Performance anxiety is associated with biases in learning from reward and punishment in skilled individuals

Andrea Erazo Hidalgo[1, §], Lisa Pearson[1], Takanori Oku[2], Yudai Kimoto[3], Shinichi Furuya[3], María Herrojo Ruiz*[1, §]

1: Department of Psychology, Goldsmiths, University of London, London, UK
2: Shibaura Institute of Technology, Tokyo, Japan
3: Sony Computer Science Laboratories Inc., Tokyo, Japan
*Corresponding author: M.Herrojo-Ruiz@gold.ac.uk
§AEH and MHR contributed equally to this work.

Abstract

Many individuals experience performance anxiety (PA) in high-stakes situations, from public speaking to the performing arts. While debilitating PA is associated with physiological, cognitive, and affective alterations, its underlying mechanisms remain unclear. Using behavioural analysis, computational modelling, and electroencephalography, we investigated whether PA predisposes individuals to learn faster from punishment than reward, particularly under high task uncertainty. Across three experiments with 95 skilled pianists, participants learned hidden melody dynamics through reinforcement with graded reward or punishment feedback. Bayesian hierarchical modelling revealed that performers with greater PA levels learn faster from punishment in low-uncertainty environments but increasingly rely on reward as uncertainty escalates. These biases were mediated by reinforcement-driven modulation of motor variability—increasing following poor outcomes—and shifts in frontal theta (4–7 Hz) activity encoding feedback changes and signalling upcoming motor adjustments. The findings reveal that PA alters the weighting of reward and punishment signals based on task uncertainty.

Introduction

Performing in high-stakes, socially evaluative settings—where individuals are judged on their abilities—is a fundamental challenge in human behaviour, spanning domains as diverse as sports, public speaking, and the performing arts. While some individuals thrive in those settings, others experience performance anxiety (PA)—a debilitating condition characterised by anxious apprehension towards performance[1]. PA affects between 25% and 40% of professionals and students across domains[2–5], significantly impacting health and career trajectories, yet its underlying neurocognitive mechanisms remain poorly understood.

PA is characterised by altered physiological, cognitive, and affective states[3,4,6–9], often impairing performance in critical moments. Competitions and stage performances heighten state anxiety, disrupting cardiorespiratory rhythms and motor control[6–8,10]. Laboratory studies show that PA increases muscle stiffness, impairs memory retrieval, and disrupts the automatic execution of well-learned actions[11–13]. Despite these advances, a major gap remains in understanding how PA interacts with fundamental learning processes.

Anxiety disorders are increasingly conceptualised as disorders of learning and decision-making, particularly under uncertainty[14–16], where information is incomplete or the environment is unstable. A prevailing hypothesis is that anxiety is associated with negative learning biases[16–18], whereby individuals exhibit a greater reliance on negative outcomes to update their behaviour or beliefs. Computational studies have shown that clinically anxious individuals learn faster from punishment than from reward[17,18], an effect attributed to biased attention for threats and suppression of reward-seeking behaviour[16,19,20]. Variations in these patterns associate cognitive symptoms of anxiety with faster threat learning, while physiological symptoms increase safety learning[21]. Decision-making studies further show that abnormal learning processes in anxiety intensify as uncertainty increases[22–26]. Similarly, in motor learning tasks involving large, intrinsically uncertain continuous action spaces, state anxiety attenuates reward learning[27]. Based on these findings, we hypothesise that PA promotes negative learning biases, leading affected individuals to rely more on punishment than reward to guide adaptation during performance—an effect that may be exacerbated under uncertainty.

In the motor domain, reward and punishment differentially modulate learning. While punishment increases learning rates during sensorimotor adaptation[28,29], reward improves retention of adaptation and motor skills[28,30,31]. However, inconsistencies in these findings indicate task-dependent effects[32,33]. Faster punishment learning has been explained by greater trial-by-trial motor variability and larger motor updates following negative outcomes[28,31]. After unsuccessful actions, increased task-related variability promotes exploration, enabling the sensorimotor system to more rapidly identify successful actions[34–36].

Motor variability arises from several sources[37,38], including neuromotor and planning noise, along with exploratory variability, which is particularly sensitive to outcomes about success and failure. By dissociating reinforcement effects from behavioural autocorrelations, recent studies demonstrate that poor outcomes causally increase exploratory variability to improve learning[36,39]. Computational modelling complements these findings by revealing how agents adjust different sources of motor variability[36,39]. Together, these approaches provide a framework to test our second hypothesis: that, in the motor domain, PA biases learning from reward and punishment through altered regulation of motor variability.

To identify the neural mechanisms underlying the hypothesised learning biases and altered motor variability regulation in PA, we examined electroencephalography (EEG) oscillations. Prefrontal and sensorimotor beta oscillations (13–30 Hz) are key modulators of motor learning[40–43], including reward-based learning[27,44,45], with beta attenuation post-feedback contributing to updating motor plans[42,45,46]. In line with this, increased beta activity in state-anxious individuals has been linked to attenuated reward processing, impairing the updating of motor predictions[27]. Additionally, frontal midline theta oscillations (4–7 Hz) have been implicated in adaptive control, adjusting behaviour under uncertainty by facilitating switching and exploration in reinforcement learning[47–50]. Theta is also associated with a predisposition to anxiety and heightened responses to punishment[51], suggesting a role in mediating reinforcement-learning biases in PA.

Beta and theta oscillations in these settings have been explained by activations in the anterior cingulate cortex (ACC), prefrontal cortex (PFC), hippocampus, and striatum[52,53]—regions crucial to decision-making, learning under uncertainty, and anxiety[14,54–57]. The striatum, a key structure in reinforcement-based motor learning[58], is part of the cortico-basal ganglia-thalamo-cortical circuits, which are proposed to regulate motor variability[35,37,39,59]. In the cortex, the PFC tracks hidden task states by predicting observations during reward-guided decision-making[60], complementing the role of the basal ganglia in learning via reward prediction errors. Here, we tested whether beta and theta modulation reflects changes in reinforcement-based motor learning in PA. We predicted a more pronounced beta attenuation during punishment learning, associated with faster avoidance learning, and enhanced medial frontal theta, encoding control signals for greater behavioural adjustments and increased motor variability regulation under punishment. If confirmed, these EEG dynamics would mark a neurophysiological signature of learning biases in PA, reflecting maladaptive learning mechanisms that may undermine skilled performance.

Despite extensive evidence linking anxiety to learning biases, investigating these mechanisms in highly trained individuals with a predisposition to PA has remained challenging. This shortfall is partly due to methodological constraints in assessing skilled performance, which requires simultaneous recording of rich performance data and neural activity. To address this, our study focused on skilled pianists, enabling a quantitative assessment of reinforcement learning and motor variability in expert sensorimotor performance.

Across three experiments with 95 pianists, we examined how trait PA influences learning from reward and punishment, reinforcement-driven motor variability, and their neural correlates. We used performance learning tasks requiring pianists to adapt keystroke dynamics (intensity or loudness) to uncover hidden target dynamics in melodies under graded reward or punishment reinforcement. Contrary to our first hypothesis, Experiments 1 and 2 revealed that increasing PA levels were associated with faster reward learning, whereas lower PA levels corresponded to greater reliance on punishment feedback. In Experiment 3, under reduced task uncertainty, these learning biases reversed. Across experiments, reinforcement effects on motor variability regulation following poor outcomes explained learning biases. At the neural level, theta activity encoded unsigned differences in graded reinforcement feedback and predicted upcoming motor variability regulation, accounting for learning biases.

These findings indicate that predisposition to PA manifests as biases in learning from reward and punishment, causally linked to regulation of reinforcement-driven motor variability and associated with changes in oscillatory dynamics. The reversal of learning biases as uncertainty escalates suggests a central connection between uncertainty and PA, with implications for understanding and mitigating its debilitating effects on skilled performance.

Results

To evaluate the dissociable effects of reward and punishment on learning in skilled performers, we developed a performance learning task adapted from previous reward-based motor learning research[27] and used it to collect behavioural and EEG data from a cohort of highly trained pianists (N = 41). The data are available online (see **Data Availability Statement**).

Participants played two piano melodies designed for the right hand on a digital piano (**Figure 1A**). The task entailed varying the dynamics (the pattern of keystroke velocity or loudness) with the aim of uncovering the melody's specific hidden target dynamics. Participants were informed that the target dynamics deviated from the natural flow of the melodies (**Figure S1**) and would not correspond with their initial expectations, requiring exploration to uncover the solution (**Methods**).

After each trial, participants received graded reinforcement feedback, either as reward (scores 0-100) or punishment (-100 to 0), over 100 trials per condition (**Figure 1B**). This feedback reflected their overall proximity to the target dynamics pattern, calculated as a single summary score comparing the full vector of performed keystroke velocities to the target dynamics vector for that melody (**Figure 1C; Methods**). The goal was to infer the hidden dynamics solution and maximise the average score across trials—coupled to a monetary incentive, by either increasing gains in the reward condition or minimising losses in the punishment condition. Concurrently, EEG and MIDI (Musical Instrument Digital Interface) performance data were recorded.

Bayesian workflow of performance analysis

To assess the effects of reinforcement condition on learning and its interaction with trait performance anxiety (PA), we used Bayesian multilevel modelling[61]. PA was evaluated using the validated Kenny music performance anxiety (MPA) Inventory[62]. See simulations for sample size estimates in **Figures S2-S3**.

Following the principled Bayesian workflow[61], we constructed Bayesian beta regression models of feedback scores (rescaled to 0–1) over trials, analysing the effects of reinforcement, PA levels, and their interaction (**Methods**; **Table S1).** Beta regressions were parametrised by the mean μ and precision ϕ of the score distribution[63]. Prior predictive checks with simulated data confirmed that model behaviour aligned with domain expertise (**Figure S4**).

The best-fit model included interactions between reinforcement condition, a monotonic function of PA categorical levels[64], and trial progression, along with random intercepts and slopes for subjects (model M6, **Table S2;** Leave-one-out cross-validation[65], LOO-CV). This model demonstrated good convergence and robust predictive accuracy (**Figure S5**, **Supplementary Materials**).

Scores increased across trials, with a positive effect of 0.00441 per trial on the log-odds scale (95% credible interval, CrI: [0.00243, 0.00651]), equivalent to an increase of 10 points over 100 trials. This validates that participants progressively approached the target dynamics (**Figure 1DE; Table S2)**. Score consistency also increased (greater precision parameter ϕ), and a credible three-way interaction between PA, condition and trial on the scores was observed.

Further analysis of this interaction revealed distinct credible effects of reinforcement condition on the median trend of scores across trials (slope) as a function of PA levels (**Figure 1FG;** effects on the percentage point scale). Low-PA participants learned faster to avoid punishment (negative median slope difference, reward – punishment: -4.81 x 10^{-4}, 95% highest density interval, HDI [-6.60, -3.04] x 10^{-4}). Conversely, individuals with medium-high to high PA learned faster to maximise reward (median slope difference: 6.83 [5.06, 8.61] x 10^{-4} and 8.52 [6.14, 11.42] x 10^{-4}, respectively), with the most pronounced difference at the highest PA level.

These results demonstrate that learning rates were distinctly modulated by reward and punishment as a function of PA, exhibiting a monotonic shift from faster learning under punishment in low PA to faster learning under reward in high PA. The interaction effects on learning trends did not extend to median scores (**Figure S6**). These findings were replicated in a second experiment with an independent sample of 18 highly trained pianists, confirming similar interaction effects on learning slopes (**Figure S7, Table S3)**.

Given previous findings that cognitive (worry) and somatic (physiological) trait anxiety symptoms can influence learning biases differently[21], we conducted control analyses to determine if distinct PA components differentially influence learning biases. Bayesian multilevel modelling revealed that the model incorporating somatic PA subscores[66,67] provided a better fit than the model including cognitive PA (negative cognitions), replicating the interaction effects (Experiments 1 and 2: **Figures S8-S9, Table S4**). This suggests that learning biases in skilled performers are better explained by the debilitating physiological dimension of PA.

Learning biases in performance anxiety are underpinned by changes in motor variability

To assess our second hypothesis, we examined trial-by-trial changes in motor variability. Our task involved a large continuous action space, requiring participants to infer hidden melody dynamics across 16 (8 x 2) keystroke velocities using reinforcement feedback. In such environments, learning can be effectively guided by reinforcement-driven regulation of motor variability[35–37,39]: variability increases following unsuccessful outcomes to promote exploration, and decreases after successful outcomes to stabilise performance.

To assess variability in keystroke velocity, we first transformed the trial-wise 16-dimensional velocity vector into a scalar variable, ΔV^n. This variable represented the magnitude of change in velocity patterns between consecutive trials (*n-1* and *n*), computed using the normalised sum of absolute differences between $\boldsymbol{V}^n$ and $\boldsymbol{V}^{n-1}$ (**Methods**; **Figure 1C**)[68]. Following ref.[36], variability was assessed by calculating the variance of ΔV^n values within moving windows of five trials.

As expected[35,36,68], participants increased variability following poor outcomes—a pattern absent for high outcomes (median split of scores; **Figure 2A-B**). Slow fluctuation trends, potentially reflecting autocorrelations in performance[40,69–73], were also evident in the variability function in renditions preceding and following conditioned trials[36] (relatively low or high scores; **Figure 2B**).

To dissociate performance autocorrelations from reinforcement-dependent variability and determine whether trial outcomes causally influence motor variability in our task, we implemented two validated approaches: statistical matching analysis and computational generative modelling[36,39] (**Methods**, and next section). Statistical matching analysis indicated that poor outcomes led to a gradual increase in variability over 3-4 trials (**Figure 2D**). This increase was significantly greater than that following high scores (paired permutation test; P_{FDR} = 0.0038; non-parametric effect size estimator, Δ_{dep} = 0.74, CI = [0.65, 0.89]). Separately, we observed that larger deviations from target velocity patterns (lower scores) resulted in greater subsequent reinforcement-related variability (**Figure 2E**), in line with previous work[39].

To determine whether learning biases arose from changes in the regulation of motor variability, we applied Bayesian Gaussian linear modelling to analyse *VarDiff*—the difference in variability following poor versus good outcomes—as a function of PA category, reinforcement condition, and their interaction. The model demonstrated good convergence (**Table S5**) and revealed a credible interaction between both variables. A negative estimate (-1.06, 95% CrI: [-2.13, -0.01]) indicated that punishment, compared to reward, reduced variability following poor outcomes as PA levels increased (**Figure 2F**). This effect was most pronounced for high PA individuals, where poor outcomes did not elicit increased variability under punishment (95% CrI overlapping zero). No other effects were observed (**Table S5**).

Thus, controlling for behavioural autocorrelations, our analysis provided consistent evidence that reinforcement-driven motor variability underpins learning biases in skilled performers as a function of PA. Moreover, high-PA individuals exhibited the most contrasting responses to reward and punishment, with preserved motor variability regulation under reward but blunted responses under punishment.

Control analyses defining low and high scores based on relative trial-to-trial score changes confirmed that motor variability regulation was primarily driven by poor outcomes below the median of the score distribution (**Supplementary Materials**). Notably, these low scores were distributed across the entire session rather than concentrated in earlier trials (**Supplementary Materials**), ruling out confounds from early-session effects.

Generative model dissociating reinforcement-sensitive and autocorrelated behaviour in motor variability

The previous results suggest that keystroke dynamics were influenced by reinforcement-driven changes in motor variability, alongside slower autocorrelations in performance. Similar patterns have been observed in humans and non-human primates[39], as well as rodents[36]. In such settings, a control strategy involves counteracting autocorrelations—which reduce reward rates—by increasing variability to explore and identify reinforced solutions[34,38,39,74,75].

To investigate whether this control strategy underlies the learning biases associated with PA, we employed a reinforcement-sensitive Gaussian process[39] (RSGP; **Methods**). The RSGP models behavioural time series by incorporating long-term autocorrelations and short-term reinforcement effects on motor variability via two kernels (**Figure 3A**). Each kernel is defined by two hyperparameters: a characteristic length-scale (*l*), indicating the length of trial-to-trial dependencies, and an output scale (σ^2), quantifying the magnitude of variability attributed to that process. In this framework, σ^2_{RS} reflects the *latent* contribution of short-term, reinforcement-sensitive processes to motor variability, while σ^2_{SE} (squared exponential kernel) reflects variability due to longer-term autocorrelations.

In line with ref.[39], we used the trial-wise *signed* error (e^n)—the difference between the produced and target keystroke velocity vectors—as the variable predicted by the RSGP at trial *n* (**Figure 3B**; **Methods**; Eq. 2), assuming a zero mean Gaussian process. Observed motor variability was quantified as the standard deviation of the error distribution, $\sigma(e^n)$, and assessed in relation to the error on the previous trial, e^{n-1}.

