would have made the manuscript easier to follow.

Agreed, we have settled for "learning". The term "training" was used in analogy to Wu et al., (2014) – learning is more appropriate.

3) Overall, a strength of the study is the use of many different modalities; however, at present, findings from these modalities are often not linked together. It would be helpful to tie the disparate methods together if some analyses were done to link the different measures. For instance, additional plots like those in Figure 3C-D could be included which correlate different measures to one another across participants. (For example, (a) correlating the model predictions (i.e. belief of environment volatility) and higher variability in cvIKI on a subject-to-subject basis to help link the more abstract model parameters to behavioral findings and (b) correlating post-feedback beta power with both volatility estimates and cvIKI variability.)

Agreed.

a) The new family of response models used allowed us to obtain the best model that links trial-by-trial behavioural responses and HGF quantities. Details are provided below in our reply to Q7.

In brief, the winning response model explains the variability of temporal intervals within the trial ( $cvIKI_{\text{trial}}$ ) as a linear function of the reward estimates,  $\mu_1$ , and the precision-weighted PE about reward,  $\varepsilon_1$ . This model outperformed other alternative response models that used  $\mu_2$ ,  $\varepsilon_2$  and different combinations of  $\mu_1$ ,  $\mu_2$ ,  $\varepsilon_1$ ,  $\varepsilon_2$ , as well as a different response measure (logarithm of the mean IKI).

Thus, an increase in the estimated reward  $\mu_1$  and an enhanced pwPE  $\varepsilon_1$  that drives belief updating about reward would contribute to a larger degree of temporal variability (less isochronous performance) on the current trial. This result is intuitively meaningful as the score was directly related to the norm of the difference IKI values across successive keystrokes and the hidden goal actually required a relatively large difference between successive IKI values, which would also be associated with larger  $cvIKI_{\text{trial}}$  values. Thus, the winning response model captured how the inferred environmental states ( $\mu_1$  and  $\varepsilon_1$ ) mapped to the observed responses ( $cvIKI_{\text{trial}}$ ) on a trial-by-trial basis.

Note that the trial-wise measure  $cvIKI_{\text{trial}}$  is different from the standard measure of motor variability across trials we used in the manuscript,  $cvIKI$ .

New Figure 6—figure supplement 1: Across all our participants, the measure of changes in across-trial temporal variability ( $cvIKI$ : difference from learning block1 to block2) was positively associated with the changes in volatility estimates ( $\mu_2$ : difference between learning block2 and block1). This was revealed in a non-parametric Spearman correlation ( $\rho = 0.398$ ,  $p = 0.002$ ), supporting that participants who performed more different timing patterns across trials in block2 relative to block1 also increased their volatility estimate in block2 as compared to block1. Conversely, participants who showed a tendency to exploit the rewarded performance decreased their estimate of volatility.

b) Correlations between post-feedback beta power and HGF estimates:

Because in the predictive coding framework the quantities that are thought to dominate the EEG signal are the pwPEs (Friston and Kiebel, 2009), we had assessed the relation between belief updates (regulated by pwPEs on level 1 and 2) and the post-feedback beta activity. The revised

manuscript also follows this approach, but we have improved the analysis by assessing simultaneously the effect of  $e_1$  and  $e_2$  on the beta power activity running a multiple linear regression in all participants. The results indicate that both  $e_1$  and  $e_2$  have a significant negative effect on the beta activity (power and rate of long bursts) across participants. Furthermore, the analysis demonstrates that using  $e_2$  as second predictor in the multiple regression analysis adds significant predictive power to using simply  $e_1$  as a predictor.

We did not expect beta activity to facilitate the “encoding” of volatility estimates directly, but only precision-weighted PEs about volatility. Accordingly, our results linking post-feedback beta activity to pwPE about reward and volatility provide a mechanism through which beliefs about volatility (and reward) are updated.

For the reviewers, we have also assessed the correlation between the mean post-feedback beta activity (power) and the degree of motor variability across trials during the learning blocks, cvIKI, and we found no significant association (Spearman  $\rho < 0.08$ ,  $P = 0.56$ ). This suggests that post-feedback beta activity is not associated on a trial-by-trial basis with the overall degree of motor variability, but rather with the step of the updates in beliefs ( $e_1$ ,  $e_2$ ).

By contrast, during the initial exploration phase, there was a significant non-parametric correlation between the averaged beta activity after the STOP signal and the degree of motor variability across trials (Spearman  $\rho < -0.4397$ ,  $P = 0.0001$ ). This result links increased use of motor variability during exploration with a reduction in beta power following trial performance. See new Figure 8—figure supplement 6.

4) In general, the figures could benefit from more labeling and clarification. Some specific examples are mentioned below, but in general, it was not always clear which electrodes data were from, what time periods were shown, which groups, etc.

Agreed. We have made the labeling of analyses and figures more explicit.

In the large figures with subplots, e.g. Figure 8, and former Figure 8—figure supplements 1-5 we had used one topographic sketch to illustrate the electrodes of the effect across all measures, although the sketch was used in only one of the subplots, in the one with more empty space to allow for the inset. We have kept this system for the figure, but we now added a clarification in the figure caption.

5) Please include model fits with the results (i.e. how well do they estimate subjects’ behavior on a trial-by-trial basis and are there any systemic differences in the model fits across groups?).

Agreed. In the revised manuscript we provide as Figure 5—figure supplement 3 the grand-average of the trial-by-trial residuals in each group. The residuals represent the trial-by-trial difference between the observed responses ( $y$ ) and those predicted by the model (predResp):  $res = y - predResp$ .

In the winning response model (see below for new response models tested), the relevant response variable that was identified was  $cvIKI_{trial}$  (cv of IKI values across keystroke positions in a trial).

We also summarise here the results from Figure 5—figure supplement 3 by computing in each group the mean residual values across trials:

cont: 0.0001 (0.0002)

anx1: 0.0001 (0.0001)

anx2: 0.0002 (0.0001)

In the second control experiment we obtained the following mean residual values per group:

cont: 0.0008 ( $10^{-6}$ )

anx3: 0.0001 (0.0008)

There were thus no systematic differences in the model fits across groups and the low mean residual values further indicate that the model captured the fluctuations in data well.

6) Please provide a summary figure showing what data is included in the model and perhaps a schematic that illustrates what the model variables are and example trajectories that the model generates.

Thanks for the suggestion. We have added a schematic in Figure 5 illustrating the model’s hierarchical structure and the belief trajectories.

In addition, we have provided the detailed update equations for belief and precision estimates in the two-level HGF perceptual model (equations 3-10). This will improve the understanding of how relevant model output variables evolve in time. Moreover, in the revised manuscript we have used

more complete response models, using as reference the work by Marshall et al., (2016), that allow us to address the next question raised by the reviewers (Q7, see below). How the response model parameters influence the input to the two-level perceptual model is also reflected in the equations and the schematic in Figure 5. Details on the new response models are provided in Q7.

In Figure 5, we indicate how model parameters  $\omega_1$  and  $\omega_2$  influence the estimates at each level. Parameter  $\omega_1$  represents the strength of the coupling between the first and second level, whereas  $\omega_2$  modulates how precise participants consider their prediction on that level (larger  $\omega_2$  or smaller  $\omega_1$ ). Thus,  $\omega_1$  and  $\omega_2$  additionally characterise the individual learning style (Weber et al., 2019).

The new Figure 5—figure supplement 1 illustrates using simulated data how different values of  $\omega_1$  or  $\omega_2$  affect the changes in belief trajectories across trials, for an identical series of input scores. In Figure 5—figure supplement 1A we can observe how smaller values of  $\omega_1$  attenuate the general level of volatility changes (less pronounced updates or reduction). By contrast, in panel Figure 5—figure supplement 1C, we note that  $\omega_2$  regulates the scale of phasic changes on a trial-by-trial basis, with larger  $\omega_2$  values inducing more sharp or phasic changes to prediction violations in the level below (changes in PE at level 1).