Simulations confirmed reliable recovery of model parameters (l_{SE}, l_{RS}, σ^2_{SE}, σ^2_{RS}; **Table S6)** and generated predictive distributions of $\boldsymbol{e}^{\mathbf{n}}$ per trial, based on reward history and prior error e^{n-1}, characterised by the mean

$\mu(e^n)$ and standard deviation $\sigma(e^n)$. The simulations revealed a U-shaped relationship between e^{n-1} and $\sigma(e^n)$, while $\mu(e^n)$ increased linearly with e^{n-1}, consistent with previous findings[39] (**Figure 3C)**.

Fitting the RSGP to empirical data revealed that the autocorrelation kernel (K_{SE}) had a significantly longer length-scale (l_{SE} = 12.31 [1.5]) and larger output scale (σ^2_{SE} = 4.05 [0.2]) than the reinforcement-sensitive kernel (K_{RS}; **Methods**; l_{RS} = 2.80 [0.3]; σ^2_{RS} = 2.75 [0.1]; P_{FDR} = 0.0002; Δ_{dep} = 0.80, CI = [0.73, 0.88] for l, Δ_{dep} = 0.76, CI = [0.60, 0.79] for σ^2). Thus, autocorrelation effects spanned ~12 trials, while reinforcement effects decayed after ~3 trials. Bayesian regression modelling (log-normal family) demonstrated a credible negative effect of PA category on σ^2_{RS} (log scale: -0.17, 95% CrI = [-0.29, -0.06]), indicating that the latent contribution of reinforcement-sensitive variability to e^n decreased with increasing PA, regardless of reinforcement type. No credible effects were observed for σ^2_{SE} (**Supplementary Materials; Figure 3D**).

Simulating $\mathbf{e}^n$ from individual parameter estimates replicated the empirical U-shaped relationship between $\sigma(e^n)$ and e^{n-1} and the linear increase in $\mu(e^n)$ with e^{n-1} (**Figure 3E**). This supports that motor variability increased more after trials with lower scores (greater e^{n-1}). To link these results to **Figure 2E**, we transformed e^{n-1} into positive values, $|e^{n-1}|$, and modelled the nonlinear relationship between $\sigma(e^n)$ and $|e^{n-1}|$. An exponential Bayesian regression model ($\sigma(e^n) = b_1 \exp(b_2 \; |e^{n-1}|)$) best explained the data (LOO-CV; **Figure 3F**; **Supplementary Materials**), with non-zero posterior estimates for both b_1 and b_2. A credible interaction between PA category and reinforcement condition on b_2 (-0.03, 95% CrI = [-0.07, -0.01]) indicated that higher PA dampened the exponential growth of $\sigma(e^n)$ under punishment compared to reward (**Figure 3G**), suggesting reduced sensitivity of observed motor variability to prior error under punishment for higher PA.

Electroencephalography markers underlie learning biases and motor variability regulation

Having established that heightened PA levels in skilled pianists are associated with increased motor variability following poor outcomes under reward, but a blunted modulation under punishment, we next examined the neural processes underlying these behavioural effects across the theta and beta bands. We assessed frequency-domain amplitude changes related to processing graded feedback and regulating motor variability in keystroke dynamics using validated linear convolution models for oscillatory responses[76]. The frequency-domain general linear model (GLM) included parametric regressors for trial-wise unsigned changes in feedback scores and the scalar variable ΔV^n denoting changes in keystroke dynamics from the current to the next trial[68]. A discrete regressor was included for feedback onset (see **Methods**). Alternative GLM models using different score representations (absolute graded scores, signed score differences) were discarded due to regressor collinearity, which risked model misspecification (**Methods**).

Theta-band activity significantly increased more following punishment than reward feedback (**Figure 4A;** P_{FWER} = 0.021, cluster-based permutation test; **Figure 4B**; N = 39), consistent with previous work[47,49]. This effect emerged between 0.2–0.45 s in frontocentral electrodes. Beyond this feedback-related response, theta activity parametrically tracked unsigned score changes but in opposite directions for reward and punishment: it increased with greater score changes under reward but decreased under punishment, with a significant between-condition difference (P_{FWER} = 0.010; 0.2–1 s; **Figure 4C**). This effect was observed in left frontocentral and right centroparietal electrodes (**Figure 4D**). Additionally, theta amplitude reflected upcoming motor adjustments in a reinforcement-dependent manner: it increased with greater dynamics changes ΔV^n in the next trial under reward but decreased under punishment, with a significant difference between conditions (P_{FWER} = 0.009; 0.2–0.9 s; **Figure 4E**), spanning midline frontal and left central electrodes (**Figure 4F**).

Next, Bayesian linear models revealed a credible PA × reinforcement interaction on theta modulation with unsigned changes in graded scores. As PA increased, theta amplitude in the relevant spatiotemporal cluster became more pronounced under reward but was increasingly suppressed under punishment (posterior estimate: –2.17, 95% CrI [–4.04, –0.34]; **Figure 4G)**. Similarly, theta activity related to upcoming changes in keystroke dynamics was modulated by PA interacting with reinforcement: as PA increased, theta was attenuated under punishment but became more elevated under reward (posterior estimate: –2.58, 95% CrI [–4.12, –1.07]; **Figure 4H**).

Dissociating the influence of reward and punishment on categorical and continuous motor decision-making

Experiment 3 examined whether the interaction effects of PA and reinforcement condition on learning rates and motor variability observed in Experiment 1 stemmed from their influence on participants' categorical decisions—such as switching between dynamics contours (e.g., U-shape) after reinforcement— or on refining keystroke velocity within the same contour. Additionally, although skilled pianists exhibit high consistency in timing and velocity across renditions[77,78], we investigated whether individual differences in motor noise, potentially modulated by PA, contributed to the observed results[79,80] .

To address these questions, a new cohort of highly trained pianists (N = 36) completed a modified task with separate baseline and reinforcement learning phases. At baseline, participants played two new melodies with either constant or varying dynamics, each across 25 trials (**Figure S10**). These conditions, respectively, allowed us to evaluate unintended variability, reflecting motor noise, and total variability, encompassing both intended (exploratory variability) and unintended (motor noise) components[80] (**Methods**).

During reinforcement learning, participants learned the hidden dynamics of the melodies (**Figure 1**) through reward (0–100) and punishment (–100 to 0) feedback over 100 trials per condition (**Figure 5A**). Each trial began with categorical action selection, where participants used piano keys (C2–F2, left hand) to choose one of four displayed dynamics contours, including the unknown correct one. Their choice represented both their predicted categorical solution and the contour they would perform (**Figure 5B**). They then played the melody, refining the intensity of their dynamics within the chosen contour (**Figure 5C**). This task thus involved a reduced action space. Pianists used reinforcement feedback to adjust both their categorical and continuous decision-making to approach the hidden target dynamics.

We hypothesised that if reward and punishment differentially influenced categorical decisions based on PA levels, this would manifest in the rate of switching between contour options. Conversely, the interaction effect could modulate the refinement of keystroke dynamics within the same contour, reflecting decision-making along a continuous scale. Learning biases might also arise from the combined effects of both categorical and continuous decision-making.

Pianists with higher PA scores achieve more consistent keystroke velocity under instruction

At baseline, unintended variability, measured by the coefficient of variation in keystroke velocity across trials, (CV_{un}, mean: 0.0628 [SEM: 0.004]), was negatively associated with PA scores (Spearman ρ = -0.43, 95% CI: [-0.67, -0.10], P_{FDR} = 0.010, BF_{10} = 6.73, indicating substantial evidence for a correlation; **Figure S11**). This suggests that pianists with higher PA scores were better at maintaining consistent keystroke dynamics across trials when instructed, reflecting reduced motor noise. Conversely, intended variability (total – unintended) showed a negligible correlation with PA (Spearman ρ = -0.04, 95% CrI: [-0.34, 0.28], P_{FDR} = 0.828, BF_{10} = 0.389, anecdotal evidence for the null hypothesis; CV_{in}: 0.1762 [0.001]).

PA and reinforcement condition do not modulate switch rates in categorical decisions.

During reinforcement learning, participants selected the correct dynamics contour 64.03 (0.05)% of the time. Bayesian regression analysis of the switch rate in categorical decisions revealed no modulation by reinforcement condition, PA levels, or their interaction (**Supplementary Materials**). Thus, categorical decisions regarding the successful dynamics contour were not biased by reinforcement type, nor were they influenced by PA or its interaction with reinforcement.

Learning biases reflect the combined effect of categorical and continuous motor decision-making

Bayesian multilevel modelling of scores from the 64% of trials in which participants selected and played the correct contour revealed no consistent interaction between PA category and reinforcement condition on marginal trends (**Figure S12**). No credible fixed effects of PA were observed either. However, refinement of keystroke dynamics within the correct contour approached the target dynamics faster under reward than punishment (marginal trend difference: 0.00204, 95%-HDI [0.00117, 0.0029])**.**

Since decision-making along a continuous scale within a fixed contour category did not account for the PA × reinforcement interaction effects on learning biases in performers, the remaining analyses focused on the full dataset. As in Experiments 1 and 2, model M6 was the best fit (LOO-CV; **Table S7**), confirming good convergence and robust predictive accuracy (**Figure S13**). The model showed a credible effect of trial on increasing average scores (equivalent to a gain of 19 points over 100 trials) and their precision, indicating greater consistency (**Figure 5DE**).

Marginal effects analysis supported that reward increased learning speed more than punishment (0.00121, 95% HDI [0.000525, 0.00187]). A trial x PA x condition interaction was also observed, while PA alone did not influence slopes (**Supplementary Materials**).

Further analysis of the three-way interaction revealed that, contrary to Experiments 1 and 2, learning was faster for reward than punishment at low to medium-high PA levels, with median trend differences decreasing across PA categories (**Figure 5FG**): low: 8.53×10^{-4} (95% HDI $[5.85, 11.13] \times 10^{-4}$); medium: 6.03×10^{-4} ($[3.69, 5.36] \times 10^{-4}$); medium-high: 3.26×10^{-4} ($[0.873, 8.33] \times 10^{-4}$). For the highest PA category, the effect reversed, showing faster learning to avoid punishment (-5.72 $[-8.19, -3.22] \times 10^{-4}$). No consistent interaction effects on marginal medians were observed **(Figure S14).**

Collectively, the findings in Experiment 3 suggest that learning biases arose from the combined effects of categorical and continuous motor decision-making, rather than either component alone. These effects were also better explained by somatic than cognitive components of PA (**Supplementary Materials; Figure S15**). The results remain robust even when accounting for potential modulatory effects of individual baseline levels of intended and unintended variability, neither of which exhibited credible effects on scores in this task (**Supplementary Materials**)**.**

Reinforcement-driven use of motor variability accounts for learning biases in Experiment 3

In Experiment 3 we observed that pianists with higher PA levels learned faster to minimise losses, while those with lower PA levels learned faster to maximise gains—a reversal of the interaction effects observed in Experiments 1 and 2. Based on this, we predicted a similar interaction would modulate the causal relationship between motor variability and performance.

Before testing this prediction, we validated that performance outcomes had a causal influence on motor variability in this task. Statistical matching analysis (**Figure 6A-D**) revealed that low scores, compared to high scores, increased motor variability over the subsequent three trials (P_{FDR} = 0.006; Δ_{dep} = 0.71, CI = [0.62, 0.85]; **Figure 6D**). An additional convergent finding was that larger deviations from target velocity patterns (observing lower scores) were followed by greater reinforcement-related variability (**Figure 6E**).

Complementing these findings, a Bayesian Gaussian linear model of *VarDiff*—changes in variability in keystroke velocity (ΔV^n) post-low minus high scores—identified a credible interaction between PA category and reinforcement condition (26.31, 95% CrI: [11.34, 40.11]; **Figure 6F**). For higher PA levels and under punishment relative to reward, the *relative* change in motor variability following low-outcome trials increased. The effect of punishment on the causal influence of outcomes on motor variability changes consistently increased across PA categories, with the largest effects observed in high PA. By contrast, posterior estimates of VarDiff under reward conditions overlapped with zero, indicating no credible effects.

Reinforcement-sensitive Gaussian Process accounts for increased variability use under punishment in higher PA

Fitting the RSGP to the time series of signed errors (e^n) in Experiment 3 replicated a key finding from Experiment 1: slow autocorrelations span ~11 trials, while reinforcement effects are shorter-lived (~3 trials). In addition, the characteristic length of the autocorrelations kernel was significantly greater than that of the reward-sensitive kernel (P_{FDR} = 0.0002; l_{SE} = 11.21 [1.3]; l_{RS} = 2.88 [0.4]; Δ_{dep} = 0.83, CI = [0.74, 0.90]), and its variance scale larger (P_{FDR} = 0.0004; σ^2_{SE} = 4.66 [0.4]; σ^2_{RS} = 2.94 [0.2]; Δ_{dep} = 0.70, CI = [0.58, 0.78]).

Additionally, Bayesian regression modelling revealed a credible positive effect of PA on σ^2_{RS}, increasing the output scale of the reinforcement-sensitive process with higher PA levels (log scale: 0.16, 95% CrI = [0.03, 0.28]; **Figure 7A**). Models including reinforcement condition or its interaction with PA performed relatively worse (**Supplementary Materials).** No effects were observed on σ^2_{SE}. As in Experiment 1, simulations of the predictive distribution of $\boldsymbol{e}^n$ using individual participant parameters (l_{SE}, l_{RS}, σ^2_{SE}, and σ^2_{RS}) reproduced the U-shaped relationship between e^{n-1} and $\sigma(e^n)$, and the linear increase of $\mu(e^n)$ with e^{n-1} (**Figure 7B).**

Last, the association between motor variability, $\sigma(e^n)$, and the unsigned error in the previous trial, $|e^{n-1}|$, was better captured by an exponential model, as in Experiment 1 (**Supplementary Materials; Figure 7C**). Posterior estimates for the exponential coefficients were positive: b_1 = 1.18 (95% CrI = [0.85, 1.50]) and b_2 = 0.26 (95% CrI = [0.10, 0.46]). A credible interaction between PA and reinforcement condition modulated b_2, increasing with PA category and under punishment relative to reward, with a posterior estimate of 0.06 (95% CrI = [0.01, 0.19]; **Figure 7D**).

These findings indicate that higher PA enhanced the exponential growth rate of $\sigma(e^n)$ under punishment compared to reward, reflecting increased sensitivity to larger errors (lower scores) from the previous trial and greater behavioural adaptation through increased variability.

Discussion

Across three experiments, we investigated learning biases from reward and punishment in skilled performers as a function of PA. Using tasks designed to preserve key features of skilled performance, we examined reinforcement learning in highly trained pianists. Contrary to our hypothesis, in Experiment 1, pianists with heightened PA levels learned faster from reward, while those with lower PA relied more on punishment feedback. This interaction was explained by the regulation of motor variability in keystroke dynamics—the variable tied to reinforcement—where variability increased following lower feedback scores. A second, independent experiment replicated these findings.

At the neural level, changes in EEG theta (4–7 Hz) oscillations during feedback processing paralleled the effects of PA on reinforcement learning in Experiment 1. Theta activity encoded unsigned differences in graded feedback and signalled upcoming motor variability regulation, showing greater amplitude changes under reward—consistent with the direction of the learning biases.

In a final experiment, we reduced the action space by presenting participants with four potential dynamics contour solutions, allowing us to assess categorical and continuous motor decision-making. Learning biases were not explained by either decision-making component alone but rather by their combined effect. The reduced action space was associated with lower task uncertainty, as participants had increased information about the hidden target solution[81]. Notably, in this setting, the interaction between PA and reinforcement condition reversed. These patterns were driven by the causal effect of outcomes on motor variability, with higher-PA individuals showing greater adaptation to poor outcomes through increased variability under punishment. These findings collectively suggest that skilled performers with higher predisposition to PA learn faster from punishment in low-uncertainty environments but increasingly rely on reward as uncertainty escalates.

The results align with prior work showing that reinforcement learning is modulated by anxiety, both in clinical and subclinical populations[17,22-24,26,82](but see refs[83,84]). Trait anxiety has been associated with faster learning in volatile environments and improved inference of hidden states during information-seeking[26,,82]. By contrast, state anxiety is associated with reduced learning[24,26] (but see ref[85]), reflecting overly precise beliefs about action-outcome contingencies and attenuated belief updating[24,27]. When directly comparing reward and punishment, mood and anxiety disorders exhibit elevated punishment learning rates, potentially reinforcing negative affective biases[17,18].

While Experiments 1 and 2 did not link higher PA to negative learning biases, they align with evidence that anxiety subcomponents differentially affect learning from reward and punishment. Somatic anxiety has been associated with increased safety learning, while cognitive anxiety enhances threat learning[21]. Our findings indicate that faster reward learning at higher PA levels was better explained by somatic PA—that is, a predisposition to heightened physiological responses to an impending performance, akin to a physical threat.