In terms of the analysis of the computational quantities, we have now added a between-group comparison in  $\omega_1$  and  $\omega_2$ . The results highlight that “In addition to the above-mentioned group effects on relevant belief and uncertainty trajectories, we found significant differences between anx1 and control participants in the perceptual parameter  $\omega_1$  (mean and SEM values for  $\omega_1$ : -4.9 [0.45] in controls, -3.7 [0.57] in anx1,  $P = 0.031$ ) but not in  $\omega_2$ : -2.8 [0.71] in controls, -2.4 [0.76] in anx1 ( $P > 0.05$ ). The smaller values of  $\omega_1$  in anx1 correspond with an attenuation of the updates in volatility (less pronounced updates or reduction). The perceptual model parameters in anx2 did not significantly differ from those in control participants either ( $P > 0.05$ ; mean and SEM values for  $\omega_1$  and  $\omega_2$  in anx2 were -5.4 [0.81] and -1.8 [0.74]).”

In the second, control experiment, the group-average values of  $\omega_1$  and  $\omega_2$  were: -4.1 (SEM 0.53) and -3.3 (0.29) for controls; -4.4 (0.38) and -3.6 (0.32) in anx3. There were no significant differences between groups in these values,  $P > 0.05$ .

7) It would be helpful to provide examples to give some intuition about what types of behavior would drive a change in “volatility”. For example, can more information be provided to help the reader understand if the results (presented in Figure 10 for instance) enable predictions about subjects’ behavior? If beta is high on one trial during the feedback period, does that mean that the model makes a small change in the volatility estimate? How does this influence what the participants are likely to do on the next trial?

Thanks for this question, which has motivated us to make a substantial improvement in the response models we use in the HGF analysis. We provide a detailed explanation below, but the summary can be stated here:

Yes, a higher value of beta power or burst rate during feedback processing is associated with a smaller update in the volatility estimate (smaller  $\Delta w_{PE}$  on level 2,  $\epsilon_2$ ) in that trial. But also, with a smaller update in the belief about reward ( $\epsilon_1$ ).

Regarding  $\epsilon_2$ , if a participant had a biased estimate of volatility (underestimation or overestimation), a drop in beta activity during feedback processing would promote a larger update in volatility (through  $\epsilon_2$ ) to improve this biased belief. Similarly, a reduction in beta activity would also increase updates in reward estimates (through  $\epsilon_1$ ), which in the winning response model is linked to the performance measure, and thus increases  $c_{\text{trial}}$ .

Following the anxiety manipulation in our study we find a combination of biased beliefs about volatility and reward and increased feedback-locked beta activity, which would be associated with reduced values of  $\epsilon_2$  and  $\epsilon_1$ . Accordingly, biased beliefs are not updated appropriately in state anxiety.

In the revised manuscript, we provide a more complete description of the two-level HGF for the perceptual and response models. The perceptual model describes how a participant maps environmental causes to sensory inputs (the scores), whereas the response model maps those inferred environmental causes to the performance output the participant generates every trial.

In the following, we provide detailed explanations on these aspects: (A) how phasic volatility is estimated in the perceptual model, and (B) how changes in volatility may influence changes in behaviour. Ultimately, we address (C) how beta power and burst rate can drive the updates in volatility estimates.

A) Concerning the perceptual model, we have included the update equations for beliefs and precision (inverse variance) estimates at each level. This helps clarify what contributes to changes in the estimation of environmental volatility. An additional illustration is provided in the new HGF model schematic (Figure 5).

Estimates about volatility in trial  $k$  are updated proportionally to the environmental uncertainty, the precision of the prediction of the level below,  $\pi_1$ , and the prediction error in the level below,  $\delta_1$ ; volatility estimates are also inversely proportional to the precision of the current level,  $\pi_2$ :

$$\mu_2^k = \hat{\mu}_2^k + \frac{1}{2} \frac{1}{\pi_2^k} w_1^k \delta_1^k,$$

With

$$w_1^k = \exp(\mu_2^{k-1} + \omega_1) \pi_1^k$$

We have dropped parameter  $k$  and the time step  $t$  from these expressions (see Mathys et al., 2011, 2014), as they take value = 1.

The expression  $\exp(\mu_2^{k-1} + \omega_1)$  is often termed environmental uncertainty, and is defined as the exponential of the volatility estimate on the previous trial (before seeing the feedback) and the coupling parameter  $\omega_1$ , also termed tonic volatility (Mathys et al., 2011, 2014).

The equations above illustrate the general property of the HGF perceptual model that belief updates depend on the prediction error (PE) of the level below, weighted by a ratio of precisions.

Thus, a larger PE about reward,  $\delta_1$ , will increase the step of the update in volatility – participants render the environment to be more unstable. However, the PE contribution is weighted with the precision ratio: when an agent places more confidence on the estimates of the current level (larger precision  $\pi_2$ ), the update step for volatility will be reduced. On the other side, a larger precision of the prediction at the level below ( $\pi_1$ ) will increase the update in volatility. If the prediction about reward is more precise, then the PE about reward will be used to a larger degree (through the product  $\pi_1 \delta_1$ ).

Therefore, in addition to constant contributions from the tonic volatility  $\omega_1$  to the update, the main quantity that drives the updates in volatility is the ratio of precision between lower and current level, thereby affecting how much the PE about reward contributes to the belief updating in volatility.

B) The revised manuscript tested several new more complete response models using as reference the work by Marshall et al. (2016). In that work, the authors described in a different paradigm how the participant's perceptual beliefs map onto their observed  $\log(\text{RT})$  responses on a trial-by-trial basis, with the responses  $\log(\text{RT})$  being a linear function of PEs, volatility, precision-weighted PEs, and other terms (multiple regression). For that purpose, they created the family of scripts `tapas_logrt_linear_whatworld` in the `tapas` software.

We have now implemented similar models, but adapted to our task (scripts `tapas_IKI_linear_gaussian_obs` uploaded to the Open Science Framework data repository). The response models we tested aimed to explain a relevant trialwise performance parameter as a linear function of HGF quantities (multiple regression). The alternative models used two different performance parameters:

- The coefficient of variation of inter-keystroke intervals,  $\text{cvIKI}_{\text{trial}}$ , as a measure of the extent of timing variability *within* the trial.
- The logarithm of the mean performance tempo in a trial,  $\log(\text{mIKI})$ , with IKI in milliseconds.

Furthermore, for each performance measure, the response model was a function of a constant component of the performance measure (intercept) and other quantities, such as: the reward estimate ( $\mu_1$ ), the volatility estimate ( $\mu_2$ ), the precision-weighted PE about reward ( $\epsilon_1$ ), or the precision-weighted PE about volatility ( $\epsilon_2$ ). See details in the revised manuscript. In total we assessed six different response models. Using random effects Bayesian model selection (BMS), we obtained a winning model that explained the performance measure  $\text{cvIKI}_{\text{trial}}$  as a linear function of  $\mu_1$  and  $\epsilon_1$ :

$$\text{cvIKI}_{\text{trial}}^k = \beta_0 + \beta_1 \mu_1^k + \beta_2 \epsilon_1^k + \zeta$$

The  $\beta$  coefficients were positive and significantly different than zero in each participant group ( $P < P_{\text{FDR}}$ , controlled for multiple comparisons arising from 3 group tests), as shown in the new Figure 5—figure supplement 1.