In Experiment 3, somatic anxiety again better explained learning biases, but in the opposite direction: higher PA was associated with faster learning from punishment. This likely reflects the reduced action space and task uncertainty. These findings extend evidence on how anxiety dimensions influence learning[21], suggesting that PA—particularly its somatic component—enhances reward learning under high task uncertainty but shifts toward punishment learning when uncertainty is low. Given that intolerance of uncertainty (IU) is a hallmark of anxiety[15,86], this shift may reflect an adaptation to perceived uncertainty. We propose that higher-PA individuals perceive highly uncertain contexts—Experiments 1–2—as aversive, increasing their reliance on reward to navigate uncertainty and achieve their performance goal.

The observed learning biases were partly explained by the effects of PA and reinforcement on the regulation of motor variability. Across tasks, keystroke velocity was actively modulated by recent reinforcement history, increasing when outcomes were low—consistent with evidence that low-reward rates trigger behavioural adjustments[27,35,80,87]. However, establishing a causal link between reinforcement and motor variability is challenging, as motor performance exhibits persistent correlated variation across trials[40,69,70,72,73], which can inflate estimates of motor variability regulation[36].

To address these biases, we isolated trials where surrounding reinforcement values—and thus performance —were comparable across low- and high-score conditions. In Experiment 1, higher PA levels were associated with greater motor variability regulation under reward, but blunted regulation under punishment. The reverse pattern in Experiment 3 confirmed that learning biases aligned with motor variability regulation in the expected direction.

Further analysis using a validated generative model[39] confirmed that motor variability, quantified as the standard deviation of a complementary variable—error (the deviation between produced and target dynamics)—was regulated by reinforcement integration over 3–4 trials, while autocorrelations persisted across 11–12 trials, aligning with prior findings[39]. Crucially, the latent contribution of reinforcement-sensitive variability (σ_{RS}^2) decreased with PA in Experiment 1 but increased in Experiment 3. These results confirm that reinforcement-sensitive scaling of motor variability, likely reflecting exploratory variability, varied with PA, aligning with the punishment-minus-reward difference observed in the marginal trends. Tentatively, these results suggest that punishment may exert more salient or consistent effects on the reinforcement-sensitive scaling of motor variability.

Moreover, error deviations in keystroke dynamics led to increased error variability in the next trial, following an exponential function. The sensitivity of this growth was modulated by the interaction of PA and reinforcement condition, with the expected directional patterns across experiments. These findings extend the statistical matching analysis, confirming that PA influenced whether individuals relied more on graded reward or punishment feedback when adjusting performance trial by trial.

Notably, the learning biases were not explained by baseline levels of intended or unintended motor variability, despite an observed association between higher PA levels and reduced baseline motor noise. This supports the idea that reinforcement-dependent adaptation in skilled performance settings depends on the contextual modulation of variability sources rather than individual baseline levels.

Our findings extend research on anxiety-related exploration from discrete decisions to continuous action spaces. While trait/cognitive anxiety may enhance directed exploration for information seeking, state/somatic anxiety may diminish it[88–90]. In continuous spaces, reinforcement learning (RL) approaches to the exploration-exploitation dilemma propose uncertainty-aware critics to guide exploration[91]—particularly relevant to anxiety, where IU and misestimation of uncertainty are central features[15,86]. Deep RL offers an alternative[92], using added noise to network weights to facilitate exploration akin to random exploration strategies. Future work should examine whether uncertainty-aware or deep RL mechanisms underlie learning biases in anxiety, particularly in continuous action spaces.

Reward and punishment learning were dissociated in the amplitude of theta-band oscillations. Using a linear convolution model[76], we assessed trial-by-trial modulation of time–frequency EEG responses by each explanatory regressor while controlling for the others. Theta activity increased 0.2–0.5 s post-feedback at a small midfrontal cluster in response to punishment relative to reward, consistent with prior work[47,49]; however, this effect was unrelated to learning performance.

Instead, a distinct theta pattern scaled parametrically with unsigned feedback changes in a reinforcement-dependent manner. During reward processing, greater left frontocentral and right centroparietal theta emerged with increasing PA, with dampened effects under punishment at 0.2–1 s. Concurrently, theta amplitude in left central and midline frontal sites at 0.2–0.9 s predicted upcoming keystroke velocity changes, increasing under reward and decreasing under punishment. This spatiotemporal dissociation mirrored the behavioural and computational learning biases, reflecting an interaction between PA and reinforcement.

These findings converge with evidence that midfrontal theta encodes unsigned prediction errors and signals the need for behavioural adjustments. Our findings further align with a proposed role for theta in configuring prefrontal control networks[47,49], and with elevated frontal-midline theta in high-anxiety individuals, particularly following performance errors requiring adaptation[51]. Theta has also been shown to synchronise prefrontal and motor regions during decision-making[51], potentially guiding action execution; in our task, this may underlie the motor adjustment effect extending to contralateral sensorimotor electrodes.

We did not model trial-wise prediction errors (PE) and cannot directly map unsigned score changes to PEs within RL or Bayesian frameworks. This may account for the absence of beta-band effects in our GLM, despite evidence linking feedback-related beta suppression to updating motor predictions[42,45,46,93]—a process dysregulated in anxiety[27]. This highlights the need for models capturing hidden-state inference via PEs in continuous action spaces to clarify the role of beta in PA-related performance updating.

To further elucidate the neural circuitry, source analysis is needed to identify whether ACC and PFC activity underlies the observed theta effects, given their roles in behavioural control[52,94] and learning under uncertainty[14]—including via modulation of midline-frontal theta[51,52]—and their consistent functional alterations in anxiety[12,56]. The striatum, through its role in reward PE encoding and renforcement-based motor learning[58,60], other basal ganglia nuclei and thalamus, implicated in motor variability regulation[39,59] should also be examined to dissociate cortical and subcortical contributions to PA-related learning biases. A key hypothesis for future work is that interactions between PFC, ACC, and motor circuits modulate learning biases in skilled performers, and may be altered in high-stakes contexts that trigger PA, potentially contributing to performance breakdowns.

While the inclusion of multiple experiments in expert performers strengthens the robustness and validity of our findings, the focus on pianists limits generalisability to other expert populations. Future studies should examine whether similar effects emerge in other high-performance domains where PA is prevalent. Additionally, while we assessed learning biases as a function of trait predisposition to PA, it is crucial to investigate how these biases manifest under experimentally induced PA, particularly in relation to different forms of uncertainty. Given that self-efficacy—an individual's belief in their ability to meet performance demands—is a key determinant of achievement in skilled performance[6,95,96], future research on PA should incorporate this factor, alongside IU metrics, to better understand contributions to learning biases in performers.

In sum, our findings demonstrate that predisposition to PA in skilled individuals modulates learning from reward and punishment in a context-dependent manner. As uncertainty increases, faster learning shifts from punishment to reward, driven by the active regulation of motor variability and modulation of theta-band activity. These findings offer important insights into how trait PA shapes learning biases and identify mechanisms relevant to understanding how high-stakes settings that induce PA may further impair expert performance.

Materials and Methods

Experiment 1. Differential learning from reward and punishment in skilled performers as a function of performance anxiety

Demographics. Forty-two participants were recruited for this study, aiming for a sample size of 40 to achieve the desired power level (simulation-based Bayesian power analysis: **Supplementary Materials)**. One participant was excluded for not adhering to the task procedure. The final sample (N = 41, 23 females, 18 males; age range: 18–66, M = 29.2, SEM = 2; 32 self-reported right-handed, 8 left-handed) comprised pianists with at least six years of formal piano training, with proficiency in sight-reading sheet music, advanced musical technique, and an understanding of music dynamics. On average participants had 19.16 (SEM 2.0) years of training and performance, and were currently playing an average of 11.95 (SEM 1.9) hours per week.

Participants did not have a history of neurological or psychiatric conditions and were not currently taking medication for anxiety or depression. Due to faulty EEG recording in one participant, the EEG analysis sample consisted of 40 participants (23 females, 17 males).

All participants gave written informed consent, and the study protocol was approved by the local ethics committee at the Department of Psychology, Goldsmiths, University of London. Participants were compensated with £35, with the possibility of increasing this sum up to £45 depending on their task performance. Recruitment was predominantly conducted using flyers on university campuses around London in addition to online posts in local music groups.

The Kenny music performance anxiety (K-MPAI) inventory evaluates cognitive, behavioural, and physiological components commonly associated with MPA and other anxiety disorders[97], demonstrating high consistency across cultures and various musician populations[98]. It consists of 40 items rated on a 7-point Likert scale (0 "strongly disagree" to 6 "strongly agree"). Scores range from 0 to 240, with values above 160 considered above average. Previous factor analysis of the K-MPAI in professional musicians identified subscores associated with proximal somatic anxiety and negative cognitions, which we used to assess the dissociation between somatic and cognitive dimensions of PA on learning in our study[99]. We also administered the trait subscale of the Spielberger State-Trait Anxiety Inventory (STAI Form Y-2[100]; Spielberger, 1983), assessing more generalised anxiety (See **Supplementary Materials)**. Participants completed the questionnaires at the beginning and end of the session, respectively.

Procedure. Upon arrival, participants were seated at a digital piano (Yamaha Digital Piano P-255, London, United) positioned in front of a screen and were given time to familiarise themselves with two short melodies for the right hand, Melody 1 and Melody 2 (**Figure 1A**). Pianists were instructed to use the predetermined finger-to-note mapping indicated on the score sheet and to memorise the melodies. Following a self-paced familiarisation phase, participants practised the two melodies at a tempo of 120 bpm using the digital piano's metronome, for approximately 5 minutes (range 3-10 min). The metronome was initially used to facilitate melody learning at a consistent tempo, minimising temporal variability across participants; however, no metronome clicks were present during the main task. After the practice session, participants were fitted with EEG recording equipment for 30-45 minutes. Before the experimental phase, we checked that participants could play both melodies from memory with the instructed fingering (five consecutive error-free melody renditions) at the recommended tempo. See **Figure 1B**.

The main task consisted of playing the two melodies with the aim of discovering their hidden target dynamics, encoded as a specific pattern of keystroke velocity values for each melody. We chose melody dynamics as the target variable because the rendition of dynamics can vary widely among performers, introducing a degree of ambiguity in how they are executed during musical performances. The specific target dynamics deviated from the melodies' conventional phrasing, ensuring that the intended solution was not the most natural choice for a pianist. Pianists were informed that the target dynamics would differ from those expected based on their musical training and were encouraged to explore different dynamics guided by the feedback scores. They were visually presented with examples of potential target dynamics for the melodies, illustrating prototypical contours (e.g., crescendo, diminuendo, or mixed shapes) to convey that hidden target solutions could be approximated through simple parametric forms (**Figure S1)**.

Participants completed 100 trials per reinforcement condition, each associated with one melody, split into two blocks of 50 trials. After each trial, they received graded performance feedback in the form of a numerical score (**Figure 1B**), reflecting the difference between the target MIDI keystroke velocity and their own keystroke velocity patterns. MIDI velocity corresponds to key press intensity or loudness. Trial scores were computed using an exponential decay function applied to the square root of the summed absolute velocity differences between the participant's MIDI keystrokes and the target dynamics, adjusted to a predetermined scale (0-100 for reward and -100 to 0 for punishment conditions) based on pilot data and simulations. MIDI velocity values ranged from 0–127, and the digital piano volume was set to a medium level for consistency across sessions.

In the reward condition, scores were displayed as monetary gains (0–100), with 100 indicating that the performed dynamics were identical to the target dynamics, and a maximum total reward of £5. In the punishment condition, scores were presented as losses from an initial £5, ranging from -100 to 0, with 0 indicating target dynamics. The scoring formula was identical for both conditions, but in the punishment condition, 100 was subtracted from the trialwise score. Pianists were instructed to either maximise gains (reward reinforcement) or minimise losses (punishment reinforcement). To ensure consistency, the researcher followed a scripted protocol when delivering instructions.

Trials containing any pitch errors were classified as incorrect, and participants were informed that such errors would result in the lowest possible score. Trained pianists have demonstrated proficiency in correctly detecting pitch errors in their performances, as evidenced by early error-detection neural signatures[101,102]. Consequently, in our task, we assumed that pianists would correctly attribute the lowest feedback scores to error trials and not use them to update their beliefs about the hidden dynamics. Accordingly, incorrect trials were excluded from the analyses (6.98 [SEM 0.74]% for reward and 7.24 [0.85] % for punishment conditions, respecively; rates were similar between reinforcement conditions, $P = 0.7662$, $\mathrm{BF}_{10} = 0.1768$, providing substantial evidence for the null hypothesis).

Trials were initiated by the participant hitting a designated key with their left index finger. The melody score was briefly displayed on the screen at the start of the trial and disappeared when a visual cue replaced it on the screen, signalling the start of the performance. Participants had 7 seconds to play the melody (**Figure 1B**). After each trial, the feedback was delivered to the participant on the screen in the form of a score and was presented for 2 seconds. The task was run using Visual Basic, and additional parallel port and MIDI libraries. The order of presentation of the melodies and their mapping to reward and punishment feedback was pseudorandomised and counterbalanced across participants.

***Stimulus materials*.** Melody 1 and Melody 2 (**Figure 1A**) were composed specifically for this study. These two short melodies, in a 4/4 time signature, consist of a pattern of 8 quavers repeated twice over 2 bars. To facilitate the task, the dynamic solution was repeated twice per melody—once for the first 8 notes and once for the other 8 notes. Participants were explicitly informed about this. The melodies were designed with the following criteria in mind: (i) they were to be played with the right hand only; (ii) their performance would not present technical challenges and would require minimal shifts in hand or finger positioning to reduce movement artifacts affecting EEG recordings; (iii) the melodies were to be atonal; (iv) they would be presented to the participants with no dynamics indicated (**Figure 1A**); (v) the melody's hidden target dynamics would not be a trained musician's initial guess (**Figure S1**), necessitating exploration of dynamics to infer the solution (**Figure 1C**). The target dynamics (**Figure 1A,C**) disrupted the traditional beat structure of the 4/4 time signature, whereby the first and third beat are strong, and the second and fourth are weak. Crescendos, decrescendos and accents that were included as part of the melodies' target dynamics furthermore clashed with the natural flow and direction of the melody.

***EEG, ECG, and MIDI Recording*.** EEG and ECG signals were recorded using a 64-channel EEG system (ActiveTwo, BioSemi Inc.), following the extended international 10–20 system, placed in an electromagnetically shielded room. During the recording, the data were high-pass filtered at 0.1 Hz. Vertical and horizontal eye movements (EOG) were monitored by electrodes placed above and below the right eye and at the outer canthi of both eyes, respectively. Additional external electrodes were placed on both left and right mastoids to serve as initial references upon importing the data into the analysis software (data were subsequently re-referenced to a common average reference, see below). The ECG was recorded using two external channels with a bipolar ECG lead II configuration: the negative electrode was placed on the chest below the right collarbone, and the positive electrode was placed on the left leg above the hip bone. The sampling frequency was 512 Hz.

As in our previous EEG studies with trained pianists[101,103], participants were instructed to minimise upper body and head movements during trial performance and outcome processing, focusing movement solely on their fingers. This was facilitated by our stimulus material, which was designed to avoid shifts in hand or finger positioning, thereby reducing movement artifacts. Participants were informed that they could briefly move between trials if needed, and that they could initiate the next trial at their own pace.

Performance was recorded as MIDI files using the software Visual Basic and a standard MIDI sequencer program on a PC with Windows XP software (compatible with Visual Basic and the MIDI sequencer libraries we used). To run the behavioural paradigm and record the MIDI data, we used a modified version of the custom-written code in Visual Basic that was employed in similar paradigms in our previous studies[27,103]. This program was also used to send synchronisation signals in the form of transistor–transistor logic (TTL) pulses —corresponding with onsets of visual stimuli, key presses, and feedback scores—to the EEG/ECG acquisition PC.

***Bayesian analysis workflow of performance data*.** Beta regression models were implemented in R (version 4.3.2), using the brms package (version 2.21.0[104,105]). Beta regression is a distributional regression designed for bounded values between 0 and 1 and can be parametrised by the mean (μ) and precision (ϕ)—similar to the inverse variance in a normal distribution. In our study, these parameters describe the distribution of observed scores. To fit the beta regression model, reward scores ranging from 0 to 100 were transformed to the 0–1 interval by dividing by 100. Punishment scores were first shifted by adding 100 and then also divided by 100. The mean μ represents the central tendency of the scores, while the precision ϕ indicates the spread of the scores around this mean. Critically, the observed score distribution naturally avoided the extremes of 0 and 1—a requirement for beta regression. Error trials (which received non-informative scores of 0 in the rescaled 0–1 interval) were excluded from analyses, as described in the *Analysis of Motor Variability* subsection.