Thus, in addition to the estimated positive constant (intercept) value of  $\text{cvKI}_{\text{trial}}$ , quantities  $\mu_1$  and  $\epsilon_1$  had a positive influence on  $\text{cvKI}_{\text{trial}}$ , such that higher reward estimates and higher pwPEs about reward increased the temporal variability on that trial (less isochronous performance).

The noise parameter  $z$  did not significantly differ between groups ( $P > 0.05$ ), and therefore we found no differences in how the model was able to estimate predicted responses to fit observed responses in each group.

Overall, the BMS results indicate that response models that defined the response parameters as a function of volatility estimates and pwPE on level 2 trial-by-trial basis were less likely to explain the data. However, because  $\mu_2$  drives the step of the Gaussian random walk for the estimation of the true state  $x_1$ , an underestimation in the beliefs about volatility (smaller  $\mu_2$  as found in anxiety groups) would drive smaller updates about  $x_1$ , ultimately leading to smaller  $\text{cvKI}_{\text{trial}}$  – as our winner response model establishes. This can also be observed in Equation 6, where smaller values of the volatility estimate on the previous trial,  $\mu_2^{k-1}$  increase the precision of the prediction about reward ( $\pi^1$ ), leading to smaller updates for  $\mu_1$  (Equation 3).

As reported in the Discussion section, “Volatility estimates impact directly the estimations of beliefs at the lower level, with reduced  $m_2$  leading to a smaller step of the update in reward estimates. Thus, this scenario would provide less opportunity to ameliorate the biases about beliefs in the lower level to improve them.”

The new HGF results are shown in Figure 6, precisely illustrating that *anx1* underestimated  $\mu_2$  relative to control participants – when using the improved winning response model – thus accounting for the smaller  $\text{cvKI}_{\text{trial}}$  found in this group.

C) In a similar fashion to the way we constructed response models in the new HGF analysis, we used a multiple linear regression analysis to evaluate the measure of feedback-locked beta power, and separately, the rate of long bursts as a linear function of two quantities,  $e_1$  and  $e_2$ . This analysis is similar to the one we did in the previous version of the manuscript, but it is an improvement in two respects: It assesses the simultaneous influence of  $e_1$  and  $e_2$  on the measures of beta activity, and it uses trial-wise data in each participant to obtain the individual beta coefficients.

8) Generally, the EEG analysis opens up a massive search space (all electrodes, several seconds of data, block-wise analyses, trial-wise analyses, sample-wise analyses, power quantifications, burst-quantifications, long bursts, short bursts, etc.), and the presentation of the findings often jump around frequently between power quantification, burst-quantifications, block-wise, and trial-wise analysis etc. It would be much easier to follow if a few measurements were focused on that were a priori justified. These could be clearly laid out in the introduction with some explanation as to why they were investigated and what each measure might tell the reader. Then, if additional analyses were conducted, these should be explained as post hoc with appropriate justifications and statistical corrections.

Thanks for this suggestion. We completely agree with the reviewers and have considerably simplified the EEG statistical analyses. In addition, we have more explicitly stated in the revised introduction all our main hypotheses. The detailed aims and measures of the EEG analyses have been included at the beginning of the Results section to provide a clear overview.

Introduction:

Now we explicitly mention that prefrontal electrode regions were one of the regions of interest, together with “sensorimotor” electrode regions. In addition, we cite more work that identifies prefrontal regions as central to the neural circuitry of anxiety.

“Crucially, in addition to assessing sensorimotor brain regions, we focused our analysis on prefrontal areas on the basis of prior work in clinical and subclinical anxiety linking the prefrontal cortex (dmPFC, dlPFC) and the dACC to the maintenance of anxiety states, including worry and threat appraisal (Grube and Nitsche, 2012; Robinson et al. 2019). Thus, beta oscillations across sensorimotor and prefrontal brain regions were evaluated.”

“We accordingly assessed both power and burst distribution of beta oscillations to capture dynamic changes in neural activity induced by anxiety and their link to behavioral effects.”

“EEG signals aimed to assess anxiety-related changes in the power and burst distribution in sensorimotor and prefrontal beta oscillations in relation to changes in behavioral variability and reward-based learning.”

Subsection “Electrophysiological Analysis”:

“The analysis of the EEG signals focused on sensorimotor and anterior (prefrontal) beta oscillations and aimed to separately assess (i) tonic and (ii) phasic (or event-related) changes in spectral power and burst rate. Tonic changes in average beta activity would be an indication of the anxiety manipulation having a general effect on the modulation of underlying beta oscillatory properties. Complementing this analysis, assessing phasic changes in the measures of beta activity during trial performance and following feedback presentation would allow us to investigate the neural processes driving reward-based motor learning and their alteration by anxiety. These analyses focused on a subset of channels across contralateral sensorimotor cortices and anterior regions (See Materials and methods section).”

Below, in the Results section of the exploration phase, when we introduce the methodology to extract bursts, we now state that due to the complementary information provided by duration, rate, and slope of the distribution of bursts, we exclusively focus on the analysis of the slope when assessing tonic burst properties. The slope is already a summary statistic of the properties of the distribution (e.g. smaller slope [absolute value] indicates a long-tailed distribution with more frequent long bursts).

This will hopefully make the Results section more concise, as general average burst properties can be characterised by the slope of their distribution of durations:

Subsection “Electrophysiological Analysis”: “Crucially, because the burst duration, rate, and slope provide complementary information, we focused our statistical analysis of the tonic beta burst properties on the slope or life-time exponent,  $\tau$ . A smaller slope corresponds to a burst distribution biased towards more frequent long bursts.”

The separate analysis of bursts into long and brief bursts was inspired by the previous burst studies in parkinson’s patients showing the presence of long bursts (> 500 ms) in the basal ganglia and linking those to motor symptoms and poorer performance. However, this was indeed a post-hoc analysis in our study, additionally motivated by the clear dissociation between long and brief bursts shown in Figure 7, and determined by the difference in slope between anx1 and controls. This analysis has now been correctly identified as post-hoc analysis:

Subsection “Electrophysiological Analysis”: “As a post-hoc analysis, the time course of the burst rate was assessed separately in beta bursts of shorter (< 300 ms) and longer (> 500 ms) duration,....”

This split analysis is important in our results, as the longer burst properties seem to align better with the power results. While brief bursts are more frequent in all participants (and physiologically relevant), they seem to be here less related to task performance.

Subsection “Electrophysiological Analysis”: “The rate of long oscillation bursts displayed a similar time course and topography to those of the power analysis, with an increased burst rate after movement termination and after the STOP signal ”

9) The EEG results could be better connected to the other findings: for instance, by correlating beta results to model volatility estimates or cvKI variability, as described above.

The measures of feedback-related beta oscillations have now been correlated across participants with the index of across-trials cvKI, reflecting motor variability (Q3b above). Another specific correlation we have computed is that between motor variability, across-trials cvKI, and volatility (Q3a above).

As explained in question Q7, we consider that the Hierarchical Bayesian model – now assessed in combination with an improved family of response models – is able to explain how in individual participants behaviour and beliefs about volatility or reward relate on a trial-by-trial basis.

In addition, now we use a multiple linear regression in individual subjects to explain trialwise power measures as a function of pwPE about volatility and reward (the main measures that are expected to modulate the EEG signal, Friston and Kiebel, 2009). This new analysis thus already is an assessment of trialwise relations between power and relevant computational quantities.

We hope the reviewers agree in that these analyses are sufficient to clarify those relationships (which in the case of the multiple regression analysis is already a type of correlation analysis).

What our analyses do not clarify is the dissociation between beta activity being related to pwPE in level 1 and 2, respectively. It is likely that a combined analysis of beta and gamma oscillations in this context could help identify different neural mechanisms (potentially with a different spatial distribution) separately driving belief updating through  $e_1$  and  $e_2$ . This an investigation that we are currently completing in the context of a different study.

10) The reviewers felt that an important contribution of this paper was the potential non-motor findings related to sensorimotor beta. However, because there were also motoric differences between conditions, it seems very important to verify whether the beta differences were driven by motoric differences or anxiety-related manipulations. We appreciate the analyses in Figure 8—figure supplement 1 to try to rule out the motoric contribution to the sensorimotor beta differences, but note that this only controlled for certain kinds of movement variability. We would like to see controls for other possible differences in movement between the conditions, for instance differences in movement length, or movement length variability.

This is a great suggestion. We have now made additional control analyses similar to the original Figure 8—figure supplement 1 to assess the differences in beta power and burst rate between a subset of control and anxious participants matched in these variables:

- Duration of the trial performance (movement length or total duration in ms) – Figure 8—figure supplement 2
- Variability of movement length (cv of movement length) – Figure 8—figure supplement 3
- Mean use of keystroke velocity in the trial – Figure 8—figure supplement 4

The results indicate that when controlling for changes in each of these motor parameters, anxiety alone could explain the findings of larger post-movement beta-band PSD and rate of longer bursts, while also explaining the reduced rate of brief bursts during performance.

In the original manuscript, we had reported that “General performance parameters, such as the average performance tempo or the mean keystroke velocity did not differ between groups, either during initial baseline exploration or learning”. This outcome also accounts for why the new control analyses support that motor parameters such as the mean performance duration or keystroke velocity are not confounding factors when explaining the beta activity effects in anxiety.

Finally, is there a way to verify if the participants moved at all after they performed the task?

The best way to address this question, in the absence of EMG recordings from e.g. neck or torso muscles, is to look at broadband high-frequency activity (gamma range above 50 Hz), which has been consistently associated in previous studies with muscle artifacts. For instance, in this review paper by Muthukumaraswamy (2013), the author identified 50-160 Hz gamma activity with postural activity of upper neck muscles, generated by participants using a joystick (Figure 2). Changes in beta activity in this task were identified as true brain activity related to neural processing of the task requirements.

The author also reported that EEG activity contaminated by muscle artifacts is typically maximal at the edges of the electrode montage (e.g. temporal electrodes) but can be also observed at central scalp positions.

In our experimental setting, we instruct participants to not move the torso or head during the total duration of the trial, from the warning signal through to the sequence performance until the end of the trial (2 seconds after the feedback presentation). And we always monitor EEG for muscle artifacts while participants familiarise with the apparatus and the sequences at the beginning of the experimental session.

We have performed a control analysis of higher gamma band activity, between 50-100Hz, and display the results in Figure 9—figure supplement 2. This figure excludes the power values at 50Hz and 100Hz related to power line noise (and harmonics).

We have evaluated these conditions in the learning blocks:

A) Gamma power within 0-1 s after feedback presentation, where participants should be at rest after completing the trial performance.

B) Gamma power within 0-1 s locked to a key press, when participants are moving their fingers.

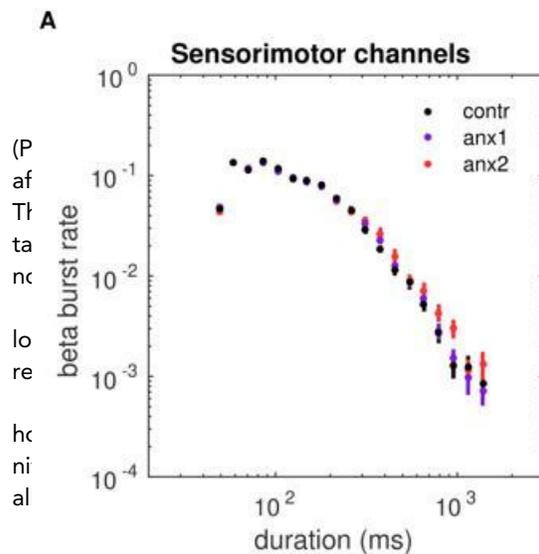
C) Gamma power within 0-1 s locked to the initiation of the trial, when participants are cued to wait for the GO response, and can be expected to be mentally preparing but otherwise at rest.

We then performed the following statistical analyses to test for differences in gamma power:

- Condition A versus Condition C in bilateral temporal electrodes
- Condition A versus Condition C in bilateral and central sensorimotor electrode regions
- Condition B versus Condition C in bilateral temporal electrodes
- Condition B versus Condition C in bilateral and central sensorimotor electrode regions

In addition, focusing now only on the target period of the manuscript, the feedback-locked changes (A), we assessed differences between experimental and control groups:

- Condition A: anx1 versus controls in bilateral temporal electrodes



**B** Essential revisions: [Editors' note: further revisions were suggested prior to acceptance, as described below.]

Related to the updated model: We have summarized some aspects of the modeling that we believe would benefit from additional explanation (to make the manuscript more broadly accessible). We apologize that we did not bring some of these up in the first submission, but these questions arose either due to the use of the new model or because of clarifications in the revision that provided new insight to us about the model.

1) Explanation/interpretation of the Bayesian modeling – Definitions:

Thank you for Figure 5 – this added clarity to the modeling work but we still are having trouble understanding the general structure of the model. It would be helpful to clearly define the following quantities that are used in the text (in the Materials and methods section before any equations are listed), and ideally also in a figure of example data.

"input" – does this mean the score for a specific trial  $k$ ? We found this a little misleading since an "input" would usually mean some sort of sensory or perceptual input (as in Mathys 2014), but in this case it actually means feedback score (if we understood correctly).

Agreed. The term "input" – as used in Mathys et al., (2014) – is now specified in the introduction to the HGF model in the Results section and the Materials and methods section. In subsection "Bayesian learning modeling reveals the effects of state anxiety on reward-based motor learning", we also give two examples of "sensory input" being replaced by a series of outcomes:

"In some implementations of the HGF, the series of sensory inputs are replaced by a sequence of outcomes, such as reward value in a binary lottery (Mathys et al., 2014; Diaconescu et al., 2017) or electric shock delivery in a one-armed bandit task (De Berker et al., 2016). In these cases, similarly to the case of sensory input, an agent can learn the causes of the observed outcomes and thus the likelihood that a particular event will occur. In our study, the trial-by-trial input observed by the participants was the series of feedback scores (hereafter input refers to feedback scores)."

In the case of the binary lottery or a one-armed bandit task, participants select one of two images and observe the corresponding outcome, which can be reward (0,1), or some other type of outcome, such as pain shocks (binary 0-1; de Berker et al., 2016). Thus, although the perceptual HGF is described in terms of "sensory" input being observed by an agent, in practice several studies use the series of feedback values or outcomes associated with the responses as input. This is also what we did in our implementation of the HGF: the input observed by participants, labeled  $u^k$  in the equations, is the feedback score associated with the response in that trial. Here, the HGF models how an agent infers the estate of the environment, which is the reward for trial  $k$ ,  $m_1^k$  (true state:  $x_1^k$ ), using the observed outcomes (observed feedback score each trial,  $u^k$ ). We have included De Berker et al., 2016, as a new reference in the manuscript.

The HGF was originally developed by C. Mathys as a perceptual model, to measure how an agent generates beliefs about environmental states. Based on those inferred beliefs, the HGF can be subsequently linked to participant's responses using a response model. This is the procedure we

d central sensorimotor electrode regions  
mporal electrodes  
d central sensorimotor electrode regions  
gamma activity in any of the assessed contrasts  
permutation test between two conditions/groups  
electrodes and the frequency range 52-98Hz).  
rted in the manuscript are confounded by simul-  
taminating the EEG signal (or by differences in  
up differences for beta power and beta bursts  
s? (For instance, if Figure 7 were generated for  
t of the reviewing process. The figure illustrates  
xperimental task there are no apparent (nor sig-  
n experimental and control groups (assessed in  
motor electrodes).