To build and evaluate these models, we followed the principled Bayesian workflow proposed by[61,106,107]. This workflow typically consists of the following steps: (1) model building, including model and prior specification, and prior predictive checks; (2) learning or conditioning the model on observed data, which includes convergence diagnostics; (3) evaluating the model fit and the implications of the resulting posterior, which includes model checking (posterior predictive checking) and validation.

Initially, for model building, we started with a minimal model M1, designed to capture the main phenomenon of interest[106]. In our study, this included the main fixed effects of reinforcement condition and trial, and their interaction, on the observed scores. We defined priors on the coefficients and performed several checks to assess the adequacy of the model.

The default priors in brms (and rstan) are intended to be weakly informative, providing moderate regularisation and facilitating stable computation. However, when additional information is available, the selection of more informative priors is encouraged. Such information could include findings from prior research or domain expertise. Based on previous work on positive/negative feedback learning in decision-making tasks[17,21], we defined our prior regressor coefficients for M1 to include: a positive effect of trial progression on μ, denoting improvements over trials; a positive interaction effect between reinforcement condition and trial, accounting for an expected faster learning with punishment than reward reinforcement across trials; no effect of punishment relative to reward reinforcement on μ at baseline. The prior on the intercept for μ was set at 0.2 on the log-odds scale, corresponding to an initial score of 0.55 (transforming log-odds to probability using plogis(); or 55 on the participants' observed scale).
Priors were Gaussian distributions centred at the chosen values (see ***Supplementary Materials***), with a standard deviation, σ, that was consistent across parameters (σ = 0.01), except for the intercept, which included a larger σ to align with its larger scale (σ = 0.1). This model also included priors for precision, ϕ, defined as Gaussian distributions centred at 0 (σ = 0.01), except for the intercept, which was centred at 2.5 (σ = 1) to allow for dispersion of scores at baseline, and for trial, which had a small positive prior mean, as we expected the precision of the beta distribution to increase across trials, reflecting participants performing more consistently close to the target solution.

We next conducted *prior predictive checks* to assess the consequences of our model and the priors, checking they are consistent with our domain expertise[106,107]. This step consists of drawing values from the prior distributions, and simulating hypothetical data using the model—without incorporating any empirical data. This can be done in brms by setting *sample_prior = "only"* when running the model. We selected minimum, maximum, and mean score values as summary statistics to visualise the prior predictive distribution. The distributions for these statistics under the Beta regression model M1 were within a suitable and plausible range (See **Figure S4**), validating our choice of M1 for subsequent data analysis without further modifications.

We constructed models of increasing complexity (Models M2-6, **Table S1**), which included random effects of subjects on the intercept (M2) and, additionally, on the slope (M3). The most complex model included a three-way interaction between condition, trial, and PA, in addition to variation by subjects on the intercept and slope (M6).

Following recommendations by Bürkner and Charpentier (2020)[108] on using ordinal regressors in Bayesian regression models, to assess PA effects, we included a monotonic function for PA categories: levels 1, 2, 3, 4, denoting low to high PA values. This allows for modelling potential nonlinear monotonic effects of PA categories on our dependent variable, scores. The quartile boundaries that split the PA scores into four partitions were 70, 107, and 125 (range in our sample 18-169; with values within 80–120 considered average in the musician population). The T-STAI ranged 20-65, median 40, where the reference median norm for Spielberger trait values is 35, and values above 45 are considered very high as they are commonly observed in clinical anxiety cases[109]. See further details in ***Supplementary Materials***.

In models M3 and M6, for the random effects structure, we included a prior on the correlation between trial-specific effects and subject-specific intercepts. Specifically, we used an LKJ prior with a shape parameter of 2 for the correlation (class = "cor"), which promotes moderate correlations but allows for flexibility. The remaining parameters, related to monotonic effects of PA, were set to default priors. We checked the prior predictive distribution in each model, similarly to M1, which confirmed their adequacy[110].

Models were estimated using 5,000 Monte Carlo Markov Chain (MCMC) samples across 4 chains, totalling 20,000, with the first 1,000 per chain being discarded as warm-up (total of 16,000 posterior samples). Model comparison was performed employing the leave-one-out cross-validation of the posterior log-likelihood (LOO-CV) with Pareto-smoothed importance sampling[65]. The best model was the one associated with the highest expected log point-wise predictive density (ELPD). Moreover, we verified that the absolute mean difference in ELPD (*elpd_diff*) between the two best-fitting models was at least 4 and larger than twice the standard error of the difference (2*se_diff*). If *elpd_diff* was smaller than 2*se_diff*, our criterion was to select the more parsimonious model.

For the best model, we assessed chain convergence using Gelman-Rubin statistics[111,112] (R-hat < 1.01). As an additional convergence diagnostic tool, we evaluated the effective sample size (ESS), which estimates the number of independent samples from the posterior distribution and should exceed 400 for four parallel chains, as recommended by Vehtari et al. (2021)[112]. The ESS was typically above 10,000 in our models. We additionally conducted posterior predictive checks to diagnose potential model misfit[61]. During this step, parameters drawn from the posterior distribution were used to simulate datasets for comparison against the empirical data.

In the winning model, we present the posterior distributions of the most relevant parameters, including posterior point estimates and their corresponding 95% credible intervals (CrI). Full details of all population-level estimates are provided in the corresponding tables in ***Supplementary Materials***, alongside R-hat values. A 95% CrI for the difference between two grouping levels (for instance, between reinforcement conditions, among anxiety levels, or across interactions such as condition*anxiety) that does not encompass zero is interpreted as indicating a credible difference.

To address our central question, whether learning rates (slope) are modulated differently for reward or punishment reinforcement as a function of PA, we used R function *emtrends* (package emmeans, version

1.10.1) to identify marginal effects on the slope considering the three way interaction between condition, trial, and PA category (`mo(anxiety_order)` in R). For linear models, function *emtrends* estimates the mean change in the DV for a unit change in a continuous predictor variable (e.g. trials), adjusted for other predictor variables in the model. For non-linear models such as Bayesian beta regressions, *emtrends* provides effects on the *median* point estimate for slope effects. The marginal effects were transformed from the logit scale, provided by emtrends(), to the response/percentage point scale by adding the regrid = "response" argument.

Complementing the marginal trends analysis, we estimated marginal medians in scores using the *emmeans* function (for simple/main effects) of the emmeans package (***Supplementary Materials***).

Bayesian multilevel modelling analysis of Experiment 2 used the same models and priors.

Analysis of Motor Variability. Given that the velocity vector comprises 16 values (8 values x 2), we integrated this multidimensional information into a scalar variable representing trial-by-trial changes in the velocity pattern using the following formula (see e.g. Banca et al 2023[68]):

$$\Delta V^n = \mathbb{E}[|\mathbf{V^n} - \mathbf{V^{n-1}}|] \tag{1}$$

The expectation operator $\mathbb{E}[\cdot]$ denotes the average across the 16 keystroke positions per trial. Here, ΔV^n represents the normalised sum of absolute differences in keystroke velocity between consecutive trials (*n-1* and *n*) across all 16 positions in the melody. $\boldsymbol{V^n}$ denotes the velocity vector at trial *n*. ΔV^n thus captures the scalar magnitude of change in the multidimensional velocity vector from one trial to the next (**Figure 1C**). Variability was assessed using the variance of ΔV^n values across running windows of five trials, as in Dhawale et al. (2019)[36].

To determine whether participants increased variability following poor outcomes relative to good ones[35,36,68], we analysed motor variability separately for low and high scores (**Figure 2A**). Low and high scores were defined using a median split for each participant and melody.

Behavioural performance, including errors in timing, press angles, and endpoint reaching, exhibits medium-to long-range (persistent) correlated variation extending across hundreds of events[40,69–73]. These previously documented autocorrelations, spanning from ten to several hundred events, can be interpreted as slow memory drifts that contribute to performance fluctuations. However, such autocorrelations confound the assessment of causal relationships between variability and performance[36].

Accordingly, to mitigate the overestimation of motor variability arising from temporal correlations in behaviour[36,39,73], we performed a statistical matching analysis[36]. This approach involved identifying (‘matching’) trials with low and high performance outcomes—termed conditioned trials—that were preceded and followed by similar reinforcement values across score conditions (**Figure 2C**). This previously validated method[36] effectively isolates reinforcement-driven changes in exploratory behaviour from autocorrelation effects, thereby providing a more accurate estimate of the causal influence of reinforcement on subsequent motor variability.

The matching analysis was applied separately for each melody. We then averaged the task-relevant variability in ΔV^n following conditioned trials across melodies, contrasting it between low and high score conditions using paired permutation tests (5,000 permutations). The results demonstrated significantly higher motor variability following low scores compared to high scores at positions +1 to +3 after the conditioned trial (see ***Results***, **Figure 2D**).

Importantly, our analysis was carefully designed to prevent incorrect trials from contaminating or confounding the results. Trials with performance errors received the lowest reinforcement feedback (0 for reward, -100 for punishment), which did not provide information about the target dynamics. Consequently, such error trials were recorded as NaN (MATLAB) in the variability analysis. Any running window of five values that included a performance error was assigned a NaN variance value, ensuring it did not influence the analysis of specific reinforcement-driven variability changes.

To investigate whether learning biases towards reward or punishment were influenced by variations in the causal relationship between reinforcement and motor variability across PA levels, we specifically employed a Bayesian Gaussian linear model. This model included fixed effects of PA category, reinforcement condition, and their interaction to analyse changes in task-relevant variability (*VarDiff*). Specifically, we examined *VarDiff* values from the first three positions following conditioned trials, as these positions exhibited significant variability differences in our statistical matching analysis (**Figure 2D**). See further details in **Supplementary Materials**.

Generative model of behaviour: reinforcement-sensitive Gaussian Process. We employed a reinforcement-sensitive Gaussian Process (RSGP) model to characterise trial-wise variability in keystroke velocity errors as a function of reinforcement history. This generative model extends the reward-sensitive Gaussian Process framework developed by Wang et al. (2020)[39] for analysing variability in reinforcement-based motor tasks. Gaussian Processes are probabilistic models that infer continuous functions from noisy observations by defining a prior distribution over functions, with dependencies between samples (here trials) captured by a covariance function[113,114]. The RSGP uses a composite kernel, combining a standard squared exponential kernel (K_{SE}) to model slow autocorrelations in behaviour with a reinforcement-sensitive kernel (K_{RS}), which modulates trial-wise covariance as a function of reinforcement scores. Specifically, K_{RS} is a squared exponential kernel with zero covariance for unrewarded (here low outcome) trials. This formulation ensures that only rewarded (high outcome) trials contribute to K_{RS}, tightening behavioural coupling to recent successes and enabling increases in observed motor variability following unsuccessful trials. The RSGP also includes a noise term with an identity matrix scaled by σ^2_0 (**Figure 3A; Supplementary Materials**).

The RSGP was fitted to trial-wise error values (e^n), defined as the expectation of the difference between the produced (***V***^{n}) and target (***T***) keystroke velocity vectors, averaged across the 16 keystroke positions for each melody rendition:

$$e^n = \mathbb{E}[\mathbf{V^n} - \mathbf{T}] \qquad (2)$$

We use lowercase e^n to distinguish this signed error metric from the error metric illustrated in **Figure 2E**, which is based on the norm of vector differences and is always positive. The model estimated the variance and mean of e^n on each trial based on prior values and reinforcement history. The kernels were parametrised by characteristic length scales (l_{SE} and l_{RS}) and output scales (σ^2_{SE}, σ^2_{RS}), which were inferred via Bayesian inference[114].

Following Wang et al. (2020)[39], we validated the model through simulations, testing its ability to recover known hyperparameters (**Supplementary Materials**). When fitting the model to empirical data, we estimated l_{SE}, l_{RS}, σ^2_{SE}, σ^2_{RS} separately in each participant and condition to capture individual differences. We used Matlab code for RSGP simulation and model fitting from ref.[39] (https://github.com/wangjing0/RSGP).

To examine the effects of PA and reinforcement condition on model parameters, we implemented Bayesian regression models with fixed effects. Model selection was based on leave-one-out cross-validation (LOO-CV), and parameter credibility was assessed using posterior predictive checks.

See further details in **Supplementary Methods**.

EEG preprocessing and analysis. EEG preprocessing was done in MATLAB R2020b using the toolboxes EEGLAB[115] and FieldTrip[116]. EEGLAB was used to import the files and filter the data, applying a 50 Hz notch filter to remove the power line noise. Data were downsampled to 250 Hz.

For the Independent Component Analysis (ICA), we applied a high-pass filter at 1 Hz to improve ICA decomposition[117]. The data were then segmented into epochs from -2 to 2 seconds locked to the outcome trigger, thereby minimising the presence of any potential movement artifacts that could emerge between trials. During trial performance and outcome processing, participants had been instructed to minimise movements. ICA was run in FieldTrip, using the runICA algorithm[118], which combines the Infomax algorithm[119] with the natural gradient learning rule[120]. After IC decomposition, we applied the resulting ICA weights to the 0.1 Hz-filtered version of the epoched data, as recommended by the EEGLAB developers. Artifacts related to eye blinks, eye movements (saccades), and cardiac artifacts, if present, were removed (3.7 on average, range 2-5).

Manual inspection of the epochs was then performed to remove any remaining artifactual epochs, such as those affected by muscle artifacts (reflected in high-frequency fluctuations[121]). This resulted in a total of 78.6 (SEM 3) and 71.3 (SEM 3) clean epochs left for the analysis of the punishment and reward conditions, respectively. In cases where a channel was faulty throughout the epoch inspection, interpolation was used to replace this channel with the average signals from neighbouring channels. This happened with 1-3 channels from 5 participants. One EEG dataset was excluded due to large muscle artifacts during the feedback presentation interval, leaving N = 39 datasets for analysis.

To model neural EEG responses to feedback scores and motor variability, we used linear convolution models for oscillatory responses[76]. This approach extends the classical general linear model (GLM) from fMRI analysis to time-frequency (TF) data and has been widely applied in EEG and MEG research[122,123]. It enables trial-by-trial assessment of TF response modulation by a specific explanatory regressor while controlling for the effects of other included regressors .

In all convolution analyses, each discrete and parametric regressor was convolved with a 20th-order Fourier basis set (40 basis functions: 20 sines and 20 cosines). This configuration enabled the GLM to resolve TF response modulations up to ~8.7 Hz (20 cycles/2.3 s; ~115 ms). The discrete regressor was modelled by convolving this chosen basis set of functions with delta functions encoding the timing of the feedback events, commonly referred to as stimulus input.

We considered three alternative GLM models, each incorporating three regressors: (i) a discrete regressor marking outcome feedback onset, (ii) a parametric regressor representing keystroke velocity changes from the current to the next trial (scalar variable ΔV^n), and (iii) a parametric regressor capturing graded scores. The three models differed in how scores were represented: (1) graded scores (scaled 0–1 for both reward and punishment conditions), (2) unsigned score differences from the previous to the current trial, or (3) signed score differences from the previous to the current trial. The difference-score models allowed us to assess neural activity related to score changes while simultaneously estimating neural representations of upcoming motor adjustments. In contrast, the graded score model assessed neural encoding of the current score and neural activity anticipating future keystroke adjustments.

To ensure GLM model robustness and avoid misspecification, we first assessed collinearity between regressors. Graded scores were highly correlated with upcoming motor variability (Pearson *R*: –0.1 to –0.8, significant in N = 36 participants after FDR correction). Similarly, moderate collinearity was observed between signed score differences and ΔV^n (*R*: –0.2 to –0.45, significant in N = 18). However, unsigned score changes showed minimal correlation with ΔV^n (only N = 2 for punishment and N = 6 for reward showed significant associations after FDR correction, R: 0.2 to 0.5). Based on these findings, we selected unsigned score change as the optimal regressor to pair with ΔV^n in our convolution model, alongside the discrete regressor for feedback onset.

The selected GLM was applied to concatenated epochs spanning -0.5 to 1.5 s around the feedback event, using Morlet wavelets for time-frequency (TF) analysis in 4–30Hz, thus covering the theta, alpha and beta ranges. We conducted this analysis using SPM12 software (http://www.fil.ion.ucl.ac.uk/spm/), adapting original code by ref.[122], as used in[26,124].

Statistical analysis of sensor-level time-frequency images used cluster-based permutation testing in the FieldTrip Toolbox[116,125] (1000 permutations). We averaged TF activity across frequency bins within each band (theta, alpha, beta). Temporal intervals of interest for statistical analyses were selected based on previous research[27,47,48,124]: 0.2–1.5 s for parametric regressors, 0.1–0.6 s for the feedback onset regressor. We controlled the family-wise error rate (FWER) at 0.05 (two-sided tests, effects considered if $P_{FWER} < 0.025$).