[Editors' note: further revisions were sug-

followed in our study: the response model explains participants' responses as a function of the inferred beliefs or related computational quantities (e.g. PEs). See below please for the implementation of new – more interesting – response models suggested by the reviewers.

Also, please define  $u^k$  before Equation 3 and Equation 4, and ideally somewhere in Figure 5;

We have included in Figure 5 the definition of input  $u^k$ , which is the observed feedback score for the trial (normalized to range 0-1). The definition is also presented at the beginning of the subsection "Computational Model":

"In many implementations of the HGF, the sensory input is replaced with a series of outcomes (e.g. feedback, reward) associated with participants' responses (De Berker et al., 2016; Diaconescu et al., 2017)."

"The HGF corresponds to the perceptual model, representing a hierarchical belief updating process, i.e., a process that infers hierarchically related environmental states that give rise to sensory inputs (Stefanics, 2011; Mathys et al., 2014). In the version for continuous inputs we implemented (see Mathys et al. 2014; function `tapas.hgf.m`), we used the series of feedback scores as input:  $u^k$ : = score; normalized to range 0-1. From the series of inputs, the HGF then generates belief trajectories about external states, such as the reward value of an action or a stimulus."

In Figure 5 we have additionally indicated which performance measure we used as response  $y^k$ , based on the winning model.

Please clarify how the precision of the input is measured? (used in Equation 3).

Here we followed Mathys et al., (2014) and the HGF toolbox that recommend to use as prior on the precision of the input ( $\pi_u^0$ : estimated in the logarithmic space) the negative log-variance of the first 20 inputs (observed outcomes). More specifically:

$\log(\pi_u^0)$  is the negative log-variance of the first 20 feedback scores.

This prior is now included in Table 1.

That is, for a participant with very stable initial 20 outcomes, the variance would be small (<1), and the log-precision on the input would be large: the participant is initially less uncertain about the input.

By contrast, a participant with larger variability in feedback scores across the first 20 trials would have a small prior value on the precision of the feedback scores: the participant attributes more uncertainty to the input.

When mentioning the precision of the input in the manuscript (subsection "Computational Model") we refer the readers to Table 1.

"predicted reward" – from the Mathys, 2014 paper we gathered that this is the mean of  $x_1$  obtained on the previous trial? Is this correct? If so, please clarify/emphasize. To increase broad accessibility of the manuscript, it would be helpful to summarize in words somewhere what the model is doing to make predictions. For example, in subsection "Computational Model" we weren't sure what the difference is between prediction of  $x_1$  and expectation of  $x_1$ . Typically this terminology would correspond to: prediction =  $E(x_1 \text{ on trial } k \mid \text{information up to trial } k-1)$ , and expectation =  $E(x_1 \text{ on trial } k \mid \text{information up to trial } k)$  but it wasn't clear to us.

Agreed. Thanks for pointing this out. Yes, the reviewers assumed correctly:

The difference between the prediction of an estimate (denoted by the diacritical mark "hat" or "̂"),  $\hat{\mu}_{ik}$  and its expectation  $\mu_{ik}$ , is that the prediction is the value of the estimate before seeing the input in the current trial  $k$ , therefore  $\hat{\mu}_{i\{k\}} = \mu_{i\{k-1\}}$ . We have made this more explicit in the equations and in the text in subsection "Computational Model":

The first term in the above expression is the change in the *expectation* or current belief  $\mu_{i\{k\}}$  for state  $x_i$ , relative to the previous expectation in trial  $k-1$ ,  $\mu_{i\{k-1\}}$ . The expectation in trial  $k-1$  is also termed *prediction*,  $\mu_{i\{k-1\}} = \hat{\mu}_{i\{k\}}$ , denoted by the "hat" or diacritical mark "̂". The term prediction refers to the expectation before seeing the feedback score on the current trial and therefore corresponds with the posterior estimates up to trial  $k-1$ . By contrast, the term expectation will generally refer to the posterior estimates up to trial  $k$ . In addition, we note that the term belief will normally concern the current belief and therefore the posterior estimates up trial  $k$ . "

In addition, when referring to Variational Bayes and the derivation of update equations (Mathys et al., 2014, appendices), we add in subsection "Computational Model":

"coupling between levels indicated above has the advantage of allowing simple variational inversion of the model and the derivation of one-step update equations under a mean-field

approximation. This is achieved by iteratively integrating out all previous states up to the current trial  $k$  (see appendices in Mathys et al., 2014)."

"variance vs. precision vs. uncertainty" – These are well defined words but it would also help immensely to only use "variance" or "precision" or "uncertainty" in the explanation/equations. Mentally jumping back and forth gets confusing.

Agreed. In the Results section we have now more consistently used uncertainty, as this is the quantity that is directly obtained in the HGF toolbox and may also be understood in a more intuitive way by the readers. In the methods and materials section, however, we have maintained the term precision in the equations, as they have a simplified form this way.

When introducing precision-weighted PEs, we have of course kept that term, as this is what all authors use. But when analyzing the HGF belief trajectories and related uncertainty we have tried to avoid using "precision".

The connection between both terms is now additionally made in subsection "Computational Model":

uncertainty  $\sigma$  or its inverse, precision  $\pi = 1/\sigma$

"belief vs. expectation" – Are these the same? What is the mathematical definition (is it Equation 3 and Equation 7)?

See above.

"In addition, the term belief will generally refer to the current belief and therefore to the posterior estimates up trial  $k$ ."

"pwPE" – please list the equation for this somewhere in the methods, ideally before use of the epsilons in the response variable models.

We have clarified this in subsection "Computational Model":

"Thus, the product of the precision weights and the prediction error constitute the precision-weighted prediction error (pwPE), which therefore regulates the update of the belief on trial  $k$ "

$$\Delta \mu_i^k = \epsilon_i^k$$

And have included Equation (14) and Equation (15) for  $\epsilon_1$  and  $\epsilon_2$ , respectively. These equations are simply a regrouping of terms in Equation (6) and Equation (10) in subsection "Computational Model".

2) Inputs/outputs of the model:

Inputs – we gather that the input to the model is the score that the participant receives. Then  $x_1$  gets updated according to Equation 3. So  $x_1$  is tracking the expected reward on this trial assuming that the reward on the previous trial must be updated by a prediction error from the current trial? Is this reasonable assumption for this task? What if the participants are exploring new strategies trial to trial? Why would they assume that the reward on the next trial is the same as the current trial (i.e. why is the predicted reward =  $\mu_{i,k-1}$ ?) Or is this the point (i.e. if trial to trial the subjects change their strategy a lot that this will end up being reflected as a higher "volatility"?) It would be helpful to outline how the model reflects different regimes of behavior (i.e. what does more exploratory behavior look like vs. what is learning expected to look like).

Using the HGF and the new response models (see below, we have followed the reviewers suggestion to link the change in responses  $\text{cv}KI_{\text{trial}}$  from trial  $k-1$  to  $k$  to computational quantities on the previous trial  $k-1$ ), we can better address the relation between a behavioral change (i.e. a change in strategy) and the belief estimates. We have also created Figure 5—figure supplement 1 for simulated responses. This figure allows us to observe how different behavioral strategies impact belief and uncertainty estimates. We considered agents whose performance is characterized by (a) small and consistent task-related behavioral changes from trial to trial, (b) larger and slightly noisier (or more exploratory) task-related behavioral changes from trial to trial, (c) very large and very noisy (high exploration) task-related behavioral changes from trial to trial.