Statistical analysis. Complementing the Bayesian multilevel and non-nested models in our study, when assessing within-subject differences in a variable (e.g., variability estimates between trials of low and high scores), we implemented paired sign permutation tests with 5,000 permutations. In those cases, we additionally provided a non-parametric effect size estimator, the probability of superiority for dependent samples (Δ_{dep}), which is the proportion of all paired comparisons in which the values for condition B are larger than for condition A[126]. 95% confidence intervals (CI) for Δ_{dep} were estimated with bootstrap methods[127]. To control for multiple comparisons arising from, for example, different permutation tests conducted on neighbouring trials or across several interrelated variables, we implemented the adaptive false discovery rate control at level $q = 0.05$.

Experiment 2. Replication study.

Demographics. A sample of 18 pianists (15 females, 3 males; 17 self-reported right-handed; age range: 18–28, mean age = 21.1, SEM = 0.8) completed the same experimental task as in Experiment 1. Participants undertook the task, which was programmed in Python, at the Sony Computer Science Laboratory (Tokyo), using a KAWAI VPC1 digital piano with keystroke velocity in range 0–127. The same inclusion and exclusion criteria pertaining to Experiment 1 were applied for this experiment.

Written informed consent was obtained from all participants, and the study protocol was approved by the local ethics committee at Sony Corporate, Tokyo. Participants received a monetary remuneration for their participation. They received a fixed amount of 3000 JPY, which could increase by an additional sum of 4000 JPY (2000 JPY for reward, 2000 JPY for punishment conditions) depending on their task performance.

As in Experiment 1, to assess trait aspects of PA, we used the Japanese version of the Kenny MPA Inventory, and the trait subscale of the STAI-Y2.

Bayesian Data Analysis. Analysisis of the evolution of scores over time as a function of PA and reinforcement condition was performed exactly as in Experiment 1. The PA scores in this sample were split into four partitions through the quartile boundary values 114, 131, and 144 (range 84–180; T-STAI scores ranged 39–55).

Experiment 3. Reinforcement effects on categorical and continuous motor decision-making in skilled performers.

This experiment employed a modified version of the paradigm used in Experiments 1 and 2, designed to isolate categorical decision-making from decisions made on a continuous scale in skilled pianists.

Participants. Thirty-six pianists (N= 36, 31 females, 5 males; age range: 19-54, M= 25.83, SD=1.3; all self-reported right-handed) were recruited for this experiment. They had not completed Experiment 2 and were naive to the task setting (Sony CSL, Tokyo). As in Experiments 1 and 2, the inclusion criterion for participants was having a minimum of six years of formal piano training. The exclusion criteria included (i) having a history of neurological or psychiatric conditions, and (ii) currently taking medication for anxiety or depression.

All participants provided written informed consent, and the study protocol received approval from the local ethics committee at Sony CSL, Tokyo. Participants were compensated with 3000 JPY, with the possibility of increasing this sum up to 4000 JPY depending on their task performance. Consistent with Experiments 1 and 2, PA levels were assessed using the Japanese version of the K-MPAI questionnaire, and the trait subscale of the STAI-Y2 was also administered. Participants completed the questionnaires at the start and end of the session, respectively.

Paradigm and procedure. The paradigm comprised a baseline variability assessment phase and a reinforcement learning phase.

Baseline variability assessment phase: This phase consisted of two blocks, each comprising 25 trials where participants were assessed on intended and unintended variability of keystroke velocity while performing two simple melodies (Melody 3 and 4, both in a 4/4 time signature, consisting of a pattern of 8 quavers repeated twice over 2 bars. **Figure S10)**. Initially, participants familiarised themselves with these melodies for 5 minutes to be able to play them from memory. For Melody 3, pianists were instructed to maintain consistent keystroke dynamics across 25 trials, which allowed us to measure unintended variability. They chose the dynamic contour freely for this melody, but had to produce it consistently across trials. For Melody 4, they were instructed to vary dynamics intentionally across the 25 trials, allowing us to assess intended variability. The order of variability conditions was pseudorandomised and counterbalanced across participants.

Reinforcement learning phase: In this phase, participants learned the hidden dynamics of the same two melodies used in Experiments 1–2 through reward (0–100) and punishment (–100 to 0) feedback, as in the earlier experiments, but with additional action selection requirements (**Figure 5AB; Figure 1A**). Participants first familiarised themselves with these melodies, in a 4/4 time signature, consisting of a pattern of 8 quavers repeated twice over 2 bars (16 notes in total; **Figure 5A**). Each trial began with participants selecting from four dynamic contour options, representing different keystroke velocity patterns. Trials started with a grey screen featuring a central plus sign, transitioning to a selection screen where participants had a 3-second window to choose a dynamic contour using designated piano keys (C2: option 1, D2: option 2, E2: option 3, F2: option 4). The chosen contour represented their prediction for the overall shape or pattern of dynamics (e.g., U-shape, inverted U-shape, etc.) they believed matched the hidden target. Failing to choose within this time resulted in a non-valid trial. After selection, participants were instructed to perform the melody with their chosen dynamics contour within approximately 8 seconds. Although participants had to align their keystroke dynamics with the chosen contour, they were instructed to use the reinforcement scores to gradually refine their performance and converge on the hidden target solution. Reward or punishment feedback scores were then presented on the screen for 2 seconds. The trial ended with a red ellipse signaling completion.

The target dynamics selected for each melody in this phase differed from those in Experiments 1 and 2, based on two criteria: (i) the hidden target dynamics for each melody matched one of the patterns and its inverted counterpart, as shown to participants at the start of each trial (patterns 1 and 2 for Melodies 1 and 2, respectively; **Figure 5AB**); and (ii) as in Experiments 1 and 2, the correct dynamics would not align with the pianists' natural choice based on their musical training. Crucially, while participants could infer the correct contour over a few trials, they still needed to use reinforcement (reward or punishment) to refine their performance and maximise scores by approaching the target solution. For example, a pianist could infer that pattern 1 was correct for Melody 1, but they would then need to determine whether, for instance, a sequence of keystroke values [60 55 50 45 65 70 75 80] (repeated twice) or [70 67 64 61 76 79 82 85] (repeated twice) more closely matched the target solution.

Analysis of baseline variability. Variability in the task-relevant dimension, keystroke velocity, was assessed using the coefficient of variation (CV) of the 25-trial distribution of keystroke velocity values. This index was first calculated for each of the 16 keystrokes of the melody and then averaged across all positions. We separately measured unintended variability (CV_{un}), representing motor noise, and intended variability (CV_{in}), integrating motor noise and exploratory variability.

Bayesian Performance analysis in Experiment 3. Following the analysis of categorical decisions using the switch rate, we used Bayesian multilevel Beta regression models to assess learning across the 64% of trials where participants chose the correct categorical dynamics contour. In addition, we conducted this analysis using the total dataset, including trials where participants chose an alternative contour. In both cases, the models were built as in Experiments 1 and 2. We retained the original priors from Experiment 1 rather than updating them based on the posterior estimates. This decision was driven by our task modifications in Experiment 3, which could render the previous posterior estimates less applicable. The initial priors, validated by prior predictive checks (**Figure S4**), provided a suitable starting point under the altered experimental conditions. The quartile values of PA scores in this sample were 114 (Q1), 131 (median), and 144 (Q3), with a range of 85 to 182. The trait STAI values ranged from 32 to 62.

Analysis of motor variability and RSGP modelling. In Experiment 3, the analysis of reinforcement-driven modulation of motor variability was conducted similarly as for Experiment 1. In addition, the same RSGP modelling approach and analysis as described for Experiment 1 was implemented.

Data availability

Behavioural and EEG data are be publicly available at the Open Science Framework (OSF) repository, https://osf.io/w7y5k/. Analysis code to reproduce the main analyses is publicly available at OSF, https://osf.io/w7y5k/.

References

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. (American Psychiatric Association, 2013).
2. Blöte, A. W., Kint, M. J. W., Miers, A. C. & Westenberg, P. M. The relation between public speaking anxiety and social anxiety: A review. *J. Anxiety Disord.* **23**, 305–313 (2009).
3. Herman, R. & Clark, T. It's not a virus! Reconceptualizing and de-pathologizing music performance anxiety. *Front. Psychol.* **14**, 1194873 (2023).
4. Fernholz, I. *et al.* Performance anxiety in professional musicians: a systematic review on prevalence, risk factors and clinical treatment effects. *Psychol. Med.* **49**, 2287–2306 (2019).
5. Miller, R. & et al. Surgical performance anxiety and wellbeing among surgeons: a cross-sectional study in the United Kingdom. *Annals of surgery* vol. **275.4** (2022).
6. Yoshie, M., Shigemasu, K., Kudo, K. & Ohtsuki, T. Effects of State Anxiety on Music Performance: Relationship between the Revised Competitive State Anxiety Inventory-2 Subscales and Piano Performance. *Music. Sci.* **13**, 55–84 (2009).
7. Gallego, A., McHugh, L., Penttonen, M. & Lappalainen, R. Measuring Public Speaking Anxiety: Self-report, behavioral, and physiological. *Behav. Modif.* **46**, 782–798 (2022).
8. Tanguy, G. *et al.* Anxiety and Psycho-Physiological Stress Response to Competitive Sport Exercise. *Front. Psychol.* **9**, 14-69 (2018).
9. Balyan, K. Y., Tok, S., Tatar, A., Binboga, E. & Balyan, M. The Relationship Among Personality, Cognitive Anxiety, Somatic Anxiety, Physiological Arousal, and Performance in Male Athletes. *J. Clin. Sport Psychol.* **10**, 48–58 (2016).
10. Chanwimalueang, T. *et al.* Stage call: Cardiovascular reactivity to audition stress in musicians. *PLOS ONE* **12**, (2017).
11. Beilock, S. L. & DeCaro, M. S. From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *J. Exp. Psychol. Learn. Mem. Cogn.* **33**, 983–998 (2007).
12. Ganesh, G., Minamoto, T. & Haruno, M. Activity in the dorsal ACC causes deterioration of sequential motor performance due to anxiety. *Nat. Commun.* **10**, 42-87 (2019).
13. Ioannou, C. I., Furuya, S. & Altenmüller, E. The impact of stress on motor performance in skilled musicians suffering from focal dystonia: Physiological and psychological characteristics. *Neuropsychologia* **85**, 226–236 (2016).
14. Grupe, D. W. & Nitschke, J. B. Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nat. Rev. Neurosci.* **14**, 488–501 (2013).

15. Pulcu, E. & Browning, M. The Misestimation of Uncertainty in Affective Disorders. *Trends Cogn. Sci.* **23**, 865–875 (2019).

16. Bishop, S. J. & Gagne, C. Anxiety, Depression, and Decision Making: A Computational Perspective. *Annu. Rev. Neurosci.* **41**, 371–388 (2018).

17. Aylward, J. *et al.* Altered learning under uncertainty in unmedicated mood and anxiety disorders. *Nat. Hum. Behav.* **3**, 1116–1123 (2019).

18. Pike, A. C. & Robinson, O. J. Reinforcement Learning in Patients With Mood and Anxiety Disorders vs Control Individuals: A Systematic Review and Meta-analysis. *JAMA Psychiatry* **79**, 303 (2022).

19. Bublatzky, F., Alpers, G. W. & Pittig, A. From avoidance to approach: The influence of threat-of-shock on reward-based decision making. *Behav. Res. Ther.* **96**, 47–56 (2017).

20. Raymond, J. G., Steele, J. D. & Seriès, P. Modeling Trait Anxiety: From Computational Processes to Personality. *Front. Psychiatry* **8**, (2017).

21. Wise, T. & Dolan, R. J. Associations between aversive learning processes and transdiagnostic psychiatric symptoms in a general population sample. *Nat. Commun.* **11**, 41-79 (2020).

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

23. Huang, H., Thompson, W. & Paulus, M. P. Computational Dysfunctions in Anxiety: Failure to Differentiate Signal From Noise. *Biol. Psychiatry* **82**, 440–446 (2017).

24. Hein, T. P., De Fockert, J. & Ruiz, M. H. State anxiety biases estimates of uncertainty and impairs reward learning in volatile environments. *NeuroImage* **224**, 117424 (2021).

25. Fan, H., Gershman, S. J. & Phelps, E. A. Trait somatic anxiety is associated with reduced directed exploration and underestimation of uncertainty. *Nat. Hum. Behav.* **7**, 102–113 (2022).

26. Hein, T. P. *et al.* Anterior cingulate and medial prefrontal cortex oscillations underlie learning alterations in trait anxiety in humans. *Commun. Biol.* **6**, 271 (2023).

27. Sporn, S., Hein, T. & Herrojo Ruiz, M. Alterations in the amplitude and burst rate of beta oscillations impair reward-dependent motor learning in anxiety. *eLife* **9**, (2020).

28. Galea, J. M., Mallia, E., Rothwell, J. & Diedrichsen, J. The dissociable effects of punishment and reward on motor learning. *Nat. Neurosci.* **18**, 597–602 (2015).

29. Song, Y., Lu, S. & Smiley-Oyen, A. L. Differential motor learning via reward and punishment. *Q. J. Exp. Psychol.* **73**, 249–259 (2020).

30. Wächter, T., Lungu, O. V., Liu, T., Willingham, D. T. & Ashe, J. Differential Effect of Reward and Punishment on Procedural Learning. *J. Neurosci.* **29**, 436–443 (2009).

31. Chen, X., Holland, P. & Galea, J. M. The effects of reward and punishment on motor skill learning. *Curr. Opin. Behav. Sci.* **20**, 83–88 (2018).

32. He, K. *et al.* The Statistical Determinants of the Speed of Motor Learning. *PLOS Comput. Biol.* **12**, (2016).

33. Roth, A. M. *et al.* Punishment Leads to Greater Sensorimotor Learning But Less Movement Variability Compared to Reward. *Neuroscience* **540**, 12–26 (2024).

34. Wu, H. G., Miyamoto, Y. R., Castro, L. N. G., Ölveczky, B. P. & Smith, M. A. Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nat. Neurosci.* **17**, 312–321 (2014).

35. Pekny, S. E., Izawa, J. & Shadmehr, R. Reward-Dependent Modulation of Movement Variability. *J. Neurosci.* **35**, 4015–4024 (2015).

36. Dhawale, A. K., Miyamoto, Y. R., Smith, M. A. & Ölveczky, B. P. Adaptive Regulation of Motor Variability. *Curr. Biol.* **29**, 3551-3562 (2019).

37. Dhawale, A. K., Smith, M. A. & Ölveczky, B. P. The Role of Variability in Motor Learning. *Annu. Rev. Neurosci.* **40**, 479–498 (2017).

38. Cashaback, J. G. A. *et al.* The gradient of the reinforcement landscape influences sensorimotor learning. *PLOS Comput. Biol.* **15**, (2019).

39. Wang, J., Hosseini, E., Meirhaeghe, N., Akkad, A. & Jazayeri, M. Reinforcement regulates timing variability in thalamus. *eLife* **9**, (2020).

40. Herrojo Ruiz, M., Brücke, C., Nikulin, V. V., Schneider, G.-H. & Kühn, A. A. 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 (2014).

41. Bartolo, R. & Merchant, H. β Oscillations Are Linked to the Initiation of Sensory-Cued Movement Sequences and the Internal Guidance of Regular Tapping in the Monkey. *J. Neurosci.* **35**, 4635–4640 (2015).

42. Tan, H., Jenkinson, N. & Brown, P. Dynamic Neural Correlates of Motor Error Monitoring and Adaptation during Trial-to-Trial Learning. *J. Neurosci.* **34**, 5678–5688 (2014).

43. Palmer, C. E., Auksztulewicz, R., Ondobaka, S. & Kilner, J. M. Sensorimotor beta power reflects the precision-weighting afforded to sensory prediction errors. *NeuroImage* **200**, 59–71 (2019).

44. HajiHosseini, A., Rodríguez-Fornells, A. & Marco-Pallarés, J. The role of beta-gamma oscillations in unexpected rewards processing. *NeuroImage* **60**, 1678–1685 (2012).

45. Zabeh, E., Foley, N. C., Jacobs, J. & Gottlieb, J. P. Beta traveling waves in monkey frontal and parietal areas encode recent reward history. *Nat. Commun.* **14**, 5428 (2023).

46. Tan, H., Wade, C. & Brown, P. Post-Movement Beta Activity in Sensorimotor Cortex Indexes Confidence in the Estimations from Internal Models. *J. Neurosci.* **36**, 1516–1528 (2016).

47. Cavanagh, J. F., Frank, M. J., Klein, T. J. & Allen, J. J. B. Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *NeuroImage* **49**, 3198–3209 (2010).

48. Cavanagh, J. F., Figueroa, C. M., Cohen, M. X. & Frank, M. J. Frontal Theta Reflects Uncertainty and Unexpectedness during Exploration and Exploitation. *Cereb. Cortex* **22**, 2575–2586 (2012).

49. Cavanagh, J. F. & Frank, M. J. Frontal theta as a mechanism for cognitive control. *Trends Cogn. Sci.* **18**, 414–421 (2014).

50. Chrastil, E. R. *et al.* Theta oscillations support active exploration in human spatial navigation. *NeuroImage* **262**, 119581 (2022).

51. Cavanagh, J. F. & Shackman, A. J. Frontal midline theta reflects anxiety and cognitive control: Meta-analytic evidence. *J. Physiol.-Paris* **109**, 3–15 (2015).

52. Andreou, C. *et al.* Theta and high-beta networks for feedback processing: a simultaneous EEG–fMRI study in healthy male subjects. *Transl. Psychiatry* **7**, (2017).

53. Algermissen, J., Swart, J. C., Scheeringa, R., Cools, R. & Den Ouden, H. E. M. Prefrontal signals precede striatal signals for biased credit assignment in motivational learning biases. *Nat. Commun.* **15**, 1-9 (2024).

54. Hayden, B. Y., Heilbronner, S. R., Pearson, J. M. & Platt, M. L. Surprise Signals in Anterior Cingulate Cortex: Neuronal Encoding of Unsigned Reward Prediction Errors Driving Adjustment in Behavior. *J. Neurosci.* **31**, 4178–4187 (2011).

55. Attaallah, B. *et al.* The role of the human hippocampus in decision-making under uncertainty. *Nat. Hum. Behav.* **8**, 1366–1382 (2024).

56. Robinson, O. J., Pike, A. C., Cornwell, B. & Grillon, C. The translational neural circuitry of anxiety. *J. Neurol. Neurosurg. Psychiatry* jnnp-2019-321400 (2019).

57. Rouault, M., Drugowitsch, J. & Koechlin, E. Prefrontal mechanisms combining rewards and beliefs in human decision-making. *Nat. Commun.* **10**, 301 (2019).

58. Vassiliadis, P. *et al.* Non-invasive stimulation of the human striatum disrupts reinforcement learning of motor skills. *Nat. Hum. Behav.* **8**, 1581–1598 (2024).

59. Woolley, S. C., Rajan, R., Joshua, M. & Doupe, A. J. Emergence of Context-Dependent Variability across a Basal Ganglia Network. *Neuron* **82**, 208–223 (2014).

60. Blanco-Pozo, M., Akam, T. & Walton, M. E. Dopamine-independent effect of rewards on choices through hidden-state inference. *Nat. Neurosci.* **27**, 286–297 (2024).

61. Gelman, A. *et al.* Bayesian Workflow. Preprint at https://doi.org/10.48550/arXiv.2011.01808 (2020).

62. Kenny, D. T. Kenny Music Performance Anxiety Inventory (K-MPAI) and scoring form. (2016).

63. Figueroa-Zúñiga, J. I., Arellano-Valle, R. B. & Ferrari, S. L. P. Mixed beta regression: A Bayesian perspective. *Comput. Stat. Data Anal.* **61**, 137–147 (2013).

64. Bürkner, P. & Charpentier, E. Modelling monotonic effects of ordinal predictors in Bayesian regression models. *Br. J. Math. Stat. Psychol.* **73**, 420–451 (2020).

65. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).

66. Kenny, D. T. *The Psychology of Music Performance Anxiety.* (Oxford University Press, 2011).

67. Kenny, D. T. The Factor Structure of the Revised Kenny Music Performance Anxiety Inventory. in 37–41 (Utrecht: Association Européenne des Conservatoires, 2009).

68. Banca, P. *et al.* Action sequence learning, habits, and automaticity in obsessive-compulsive disorder. *eLife* **12**, (2024).

69. Gilden, D. L., Thornton, T. & Mallon, M. W. 1/f Noise in Human Cognition. (1995).

70. Hausdorff, J. M., Peng, C. K., Ladin, Z., Wei, J. Y. & Goldberger, A. L. Is walking a random walk? Evidence for long-range correlations in stride interval of human gait. *J. Appl. Physiol.* **78**, 349–358 (1995).

71. Chen, Y., Ding, M. & Kelso, J. A. S. Long Memory Processes (1/ f^{α} Type) in Human Coordination. *Phys. Rev. Lett.* **79**, (1997).

72. Hennig, H. *et al.* The Nature and Perception of Fluctuations in Human Musical Rhythms. *PLoS ONE* **6**, (2011).

73. Chaisanguanthum, K. S., Shen, H. H. & Sabes, P. N. Motor Variability Arises from a Slow Random Walk in Neural State. *J. Neurosci.* **34**, 12071–12080 (2014).

74. Shmuelof, L., Krakauer, J. W. & Mazzoni, P. How is a motor skill learned? Change and invariance at the levels of task success and trajectory control. *J. Neurophysiol.* **108**, 578–594 (2012).

75. McDougle, S. D. *et al.* Credit assignment in movement-dependent reinforcement learning. *Proc. Natl. Acad. Sci.* **113**, 6797–6802 (2016).

76. Litvak, V., Jha, A., Flandin, G. & Friston, K. Convolution models for induced electromagnetic responses. *NeuroImage* **64**, 388–398 (2013).

77. Furuya, S., Klaus, M., Nitsche, M. A., Paulus, W. & Altenmüller, E. Ceiling Effects Prevent Further Improvement of Transcranial Stimulation in Skilled Musicians. *J. Neurosci.* **34**, 13834–13839 (2014).

78. Hirano, M., Sakurada, M. & Furuya, S. Overcoming the ceiling effects of experts’ motor expertise through active haptic training. *Sci. Adv.* **6**, (2020).

79. Cohen, R. G. & Sternad, D. Variability in motor learning: relocating, channeling and reducing noise. *Exp. Brain Res.* **193**, 69–83 (2009).

80. Chen, X., Mohr, K. & Galea, J. M. Predicting explorative motor learning using decision-making and motor noise. *PLOS Comput. Biol.* **13**, (2017).

81. Walker, E. Y. *et al.* Studying the neural representations of uncertainty. *Nat. Neurosci.* **26**, 1857–1867 (2023).

82. Zika, O., Wiech, K., Reinecke, A., Browning, M. & Schuck, N. W. Trait anxiety is associated with hidden state inference during aversive reversal learning. *Nat. Commun.* **14**, 4203 (2023).

83. Gillan, C. M. *et al.* Experimentally induced and real-world anxiety have no demonstrable effect on goal-directed behaviour. *Psychol. Med.* **51**, 1467–1478 (2021).

84. Ting, C.-C., Palminteri, S., Lebreton, M. & Engelmann, J. B. The Elusive Effects of Incidental Anxiety on Reinforcement-Learning. *J. Exp. Psychol. Learn. Mem. Cogn.* **48**, 619–642 (2022).

85. Lawson, R. P., Bisby, J., Nord, C. L., Burgess, N. & Rees, G. The Computational, Pharmacological, and Physiological Determinants of Sensory Learning under Uncertainty. *Curr. Biol.* **31**, 163-172 (2021).

86. Carleton, R. N., Collimore, K. C. & Asmundson, G. J. G. "It's not just the judgements—It's that I don't know": Intolerance of uncertainty as a predictor of social anxiety. *J. Anxiety Disord.* **24**, 189–195 (2010).

87. Van Mastrigt, N. M., Smeets, J. B. J. & Van Der Kooij, K. Quantifying exploration in reward-based motor learning. *PLOS ONE* **15**, (2020).

88. Chou, K.-P., Wilson, R. C. & Smith, R. The influence of anxiety on exploration: A review of computational modeling studies. *Neurosci. Biobehav. Rev.* **167**, 105940 (2024).

89. Aberg, K. C., Toren, I. & Paz, R. A neural and behavioral trade-off between value and uncertainty underlies exploratory decisions in normative anxiety. *Mol. Psychiatry* **27**, 1573–1587 (2022).

90. Smith, R. *et al.* Lower Levels of Directed Exploration and Reflective Thinking Are Associated With Greater Anxiety and Depression. *Front. Psychiatry* **12**, 782136 (2022).

91. Likmeta, A., Sacco, M., Metelli, A. M. & Restelli, M. Directed Exploration via Uncertainty-Aware Critics. in (2022).

92. Gershman, S. J. & Ölveczky, B. P. The neurobiology of deep reinforcement learning. *Curr. Biol.* **30**, 629–632 (2020).

93. Weber, J. *et al.* Ramping dynamics and theta oscillations reflect dissociable signatures during rule-guided human behavior. *Nat. Commun.* **15**, 6-37 (2024).

94. Domenech, P., Rheims, S. & Koechlin, E. Neural mechanisms resolving exploitation-exploration dilemmas in the medial prefrontal cortex. *Science* **369**, (2020).

95. Ritchie, L. & Williamon, A. Measuring distinct types of musical self-efficacy. *Psychol. Music* **39**, 328–344 (2011).

96. Đurović, D., Popov, S., Sokić, J., Grujić, S. & Aleksić Veljković, A. Z. Rethinking the role of anxiety and self-efficacy in collective sports achievements. *Primenj. Psihol.* **14**, 103–115 (2021).

97. Kenny, D. T. Music Performance Anxiety: Origins, Phenomenology, Assessment and Treatment. *J. Music Res.* **31**, 51–64 (2006).

98. Kenny, D. T. The Kenny music performance anxiety inventory (K-MPAI): Scale construction, cross-cultural validation, theoretical underpinnings, and diagnostic and therapeutic utility. *Front. Psychol.* **14**, 1143359 (2023).

99. Kenny, D., Driscoll, T. & Ackermann, B. Psychological well-being in professional orchestral musicians in Australia: A descriptive population study. *Psychol. Music* **42**, 210–232 (2012).

100. Spielberger, C. D. State-Trait Anxiety Inventory for Adults (STAI-AD). *APA PsycTests* (1983).

101. Ruiz, M. H., Jabusch, H.-C. & Altenmüller, E. Detecting Wrong Notes in Advance: Neuronal Correlates of Error Monitoring in Pianists. *Cereb. Cortex* **19**, 2625–2639 (2009).

102. Maidhof, C., Rieger, M., Prinz, W. & Koelsch, S. Nobody Is Perfect: ERP Effects Prior to Performance Errors in Musicians Indicate Fast Monitoring Processes. *PLoS ONE* **4**, (2009).

103. Bury, G., García-Huéscar, M., Bhattacharya, J. & Ruiz, M. H. Cardiac afferent activity modulates early neural signature of error detection during skilled performance. *NeuroImage* **199**, 704–717 (2019).

104. Bürkner, P.-C. Advanced Bayesian Multilevel Modeling with the R Package brms. *R J.* **10**, 395 (2018).

105. Bürkner, P.-C. **brms** : An *R* Package for Bayesian Multilevel Models Using *Stan*. *J. Stat. Softw.* **80**, (2017).

106. Schad, D. J., Betancourt, M. & Vasishth, S. Toward a principled Bayesian workflow in cognitive science. *Psychol. Methods* **26**, 103–126 (2021).

107. Gabry, J., Simpson, D., Vehtari, A., Betancourt, M. & Gelman, A. Visualization in Bayesian Workflow. *J. R. Stat. Soc. Ser. A Stat. Soc.* **182**, 389–402 (2019).

108. Bürkner, P. & Charpentier, E. Modelling monotonic effects of ordinal predictors in Bayesian regression models. *Br. J. Math. Stat. Psychol.* **73**, 420–451 (2020).

109. Shadli, S. M. *et al.* Right frontal anxiolytic-sensitive EEG 'theta' rhythm in the stop-signal task is a theory-based anxiety disorder biomarker. *Sci. Rep.* **11**, 19746 (2021).

110. Shiffrin, R. M., Lee, M. D., Kim, W. & Wagenmakers, E. A Survey of Model Evaluation Approaches With a Tutorial on Hierarchical Bayesian Methods. *Cogn. Sci.* **32**, 1248–1284 (2008).

111. Gelman, A. & Rubin, D. B. Inference from Iterative Simulation Using Multiple Sequences. *Statist. Sci.* vol. **7**, 457–472 (1992).

112. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P.-C. Rank-Normalization, Folding, and Localization: An Improved Rˆ for Assessing Convergence of MCMC (with Discussion). *Bayesian Anal.* **16**, (2021).

113. Bishop, C. M. *Pattern Recognition and Machine Learning*. (Springer New York, NY, 2006).

114. Rasmussen, C. E. & Williams, C. K. I. *Gaussian Processes for Machine Learning*. (MIT Press, Cambridge, Mass., 2006).

115. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).

116. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Comput. Intell. Neurosci.* **2011**, 1–9 (2011).

117. Klug, M. & Gramann, K. Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments. *Eur. J. Neurosci.* **54**, 8406–8420 (2021).

118. Makeig, S., Bell, A. J., Jung, T.-P. & Sejnowski, T. J. Independent Component Analysis of Electroencephalographic Data. *MIT Press* (1996).

119. Bell, A. J. & Sejnowski, T. J. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Comput.* **7**, 1129–1159 (1995).

120. Amari, S., Cichocki, A. & Yang, H. H. A New Learning Algorithm for Blind Signal Separation. *Neural Inf. Process. Syst.* **8**, (1996).

121. Muthukumaraswamy, S. D. High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. *Front. Hum. Neurosci.* **7**, (2013).

122. Spitzer, B., Blankenburg, F. & Summerfield, C. Rhythmic gain control during supramodal integration of approximate number. *NeuroImage* **129**, 470–479 (2016).

123. Auksztulewicz, R., Friston, K. J. & Nobre, A. C. Task relevance modulates the behavioural and neural effects of sensory predictions. *PLOS Biol.* **15**, (2017).

124. Hein, T. P. & Herrojo Ruiz, M. State anxiety alters the neural oscillatory correlates of predictions and prediction errors during reward-based learning. *NeuroImage* **249**, 118895 (2022).

125. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).

126. Grissom, R. J. & Kim, J. J. *Effect Sizes for Research: Univariate and Multivariate Applications, Second Edition*. (Routledge, 2012).

127. Ruscio, J. & Mullen, T. Confidence Intervals for the Probability of Superiority Effect Size Measure and the Area Under a Receiver Operating Characteristic Curve. *Multivar. Behav. Res.* **47**, 201–223 (2012).

Acknowledgements

The research was partially funded by the Strategic Research Fund of Goldsmiths University of London (MHR), by the Economic and Social Research Council (ESRC) and the South East Network for Social Sciences (SeNSS) through grant ESRC grant number ES/Y001834/1 (AEH, MHR), and by JST CREST (JPMJCR20D4) and JST Moonshot R&D (JPMJMS2012) (SF, TO, YH). The authors thank Jenny Halseid for support in part of the data collection (Experiment 1, Goldsmiths University of London).

Author contributions

Conceptualization: MHR, LP, SF. Methodology: MHR, AEH, TO, YH, LP. Formal analysis: MHR, AEH. Investigation: AEH, YH, MHR. Data curation: AEH, MHR. Writing original draft/visualization: AEH, MHR. Review and editing: all authors. Supervision: MHR, SF. Funding: MHR, AEH, SF.