We explain below in our answer to point 3, the details of how these types of behavior influence belief and uncertainty estimates but the summary is:

If "the participants are exploring" more "new strategies trial to trial" then they will observe more different types of scores, and the distribution of feedback scores will be broader. This leads to a broader distribution of the expectation of reward,  $m_1$ , and therefore higher uncertainty about reward. Simultaneously this is associated with increased volatility estimates and smaller uncertainty

about volatility. The higher volatility estimates obtained in agents that exhibit a more exploratory behavior do not necessarily reflect pronounced increases across time in volatility but rather a lack of reduction in volatility. This effect results from smaller update steps in volatility estimates, due to both high  $s_1$  in the denominator of the update equations for volatility and low  $s_2$  in the numerator, see Equation (5).

So the main link is between a more exploratory behavior leading to more variable reward estimates (which feedback into the update equations as prediction errors at the lower level and as an enhanced uncertainty in volatility,  $s_1$ ). These effects ultimately maintain volatility estimates to a high level, or may even increase them.

Please, see below question 3 as we provide a more detailed explanation of Figure 5—figure supplement 1 and also of the new response model – which was suggested by the reviewers and it is actually a much better model (in terms of log-model evidence and also in terms of allowing to understand better the between-group differences).

Outputs – response models; please clarify why  $cvIKI_{\text{trial}}$  and  $\log(mIKI)$  are the chosen responses since these are not variables that are directly responsible for the reward? We thought that the objective of this response modeling was to determine how a large prediction error on the previous trial would influence action on the next trial? Perhaps an output metric could be [similarity between trial  $k$ , trial  $k-1$ ] =  $B_0 + B_1(u_{ik-1}) + B_2(pwPE_{k-1})$ ? So, depending on the reward and previous prediction error, you get a prediction of how similar the next trials' response is to the current trials' response? Right now, we don't understand what is learned from seeing that  $cvIKI_{\text{trial}}$  is higher with higher reward expectation (this is almost by necessity right? because the rewarded pattern needs high  $cvIKI$ ) or higher prediction error.

Yes, we completely agree that this type of response model is more interesting. In the last manuscript we followed Marshall et al., which explain responses  $\log(RT)$  in trial  $k$  as a function of HGF quantities in trial  $k$ . However, in our paradigm it is more interesting to link the HGF perceptual beliefs and their precision-weighted prediction errors to the “change” in behavior. We have now replaced as suggested the original response variables ( $cvIKI_{\text{trial}}$  and  $\log(mIKI_{\text{trial}})$  at trial  $k$ ) with their trial-wise difference:  $\Delta cvIKI_{\text{trial}}$  or  $\Delta \log(mIKI_{\text{trial}})$  reflecting the difference between current trial  $k$  and previous trial  $k-1$ .

First, a clarification on why we had chosen as performance variables  $cvIKI_{\text{trial}}$  and  $\log(mIKI_{\text{trial}})$ , see subsection “Bayesian learning modeling reveals the effects of state anxiety on reward-based motor learning”:

“Variable  $cvIKI_{\text{trial}}$  was chosen as it is tightly linked to the variable associated with reward: higher differences in IKI values between neighboring positions lead to a higher vector norm of IKI patterns but also to a higher coefficient of variation of IKI values in that trial (and indeed  $cvIKI_{\text{trial}}$  was positively correlated with the feedback score across participants, nonparametric Spearman  $\rho = 0.69$ ,  $P < 10e - 5$ ). Alternatively, we considered the scenario in which participants would speed or slow down their performance without altering the relationship between successive intervals. Therefore, we used a performance measure related to the mean tempo,  $mIKI$ . ”

Now we use those performance variables as well however the new response models include the difference between trial  $k$  and trial  $k-1$  in those performance variables and link them to the belief estimates and  $pwPE$  in the trial before,  $k-1$ . The code is provided at the Open Science Framework, under the accession number [sg3u7](https://osf.io/sg3u7).

We have done family-level Bayesian model comparison (one family of models for  $\Delta cvIKI_{\text{trial}}$  and a separate family of models for  $\Delta \log(mIKI)$ ), followed by additional BMC within the winning family. The response model that had more evidence is based on the  $pwPE$ s (model HGF14, Equation 2):

$$cvIKI_{\text{trial}k} = \beta_0 + \beta_1 \mu_{1k} + \beta_2 \epsilon_{1k} + \zeta$$

This model explains the change in  $cvIKI_{\text{trial}}$  from trial  $k$  to  $k-1$  as a function of  $pwPE$  on reward and volatility on the preceding trial. Moreover, we obtained an interesting between-group difference in the  $\beta_2$  coefficients of the response model, supporting that large  $pwPE$  on volatility promote larger behavioral changes on the following trial in control participants, yet they inhibit or constrain behavioral changes in *anx1* and *anx2* participants (see Figure 5 —figure supplement 3). In addition, in all groups,  $\beta_1$  is negative, indicating that smaller  $pwPE$  on reward on the last trial (reduced update step in reward estimates) promotes an increase in the changes in the relevant performance variable, thus an increase in exploration. By contrast, an increase in  $m_1$  updates through large  $pwPE$  on reward, is followed by a reduction in  $cvIKI_{\text{trial}}$  (more exploitation).

Additional examples illustrating the implications of the winning response model are included as Figure 5—figure supplement 4 and Figure 5—figure supplement 5.

The former response models that assessed whether  $\text{cvIKI}_{\text{trial}}$  and  $\log(\text{mIKI}_{\text{trial}})$  in trial  $k$  can be explained by pwPE or belief estimates in the same current trial  $k$  have not been included in the new manuscript. However, for the reviewer team we provide the results of the BMS applied to the total of four families of models (two old families  $F_1$  and  $F_2$  for  $\text{cvIKI}_{\text{trial}}$  and  $\log(\text{mIKI}_{\text{trial}})$  in trial  $k$ , HGF quantities in trial  $k$ ; and two new families  $F_3$  and  $F_4$  for the change  $k-1$  to  $k$  in  $\text{cvIKI}_{\text{trial}}$  and  $\log(\text{mIKI}_{\text{trial}})$ , HGF quantities in trial  $k-1$ ). BMS using the log-family evidence in each family provided more evidence for the new families,  $F_3$  and  $F_4$ , as indicated by an expected frequency of:

0.0160 0.0165 0.9335 0.0340

And an exceedance probability of

0 0 1 0

This demonstrates that the third family of models (related to  $\Delta\text{cvIKI}_{\text{trial}}$ ) outperforms the other families.

### 3) Interpretation of the model:

Is the message from Figure 6A that the expected reward is lower for anx1 than anx2 and control? Since the model is trying to predict scores from actual score data, isn't this result expected given Figure 4A. Can the authors please clarify this?

Correct. The HGF as a generative model of the observed data (feedback scores) provides a mapping from hidden states of the world (i.e. true reward  $x_1$ ) to the observed feedback scores ( $\mu$ ). Anx2 and control participants achieved higher scores (Figure 4) and therefore the HGF perceptual model naturally provides trajectories of beliefs about reward with higher expectation values,  $\mu_1$ , than in anx1. We acknowledge that this result is a kind of "sanity check" and is not the emphasis of the interpretation and discussion in the new manuscript. A mention of this expected result is included in the new manuscript, subsection "Bayesian learning modeling reveals the effects of state anxiety on reward-based motor learning":

"Participants in the anx1 relative to the control group had a lower estimate of the tendency for  $x_1$ . . . . This indicates a lower expectation of reward on the current trial. Note that this outcome could be anticipated from the behavioral results shown in Figure 4A."

Using the new winning response model and associated results, the manuscript now places more emphasis on the obtained between-group differences in the response model parameters ( $\beta$  coefficients, Figure 5—figure supplement 3; see also Figure 10, Figure 10—figure supplement 1), as well as on the parameters of the perceptual HGF model ( $w_1$  and  $w_2$ , with  $w_2$  being different between anx1 and control participants, and thus reflecting a different learning style or adaptation of volatility estimates in anx1).

We noticed that  $\log(\mu_2)$  is lower for anx1 and anx2 than control. Should this correspond to a shallower slope (or a plateau in score that is reached more quickly) in Figure 4A over learning for anx1? If so, why don't we see that for anx2? If this is true, and given that  $\text{cvIKI}$  is no different for anx1, anx2, and control, wouldn't that mean that the reward rate is plateauing faster for anx1 and anx2 while they are still producing actions that are equally variable to control? So, are participants somehow producing actions that are variable yet getting the same reward – so they're getting "stuck" earlier on in the learning process? Can the authors provide some insight into what type of behavior trends to expect given the finding of Figure 6B-C? Right now all the reader gets as far as interpretation goes is that the anx1 group underestimates "environmental volatility" and that the mean behavior and  $\text{cvIKI}$  is the same across all groups.

To answer this question we have created Figure 5—figure supplement 1 for simulated responses (see legend for details).

The simulated responses have been generated by changing the pattern of inter-keystroke intervals on a trial by trial basis to a different degree, e.g. leading to a steeper (green lines) or shallower (pink lines) slope of change in  $\text{cvIKI}_{\text{trial}}$  (Figure 5—figure supplement 1B) and associated feedback score (Figure 5—figure supplement 1A). The feedback scores are illustrated in Figure 5—figure supplement 1A to align it to Figure 5—figure supplement 1C below displaying reward estimates,  $m_1$ .

The figure demonstrates that a shallower slope in the feedback score function is associated with a shallower slope in the trajectory of reward estimates,  $m_1$ , and smaller estimation uncertainty on that level,  $s_1$  (Figure 5—figure supplement 1E). More importantly, this scenario is also associated with smaller  $\log(m_2)$  estimates (Figure 5—figure supplement 1D) and greater estimation uncertainty  $s_2$

(Figure 5—figure supplement 1F). This case of shallower slope could represent anx1 participants (Figure 6).

These results also confirm the relationship between higher estimation uncertainty on one level,  $s_i$ , and larger updates in the beliefs on that level,  $m_i$ , that characterize the HGF. See Equation (5).

In addition to simulating responses that lead to different slopes of the feedback score trajectory, we have also simulated responses with different levels of noise or variation from trial to trial (while keeping the slope constant as underlying trend: green and pink trajectories). We considered these three scenarios:

i) Smooth trial-by-trial change in  $cvKI_{\text{trial}}$  and corresponding feedback scores (linear trends in panels A and B)

ii) Slightly noisy or variable transition from trial to trial in  $cvKI_{\text{trial}}$  and corresponding feedback scores – moderate noise level (slightly jerky trajectories, shown as darker green or pink lines)

This scenario represents an agent changing slightly more randomly their responses from trial to trial.

iii) Highly noisy or variable transition from trial to trial in  $cvKI_{\text{trial}}$  and corresponding feedback scores – high noise level (pronounced jerky trajectories, shown as the darkest green or pink lines).

This scenario represents an agent changing significantly more randomly their responses from trial to trial.

Green lines, constant steep slope: Increasing level of noise in the behavioral responses associated with higher variation in trial-by-trial changes leads to higher  $\log(\mu_2)$  and reduced uncertainty about volatility,  $\sigma_2$ . In addition, the more variable changes in reward estimates have higher uncertainty,  $\sigma_1$

Pink lines, constant shallow slope: Similar results for increasing level of noise as described for the steep slope trajectories.

Thus, based on these simulation results, higher expectation on volatility in the HGF for continuous inputs can result from:

1) A steeper slope in feedback scores and therefore a steeper slope in the trajectory of perceptual beliefs for reward,  $\mu_1$ .

2) More variable trial-to-trial changes in the observed feedback scores (corresponding with a more exploratory or noisier performance). This would also lead to more variable trial-to-trial changes in the perceptual beliefs for reward,  $\mu_1$ .

These two cases come down to one single general case:

A broader range of values in the distribution of observed inputs ( $u$ ) that lead to a broader distribution of reward estimates,  $\mu_1$ .

With regard to the HGF belief trajectories for volatility,  $\mu_2$ , in our experimental and control groups, we have noted in subsection “Bayesian learning modeling reveals the effects of state anxiety on reward-based motor learning” that:

“As indicated above, volatility estimates are related to the rate of change in reward estimates, and accordingly we predicted a higher expectation of volatility  $\mu_2$  in participants exhibiting more variation to  $\mu_1$  values.”

This is interesting but also simply implies that in participants achieving more different feedback score values (i.e. because they encounter all values from low to high scores), the volatility estimate will be higher (control group). By contrast, participants getting stuck at low score values (anx1) will have a reduced volatility estimate (due to a smaller rate of change of the estimate on the level below). This is what our findings in Figure 5 confirm, in line with the results for simulated responses in Figure 5—figure supplement 1. We anticipate this behavior of the HGF model in subsection “Bayesian learning modeling reveals the effects of state anxiety on reward-based motor learning”:

“Additionally, the HGF estimation of volatility (as change in reward tendency) was expected to be higher in participants modulating more their performance across trials and thereby observing a broader range of feedback scores (see different examples for simulated performances in Figure 5—figure supplement 1).”

The case of anx2 is interesting as these participants had a similarly steep slope in feedback scores and in the trajectory for  $\mu_1$  as the control group, however their log-volatility estimates  $\mu_2$  and their uncertainty  $\sigma_2$  resemble more the trajectories observed in anx1.

Accordingly, from the two cases contributing to higher volatility estimates indicated above, the likely explanation for the results in anx2 is that these participants must have a narrower distribution

of encountered scores than control participants, and/or a smaller trial to trial change in the performance measure  $cvKI_{\text{trial}}$ .

We tested this prediction and found:

- The mean difference between trial  $k-1$  and  $k$  in  $cvKI_{\text{trial}}$  (our performance measure  $\Delta cvKI_{\text{trial}}^k$ ) was significantly smaller in anx2 than control participants: mean 0.005 (SEM 0.0011) in controls, 0.0032 (0.0007) in anx2,  $P_{\text{FDR}} < 0.05$ . In anx1 participants this parameter was also smaller than in control participants: 0.0013 (0.0009),  $P_{\text{FDR}} < 0.05$ .

- The variance of the observed feedback scores was significantly smaller in anx2 than in control participants: mean 0.064 (SEM 0.004) in controls; 0.052 (SEM 0.003) in anx2,  $P_{\text{FDR}} < 0.05$ . A non-parametric Spearman correlation between these two parameters ( $\rho = 0.4563$ ,  $P = 0.0282$ ) further confirmed that higher volatility estimates were associated with a larger variance of the distribution of feedback scores.

This is now presented as a post-hoc analysis in subsection “Bayesian learning modeling reveals the effects of state anxiety on reward-based motor learning”:

“...Thus, anx2 participants achieved high scores, as did control participants, yet they observed a reduced set of scores. In addition, their task-related behavioral changes from trial to trial were more constrained but also goal-directed as they indicated a tendency to exploit their inferred optimal performance, leading to consistently high scores. This different strategy of successful performance ultimately accounted for the reduced estimation of environmental volatility in this group, unlike the higher  $\mu_2$  values obtained in control participants.”

Anx2 participants therefore showed a tendency to exploit more their inferred best response and thus observed fewer outcomes: they moved quickly from low to high feedback.