Figures

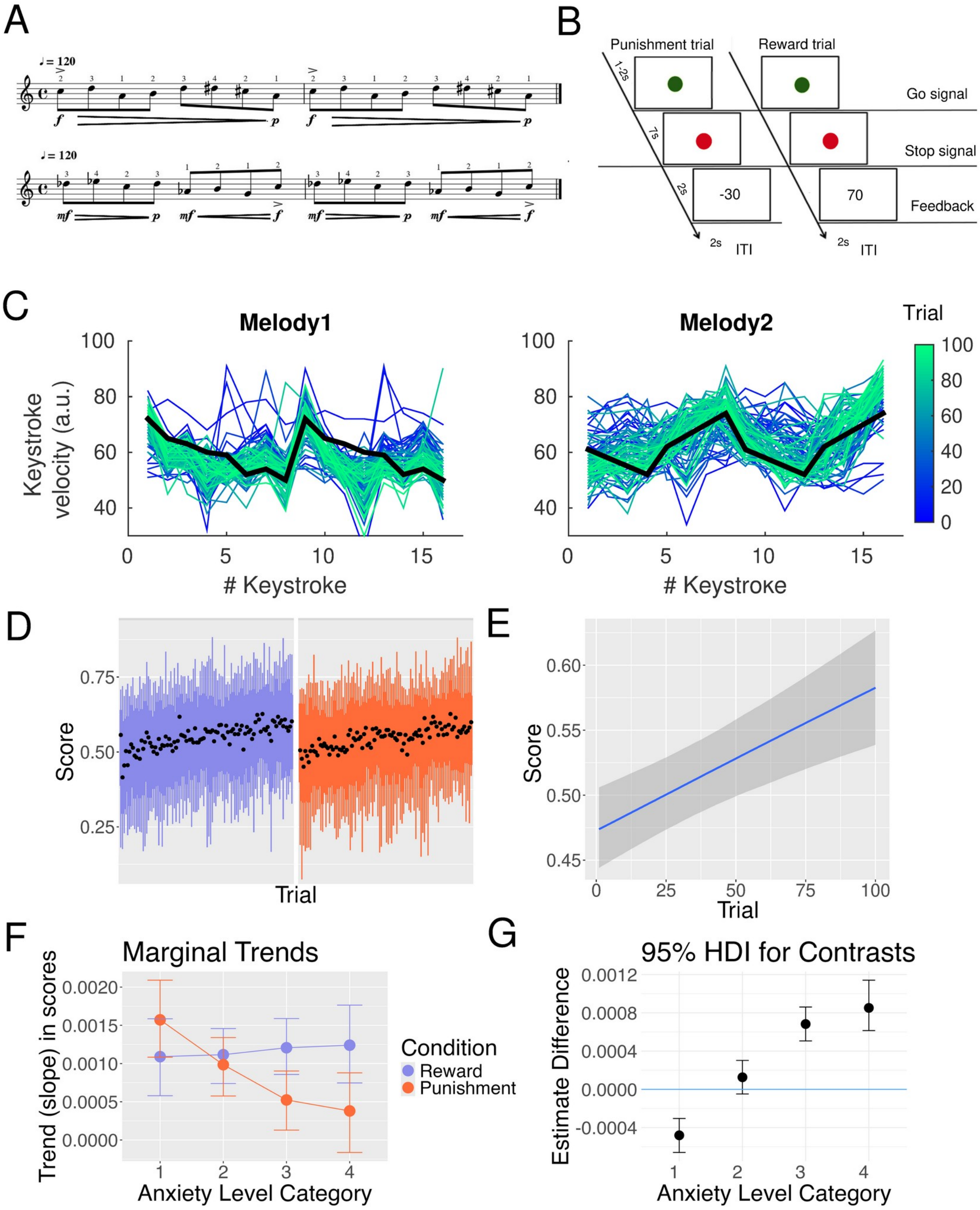

Figure 1 | Task and Performance Analysis for Experiment 1. A. Participants (N = 41 skilled pianists) played two right-hand piano melodies on a digital keyboard, adjusting their keystroke dynamics (intensity of key press) to uncover a hidden target dynamics pattern. Beneath the musical scores, the flow of the target dynamics is represented in musical notation, with *p* denoting *piano*, *f forte*, and *mf mezzo-forte*. **B.** Trial timeline. Each trial yielded graded reinforcement feedback in the form of reward (0–100) or punishment (–

100–0) scores, separately in each reinforcement condition. Feedback was a function of the proximity of performed dynamics to the hidden target. **C.** Example of melody dynamics performed by one participant (graded blue to green for trials 1 to 100), with target dynamics denoted by the bold black line. **D.** Score progression across trials in the group, representing mean with 66% and 95% confidence intervals for the reward (purple) and punishment (orange) conditions. **E.** Bayesian multilevel beta regression modelling revealed a credible positive effect of trial progression on score (posterior median slope = 0.00441; 95% credible interval [0.00243, 0.00651]; log-odds scale), reflecting learning to approach the target dynamics. **F–G.** Marginal trends. A three-way interaction between reinforcement condition, trial, and trait levels of performance anxiety (PA) showed a credible dissociation: lower-PA participants learned faster under punishment than reward (reward minus punishment median slope estimate: -4.81 x 10^{-4}, 95% highest density interval, HDI [-6.60, -3.04] x 10^{-4}). By contrast, medium-high and higher-PA individuals showed steeper learning under reward (median slope difference: 6.83 [5.06, 8.61] x 10^{-4} and 8.52 [6.14, 11.42] x 10^{-4}, respectively). Coloured circles denote median estimates, and shaded intervals indicate 95% HDI of the posterior distribution of median trends for reward and punishment conditions, as well as for the contrast in panel G (reward minus punishment difference estimates).

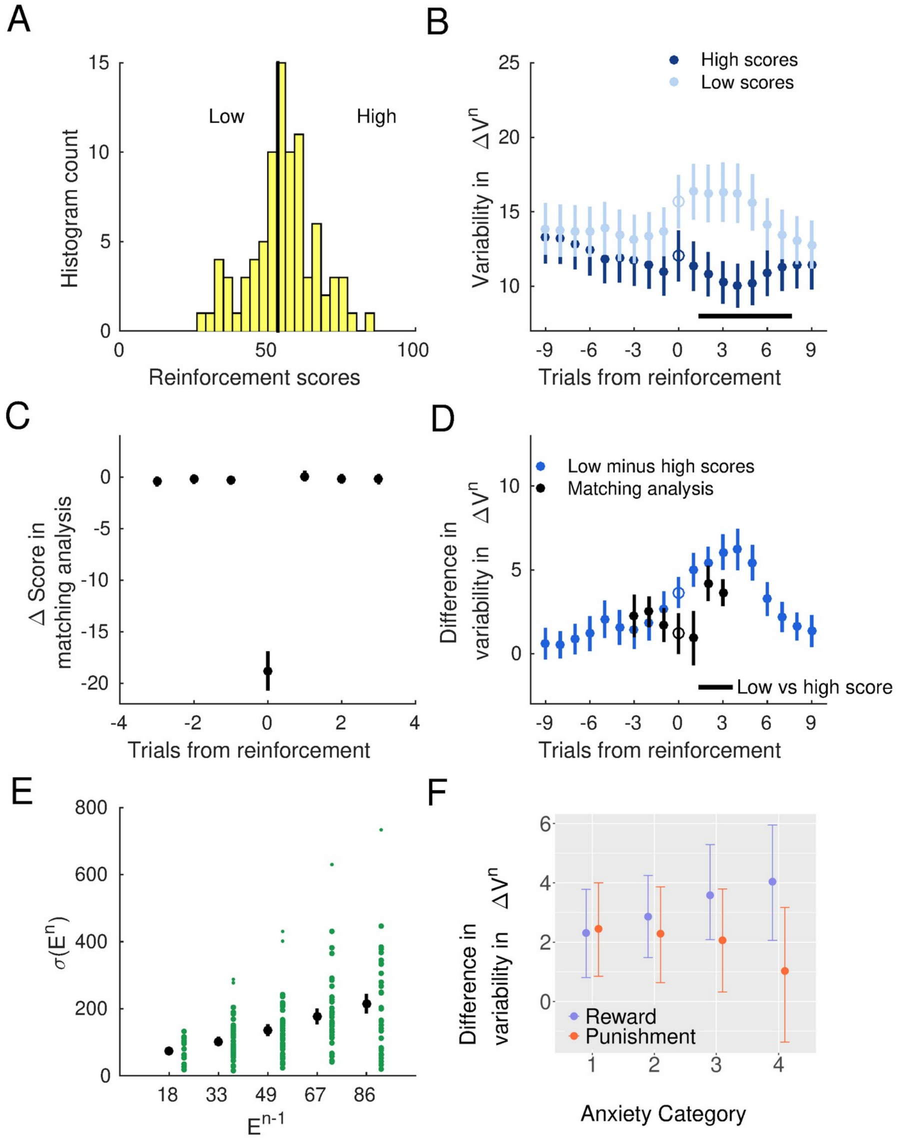

Figure 2 | Reinforcement-related modulation of motor variability in Experiment 1. A. Histogram of reinforcement scores (in one example participant and condition) illustrating the median split used to define low and high score conditions. **B.** Time course of task-relevant motor variability surrounding high (dark blue) and low (light blue) score trials (relative to median split), aligned to the conditioned trial at position 0. Motor variability was assessed using the variance of *ΔVⁿ* values across running windows of five trials, where *ΔVⁿ* represents the normalised sum of absolute differences in keystroke velocity at each of the 16 positions in a melody between consecutive trials (*n–1* to *n*). Variability was significantly higher following low scores compared to high scores at positions +1 to +3 (N = 41; paired permutation test; P_{FDR} = 0.001, significant effects denoted by the horizontal bar at the bottom). Large coloured dots indicate means, with error bars denoting ± SEM. **C.** Statistical matching analysis: we selected trials of low and high scores (median split) that were preceded and followed by similar reinforcement values (and thus performance). The black dots represent the mean performance score difference (SEM) at each trial. **D.** Using trials obtained from the matching analysis (black), we found that motor variability was significantly greater at lags +1 to +3 following conditioned trials (position 0) associated with low compared to high scores (P_{FDR} = 0.0038; dots represent mean, and bars SEM). The difference in uncorrected (all trials) motor variability between low and high scores, not accounting for performance autocorrelations, as shown in panel B, is depicted in blue. **E.** Larger deviations from target velocity patterns—measured as the norm of vector differences and represented as unsigned error E^{n-1}—were followed by greater subsequent reinforcement-related variability, *σ*(E^n). In our task, larger deviations were associated with lower scores. **F.** The difference in motor variability following poor versus good outcomes, labelled *VarDiff* in the main text and obtained from the matching analysis trials, was modulated by the interaction between PA categorical level and reinforcement condition. A negative estimate (-1.06, 95% CrI: [-2.13, -0.01]) indicated that punishment, compared to reward, reduced variability following poor outcomes as PA levels increased. Coloured dots represent posterior point estimates (reward in purple, and punishment in orange) and bars denote 95% CrI.

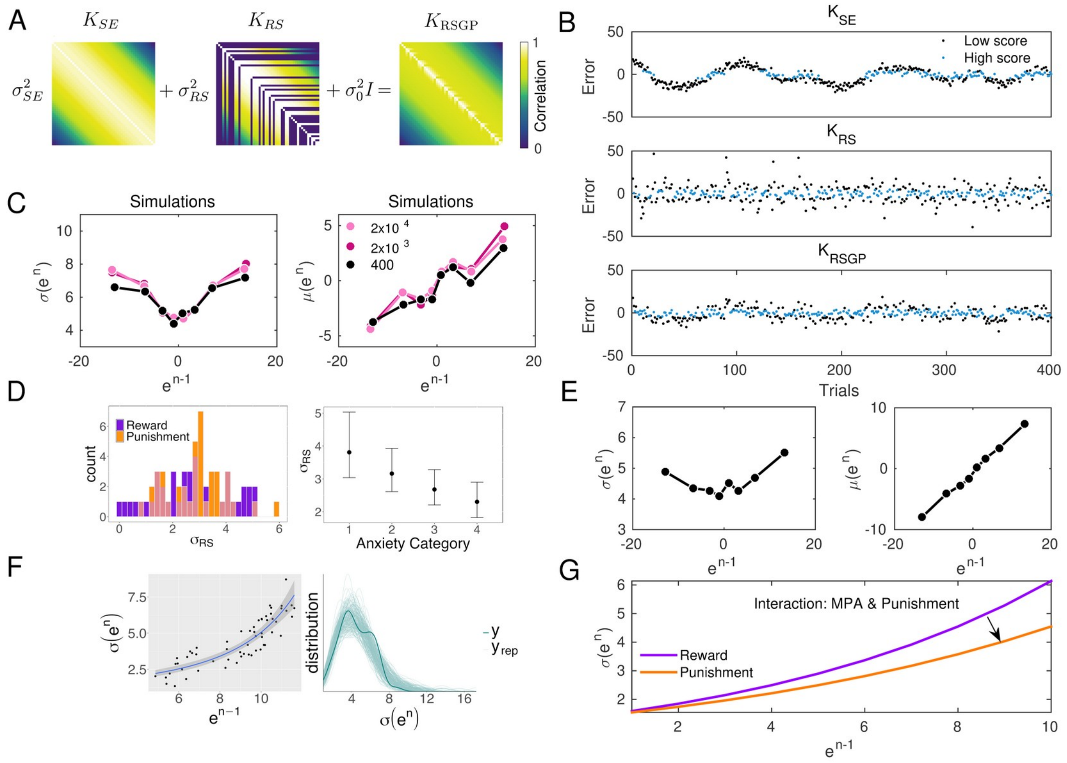

Figure 3 | Reinforcement-sensitive Gaussian process dissociates the effects of autocorrelations and the short-term influences of reinforcement on motor variability in Experiment 1. A. Correlation structures of the squared exponential (K_{SE}), reinforcement-sensitive (K_{RS}), and combined reinforcement-sensitive Gaussian process (K_{RSGP}) kernels used to model the time series of deviations between the produced and target keystroke velocity vectors, the trial-wise signed error e^n. **B.** Simulated trial-wise errors (e^n) under K_{SE}, K_{RS}, and the combined K_{RSGP}. Black dots indicate trials associated with low scores (median split), while blue dots correspond to high-score trials. **C.** Simulations from the generative RSGP model reproduce key empirical patterns from previous work[39]. Left: standard deviation of error ($\sigma(e^n)$) shows a U-shaped dependence on the error on the previous trial (e^{n-1}). Right: mean error ($\mu(e^n)$) increases linearly with e^{n-1}. Coloured lines (dots) reflect the number of samples used in the simulations. **D.** Left: distribution of σ^2_{RS} estimates across participants, split by reinforcement condition (orange for punishment, purple for reward). Right: Bayesian regression (log-normal family) revealed that σ^2_{RS} declined with increasing performance anxiety (PA) category (posterior estimate in log scale: -0.17, 95% CrI = [-0.29, -0.06]). This implies that the latent contribution of reinforcement-sensitive variability to e^n decreased as PA increased. **E.** RSGP fit to empirical data replicates simulation results from C: $\sigma(e^n)$ (left) and $\mu(e^n)$ (right) as functions of e^{n-1}. Mean and SEM are shown; however, SEM values are very small and not visually noticeable. **F.** Left: Exponential fit (line, 95% CrI shaded) to the relationship between $\sigma(e^n)$ and $|e^{n-1}|$. Right: posterior predictive distribution of $\sigma(e^n)$, with individual posterior draws (light green) and empirical data (dark green). **G.** Illustration of the interaction between PA category and reinforcement condition on parameter b_2 in the exponential relationship between $\sigma(e^n)$ and $|e^{n-1}|$. Posterior estimate: b_2 (-0.03, 95% CrI = [-0.07, -0.01]). This revealed that the exponential growth in observed motor variability was attenuated under punishment (orange) relative to reward (purple) as PA levels increased (arrow), suggesting reduced sensitivity to prior error under punishment.