Interestingly, however, volatility estimates  $\log(\mu_2)$  and  $\Delta cvKI_{\text{trial}}^k$  were not correlated in the  $N = 60$  population. We only found a correlation between  $\log(\mu_2)$  and the variance of the feedback scores distribution,  $r = 0.30$ ,  $p = 0.019$ . This also explains why there were no significant effects between groups in the degree of across-trials variability ( $cvKI$ , Figure 4). So it seems that, although behavioral changes directly fed to the score modulation across trials, the most robust association was between the variance of the distribution of scores and volatility estimates.

In the adapted manuscript, following other papers using the HGF (see e.g. Marshall et al., 2016, Weber et al., 2019), the emphasis is placed now on the between-group differences in perceptual or response model parameters. Additionally, we maintain our emphasis on the analysis of pwPEs and how they relate to beta oscillatory activity and behavioral responses.

Does underestimating volatility mean that subjects just keep repeating the same sequence over and over? If so, can that be shown? Or does it mean that they keep trying new sequences but fail to properly figure out what drives a higher reward? Since the model is fit on the behavior of the participants, it should be possible to explain more clearly what drives the different model fits.

See above, please.

Related EEG Analysis: We greatly appreciated the clarified EEG analysis. Re-reading this section, we were able to understand what was done much better, but had two queries related to the analysis.

1) We noted that the beta envelope in Figure 7A looks unusual. It looks almost like the absolute value of the beta – filtered signal rather than the envelope, which is typically smoother and does not follow peaks and troughs of the oscillation. Can the authors please clarify how this was calculated?

Thanks for spotting this. Yes, the figure was not correct. We have amended it and also uploaded to the OSF (<https://osf.io/nv4m3/>) the original code we used to compute the amplitude envelope from the band-pass filtered and Hilbert-transformed data. As in our earlier work (e.g. Herrojo Ruiz et al., 2014), the amplitude envelope  $A(t)$  of the instantaneous analytic signal was computed after applying the Hilbert transform to the bandpass-filtered raw data (12–35 Hz; two-way least-squares FIR filter applied with the `eegfilt.m` routine from the EEGLAB toolbox, Delorme and Makeig, 2004) spanning the full continuous recording of the task performance. Next, from the total beta-band amplitude envelope we extracted data segments corresponding with the epochs locked to the feedback presentation from -9 to 2 s.

We highlight here the main MATLAB steps:

```
% EEGdata: dimensions 64 channels x Nsampl, continuous data
% srate: sampling rate, 512Hz
f1=12; f2=35;% bounds for band-pass filter
```

```
betatot = eegfilt(EEGdata,srate,f1,f2);
amplitudebetatot=transpose(abs(hilbert(betatot')));
% after this step we extracted the epochs that were used to detect oscillation bursts
```

2) In subsection "Analysis of power spectral density", the authors write: "The time-varying spectral power was computed as the squared norm of the complex wavelet transform, after averaging across trials within the beta range." This sounds like the authors may have calculated power after averaging across trials? Is this correct (i.e. was the signal averaged before the wavelet transform, such that trial to trial phase differences may cancel out power changes)? Or do the authors mean that they averaged across trials after extracting beta power for each trial? If the former the author should emphasize that this is what they did, since it is unconventional.

We have clarified this in the new version of the manuscript. In brief, the time-frequency transformation is first performed for each trial separately, followed by averaging. This is the standard practice to obtain the total oscillatory activity (induced + evoked). This thus converges with the reviewers' expectations.

The analysis was done using Morlet wavelets based on convolution in the time domain. After the time-frequency transformation of each epoch, we obtained for each trial the wavelet energy, which was computed as the squared norm of the complex wavelet transform of signal  $x$  (for each trial):

$$E_x(t,f) = |W_x(t, \eta/2\pi f)|^2$$

In this expression equation,  $h$  is the wavelet family function or number of cycles. The expression is taken from our earlier work e.g. Herrojo Ruiz et al. (2009).

Next, we assessed the spectral content of the oscillatory activity using the trial-average of the wavelet energy.

We have modified the text in the manuscript to clarify the analysis steps (and corrected a typo: the windows were set every 50ms). Subsection "Analysis of power spectral density":

"Artefact-free EEG epochs were decomposed into their time-frequency representations using a 7-cycle Morlet wavelet in successive overlapping windows of 50ms within the total 12s-epoch. The frequency domain was sampled within the beta range from 13 to 30 Hz at 1 Hz intervals. For each trial, we thus obtained the complex wavelet transform, and computed its squared norm to extract the wavelet energy (Ruiz et al., 2009). The time-varying spectral power was then simply estimated by averaging the wavelet energy across trials within the beta range. "

In our earlier work we had used our own code to obtain the wavelet transformation with Morlet wavelets. Accordingly, we manually coded the trial-based time-frequency analysis followed by the calculation of the squared norm and then trial-averaging.

For this study, however, we used the built-in functions in the fieldtrip toolbox, which also follow this approach. The link to the uploaded code is provided in the next question. Here we only highlight the details of the fieldtrip analysis configuration:

```
cfg = [];
cfg.output = 'pow';
cfg.channel = 'all';
cfg.precision = 'single'
cfg.method = 'tfr';% implements wavelet time frequency transformation
% (using Morlet wavelets) based on convolution in the
% time domain.
cfg.foi =[13:1:30];
cfg.toi= -9:0.05:3;
cfg.width = 7;% default
cfg.trials = 1: length(EEG.trial);
cfg.keeptrials = 'yes'
TFRwav7 = ft_freqanalysis(cfg, EEG);
```

3) To try to understand point 2 above, we checked if the authors had shared their code, and found that, although data was shared, code was not, as far as we could tell. eLife does require code sharing as part of their policies (<https://reviewer.elifesciences.org/author-guide/journal-policies>) so please include that.

We have now included the code in the folder “Code for analysis of bursts and time-varying spectral power” of the Open Science Framework website for this study:

<https://osf.io/nv4m3/>

The script `get_timecourse_wavelet.m` (and Wiki) illustrates how to compute the time-varying spectral power in the beta-band (13-30Hz) after implementing the wavelet time frequency transformation (using Morlet wavelets) based on convolution in the time domain. It calls fieldtrip function `ft_freqanalysis.m`

---

## Additional information

### Funding

Funder	Grant reference number	Author
British Academy	R134610	Maria Herrojo Ruiz
Economic and Social Research Council	ES/P00072X/1	Thomas Hein
National Research University Higher School of Economics	Basic Research Program	Maria Herrojo Ruiz
Ministry of Science and Higher Education of the Russian Federation	Russian Academic Excellence Project 5–100	Maria Herrojo Ruiz

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

Sebastian Sporn, Conceptualization, Investigation, Writing - review and editing; Thomas Hein, Investigation, Writing - review and editing; Maria Herrojo Ruiz, Conceptualization, Software, Formal analysis, Supervision, Funding acquisition, Investigation, Visualization, Methodology, Writing - original draft, Project administration, Writing - review and editing

### Author ORCIDs

Maria Herrojo Ruiz  <https://orcid.org/0000-0001-8948-9444>

### Ethics

Human subjects: Participants gave written informed consent prior to the start of the experiment, including written consent to potentially share de-identified data with other researchers. Experimental procedures were approved by the research ethics committee of Goldsmiths University of London.

### Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.50654.sa1>

Author response <https://doi.org/10.7554/eLife.50654.sa2>

---

## Additional files

### Data availability

MIDI (performance) and EEG data, as well as new response model scripts, have been deposited in the Open Science Framework Data Repository under the accession code `mfe2j`.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Ruiz MH	2019	Motor Learning and Anxiety - Data repository - behavioral, electrophysiological	<a href="https://osf.io/mfe2j/">https://osf.io/mfe2j/</a>	Open Science Framework, <code>mfe2j</code>