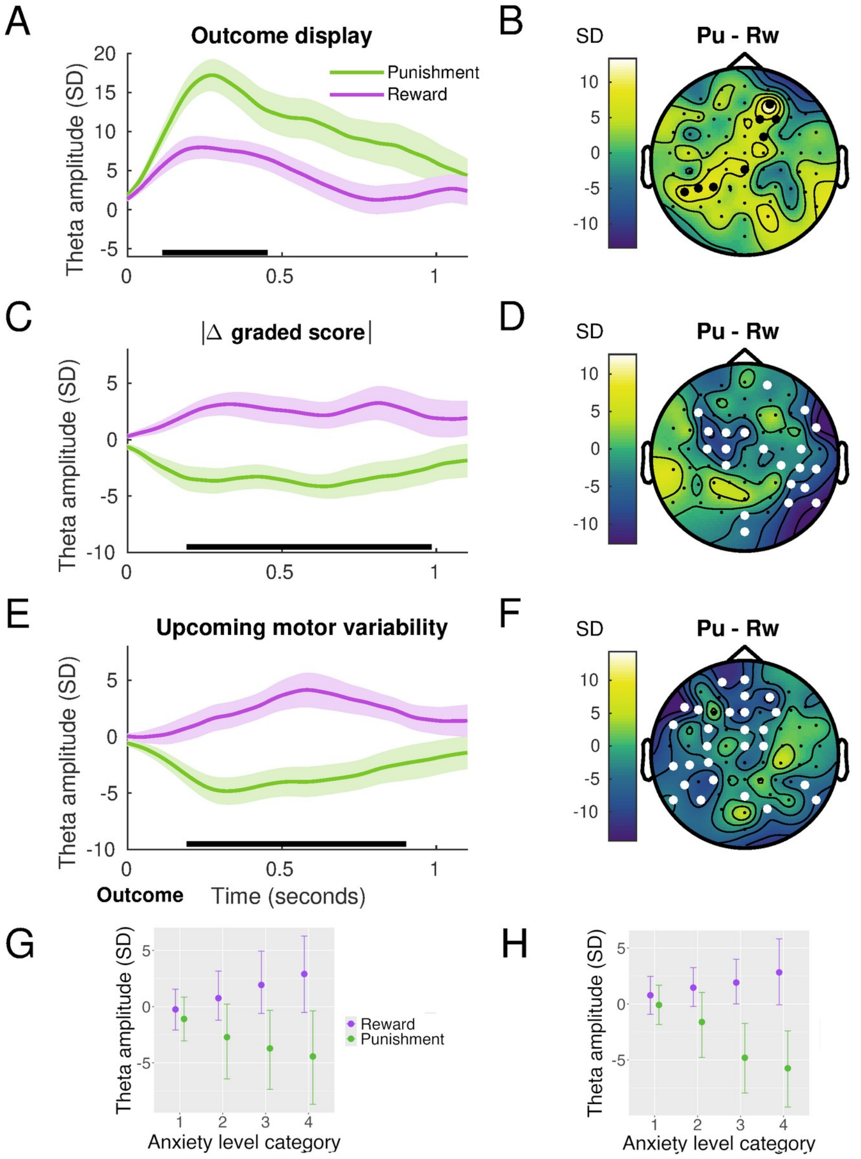

Figure 4 | Theta activity is modulated by unsigned score differences and the amount of upcoming motor adjustments in a reinforcement-dependent manner. A. Time course of feedback-locked theta amplitude (4–7 Hz) in response to outcome onset (reward, magenta; punishment, green), showing greater theta power following punishment than reward between 0.2–0.45 s. Shaded regions denote ±1 SEM; black horizontal bars indicate significant cluster intervals (N = 39; cluster-based permutation test, $P_{FWER} = 0.021$, FWER-corrected). **B.** Topographic map of the difference (punishment – reward) in theta amplitude during the significant time window in panel A, showing a frontocentral distribution (N = 39). **C.** Theta activity parametrically tracked unsigned changes in feedback scores (labelled |Δ graded score|), increasing under reward and decreasing under punishment. A significant condition difference was observed within 0.2–1 s ($P_{FWER} = 0.010$). **D.** Topography of the effect in panel C, with condition differences localised to left frontocentral and right centroparietal electrodes. **E.** Theta amplitude as a function of the amount of upcoming changes in keystroke dynamics (ΔV^n), showing increased amplitude under reward and decreased amplitude under punishment between 0.2–0.9 s ($P_{FWER} = 0.009$). **F.** Spatial distribution of the condition difference in panel E, peaking in midline frontal and left central regions. **G.** Bayesian regression analysis. Posterior estimates of theta amplitude, averaged over the significant spatiotemporal cluster identified in C, revealed a credible interaction between PA category and reinforcement condition in relation to unsigned score changes: with increasing PA, theta increased under reward but was suppressed under punishment (posterior estimate: –2.17, 95% CrI [–4.04, –0.34]). **H.** Same as G, but for theta activity related to upcoming motor adjustments (ΔV^n). Theta amplitude, averaged over the significant cluster identified in F, also showed a PA × reinforcement interaction: as PA increased, theta decreased under punishment but increased under reward (posterior estimate: –2.58, 95% CrI [–4.12, –1.07]). Dots represent posterior means, and bars denote 95% highest density intervals.

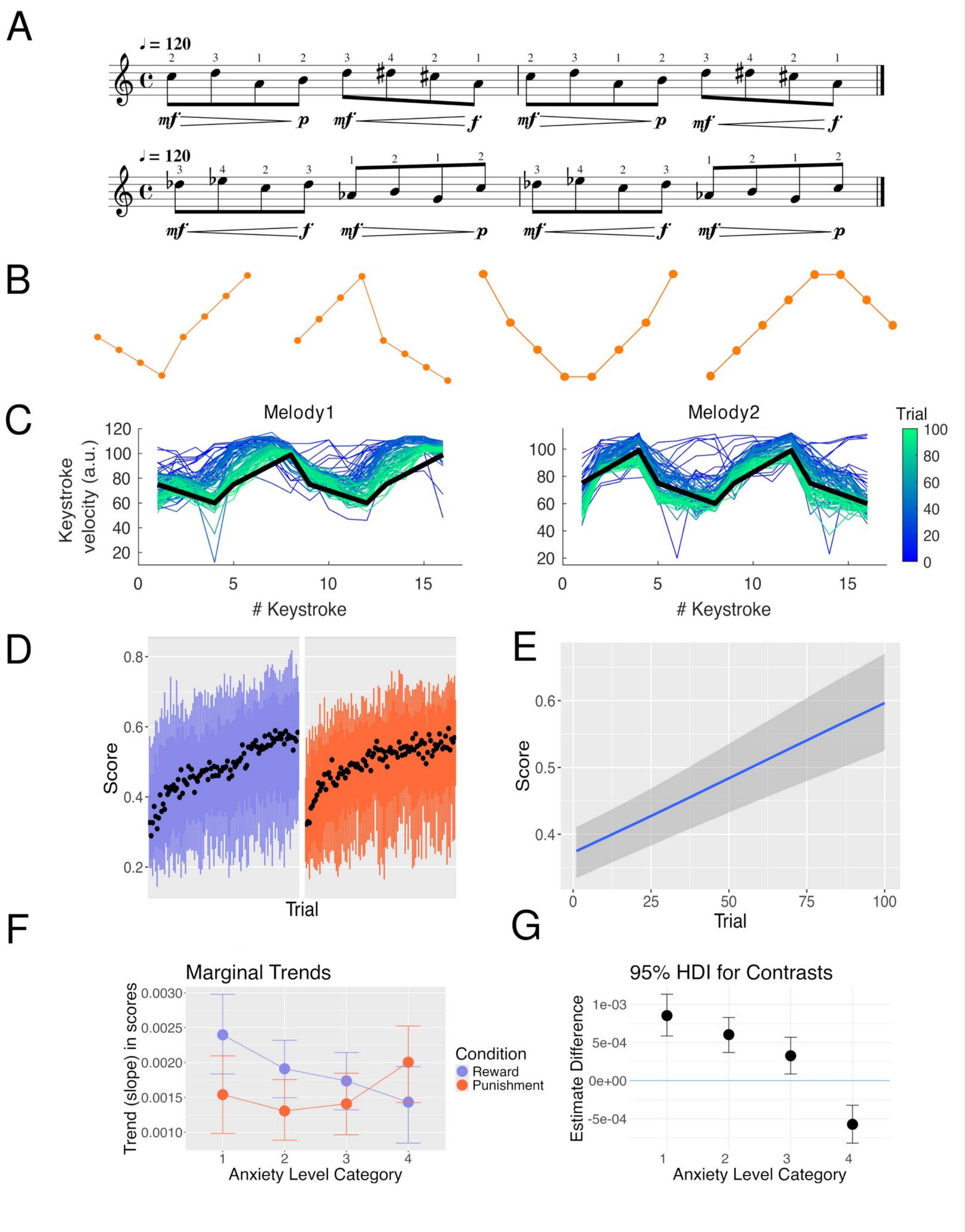

Figure 5 | Task and Performance Analysis for Experiment 3. A. Right-hand melodies for the reinforcement-based performance task (same as in Experiments 1 and 2), each associated with a different hidden target pattern of keystroke velocity values, indicated in musical notation beneath the musical score. **B.** At the start of each trial, participants selected one of four displayed dynamics contours using left-hand piano keys (C2–F2), indicating both their categorical prediction of the correct contour and the one to be executed. The correct contour was 2 for Melody 1 and 1 for Melody 2. **C.** Example performance data from one participant across trials 1–100, showing keystroke velocity profiles for each melody (graded from blue to green) overlaid on the correct target dynamics (bold black line). **D.** Mean feedback scores over trials for the reward (purple) and punishment (orange) conditions, with 66% and 95% confidence intervals. The sample consisted of N = 36 skilled pianists. **E.** The posterior estimate of the overall trial effect on scores revealed a credible learning effect (slope = 0.00963, 95% CrI [0.00737, 0.01202], log-odds scale). **F–G.** Marginal trends. A credible three-way interaction between trait performance anxiety (PA) categorical level, trial, and reinforcement condition was observed. Contrary to Experiments 1–2, reward sped up learning more than punishment at low and medium PA levels. Median slope differences (reward – punishment) decreased across PA categories: low: 8.53 × 10^{-4} (HDI [5.85, 11.13] × 10^{-4}); medium: 6.03 × 10^{-4} ([3.69, 5.36] × 10^{-4}); medium-high: 3.26 × 10^{-4} ([0.873, 8.33] × 10^{-4}). At the highest PA level, learning was faster under punishment (−5.72 × 10^{-4}, [−8.19, −3.22] × 10^{-4}).

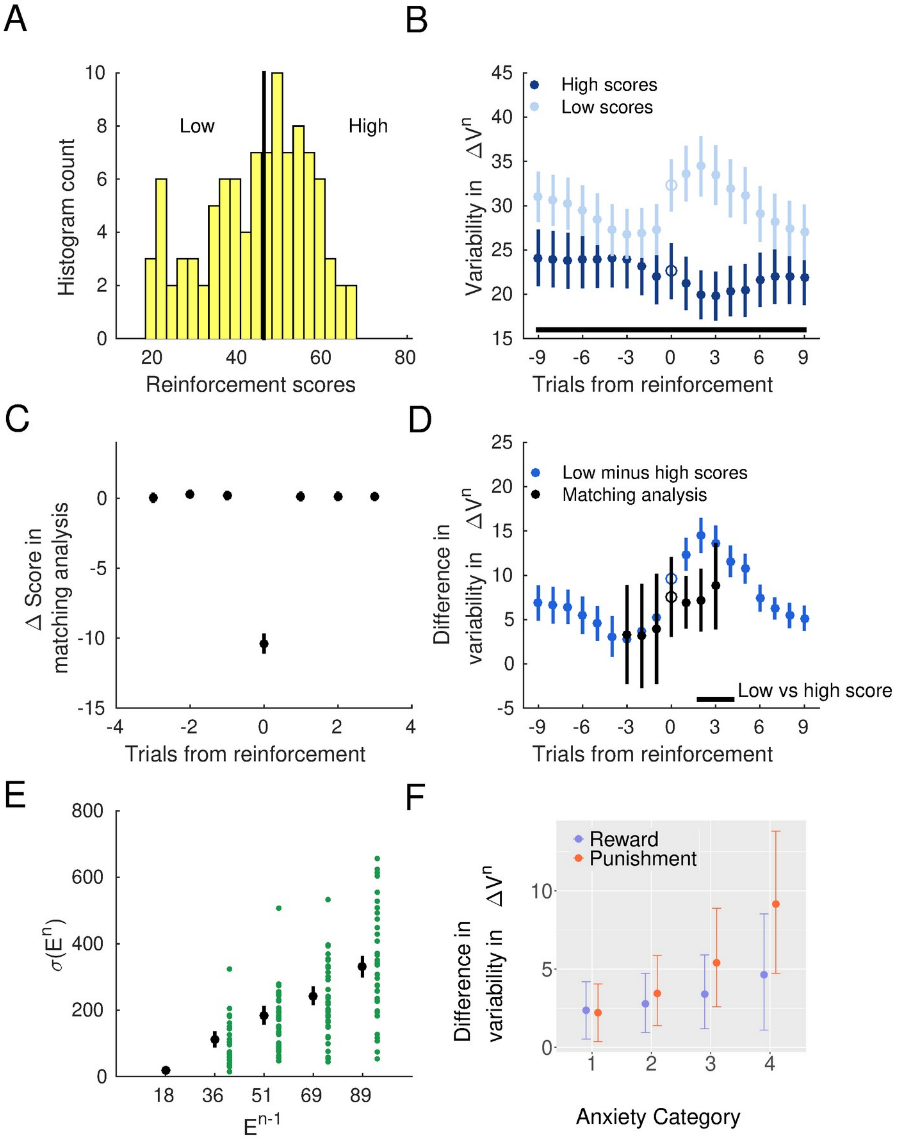

Figure 6 | Reinforcement-related modulation of motor variability in Experiment 3. A. Histogram of reinforcement scores (example participant) with median split used to define low and high score conditions. **B.** Time course of motor variability (ΔV^n) surrounding relatively high (dark blue) and low (light blue) score trials, aligned to the conditioned trial at position 0. Variability was greater following low scores at positions +1 to +3 (N = 36; paired permutation test; P_{FDR} = 0.001; significant cluster denoted by the black bar). **C.** Statistical matching analysis: trials of low and high scores (median split) were matched for surrounding reinforcement values. Black dots indicate mean performance score difference (± SEM) at each time point. **D.** Using matched trials, motor variability was significantly greater following low- than high-score trials at lags +1 to +3 (P_{FDR} = 0.006). Uncorrected differences from all trials (as in panel B) are shown in blue. **E.** Larger deviations from target dynamics (unsigned error E^{n-1}) were associated with greater subsequent reinforcement-related variability, σ(E^n), replicating the relationship found in Experiment 1. **F.** Bayesian linear modelling of *VarDiff* (post-low minus high score variability) showed a credible PA × reinforcement condition interaction (posterior estimate = 26.31, 95% CrI [11.34, 40.11]). Under punishment, the reinforcement-driven increase in motor variability became larger as PA levels increased, whereas under reward, posterior estimates overlapped with zero. Dots represent posterior means, bars denote 95% credible intervals.

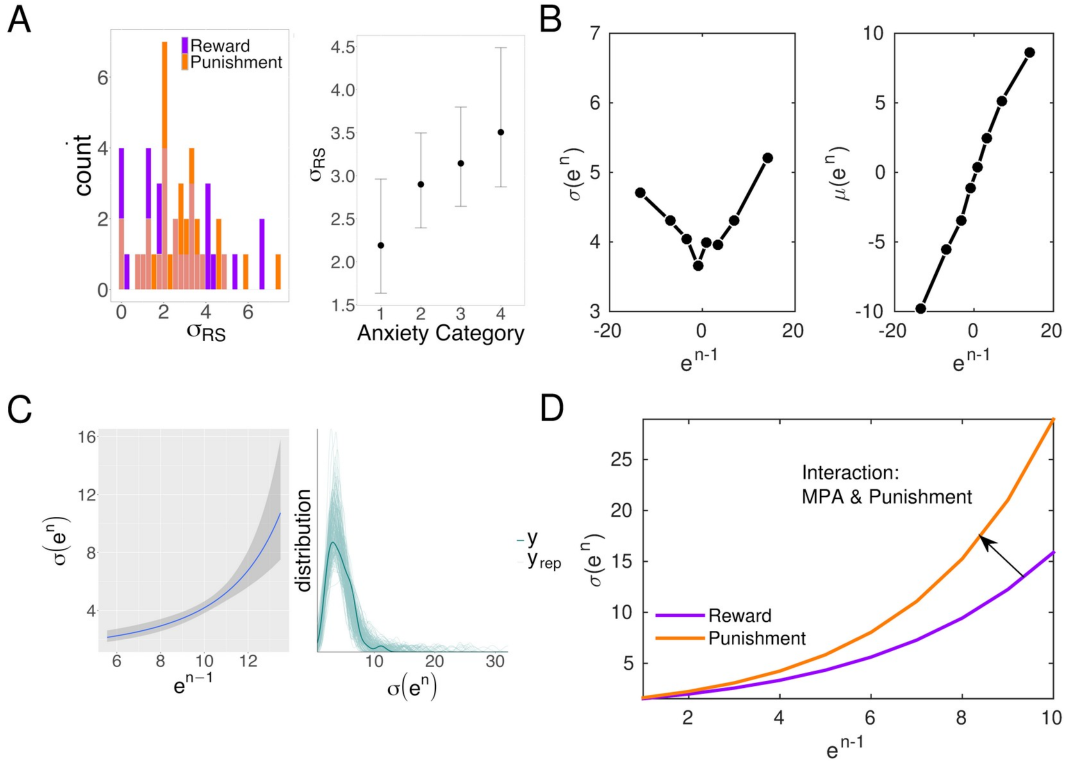

Figure 7 | Reinforcement-sensitive Gaussian process accounts for increased motor variability under punishment in higher performance anxiety. A. Left: distribution of σ^2_{RS} estimates across participants by reinforcement condition (orange: punishment; purple: reward). Variable σ^2_{RS} is the output scale of the reinforcement-sensitive kernel, which captures short-term variability that is modulated by reinforcement. Right: σ^2_{RS} increased with PA category, reflecting greater contribution of reinforcement-sensitive processes to variability in individuals with higher PA levels. Estimates are in the log scale. **B.** RSGP fit to empirical data replicates simulation results: $\sigma(e^n)$ shows a U-shaped dependence on e^{n-1} (left), and $\mu(e^n)$ increases linearly with e^{n-1}(right). Mean and SEM are shown; SEM values are very small and not visible. **C.** Left: exponential fit (line with 95% CrI shading) to the relationship between $\sigma(e^n)$ and $|e^{n-1}|$. Right: posterior predictive distribution of $\sigma(e^n)$ values (light green lines: individual posterior draws; dark green line: empirical density). **D.** Illustration of the interaction between PA category and reinforcement condition on the exponential growth parameter b_2. Under punishment (orange), the growth in $\sigma(e^n)$ with $|e^{n-1}|$ was amplified in higher PA individuals relative to reward (purple), indicating increased error sensitivity and behavioural adaptation through variability. Posterior estimate: b_2 = 0.06, 95% CrI [0.01, 0.19]