Sensory expectations and prediction error during feedback control in the human brain

Marco Emanuele[1,5], Paul L. Gribble[1,2,4], J. Andrew Pruszynski[1,2,4], Jonathan A. Michaels[1,2,3,*], Jörn Diedrichsen[1,5,6,*]

*co-senior authors
[1]Western Centre for Brain and Mind, Western University, London, Ontario, Canada
[2]Department of Physiology and Pharmacology, Western University, London, Ontario, Canada
[3]School of Kinesiology and Health Science, York University, Toronto, Ontario, Canada
[4]Department of Psychology, Western University, London, Ontario, Canada
[5]Department of Computer Science, Western University, London, Ontario, Canada
[6]Department of Statistical and Actuarial Sciences, Western University, London, Ontario, Canada
Corresponding author: Marco Emanuele (memanue5@uwo.ca)

Abstract

External disturbances to the body are better counteracted when their nature can be predicted in advance. Here, we investigated the neural mechanisms through which probabilistic predictions shape feedback responses using functional magnetic resonance imaging (fMRI) in humans. We show that, prior to a mechanical perturbation applied to a finger, the primary motor (M1) and somatosensory (S1) cortices receive a signal that linearly encodes the expected sensory input. When perturbations reach these areas, expectations are combined with the sensory input through a simple additive mechanism, yielding motor commands that reflect a weighted sum of the two signals. At the same time, M1 and S1 receive a prediction error signal, likely from upstream regions, encoding the difference between expectations and actual sensory input. This signal is visible in fMRI data in humans and in the local field potentials in non-human primates, but not in M1-S1 spiking activity.

Main

From skiing to cycling on uneven ground, many activities require rapid compensations for external disturbances to the body. Yet, feedback responses to these mechanical perturbations would arrive too late if exclusively driven by delayed[1–3] sensory input. Performance can improve if the nature (e.g., direction, intensity) of upcoming perturbations is predicted from contextual cues or prior experience. For example, a passenger standing on a bus can predict being pulled to the right when the bus is about to turn left at a light. However, if the driver suddenly swerves right to avoid a pedestrian, the passenger will be pushed left, having to reverse the expected response. Preparing movements in advance typically improves performance[4–9], but it is unclear how the brain shapes feedback responses based on probabilistic knowledge of future perturbations.

Recent work from our group showed that both humans and non-human primates were able to counter elbow perturbations (flexion or extension) more quickly when receiving valid probabilistic information about the perturbation direction[9]. In monkeys, the expected direction was probabilistically represented in the spiking activity of neurons in the dorsal premotor (PMd) and primary motor cortex (M1). At perturbation onset, sensory expectations shaped the initial response, which was then increasingly dictated by the accumulating sensory input that signalled the actual perturbation direction. These findings point to a simple mechanism that pre-activates cortical motor regions based on expectations, and then additively combines expectations with incoming sensory information to generate the motor commands for feedback responses.

Are there additional computations involved in the manipulation of probabilistic information for feedback control that eluded the limited spatial coverage offered by electrophysiological recordings? Here, we investigated this question in humans using functional magnetic resonance imaging (fMRI). Previous work suggests that the brain may also represent the uncertainty and surprise[10–13] associated with expectations and incoming sensory input, together with a prediction error signal useful for control and subsequent learning[14–17]. However, this information was mostly absent in the spiking activity recorded in monkeys[9]. The broader spatial coverage offered by fMRI allowed us to assess how sensory expectations are represented and then combined with incoming perturbations across cortical motor regions, including those not reached by our previous electrophysiological recordings.

Importantly, the blood-oxygen-level-dependent (BOLD) signal mostly reflects the synaptic input to neural populations[18–20] and therefore provides a complementary measure to spiking activity. Accordingly, to achieve a meaningful comparison with our previous electrophysiological recordings[9], we corroborated our fMRI results in humans by assessing the local field potentials (LFPs; also a measure of synaptic input[21]) recorded simultaneously with spiking activity in our non-human primates dataset[22].

Results

Sensory expectations bias responses to mechanical perturbations in a finger perturbation task

To adapt our previous probabilistic perturbation paradigm[9] for fMRI, we chose a task in which human participants (Experiment 1, N=14; Experiment 2, N=10) countered sudden mechanical perturbations (~3.5N) to their right index or ring finger (Fig. 1a). Compared to elbow perturbations, finger perturbations are easier to deliver without causing motion artifacts in fMRI data. More importantly, at the relatively low spatial resolution of fMRI, finger representations are more spatially distinct than movement directions[23], enabling a clearer differentiation of the neural activity patterns associated with the two perturbations.

Each trial began with the presentation of a visual cue (preparation epoch) signalling the probability that either the index or ring finger would receive the perturbation. Following a variable delay of 1.5-2.5s, a perturbation randomly drawn from the cued probability was applied to one of the two fingers using a pneumatic piston. The participant had to respond as quickly as possible by pushing down the piston with the perturbed finger (execution epoch).

The active force response began 163±21ms after the perturbation and was modulated by the probability cue (Fig. 1b). Specifically, between 0.2-0.4s after the perturbation, the stimulated finger produced a larger force if cued with higher probability during preparation (Fig. 1c; index: $F_{3,39}$=21.134, P<0.001; ring: $F_{3,39}$=6.109, P=0.002). The response often began with the finger cued with higher probability even when the perturbation was applied to the other (e.g., 75:25% probability in favour of the index followed by ring perturbation), with participants switching to the perturbed finger only after sensory evidence had accumulated. The overall response can be described as a 2-dimensional trajectory defined by the index and ring finger force (Fig. 1d). Responding only with the perturbed finger produces a straight trajectory. Trials in which the perturbation was applied to the finger cued with lower probability showed a significantly larger mean deviation (see Methods) from the ideal straight trajectory compared to those in which the perturbation was applied to the high-probability finger (Fig. 1e; index: t_{13}=2.743, P=0.008; ring: t_{13}=3.200, P=0.003).

Similar to what we recently reported for the arm[9], these results show that rapid finger responses are generated by combining probabilistic information available before the perturbation with the incoming sensory input.

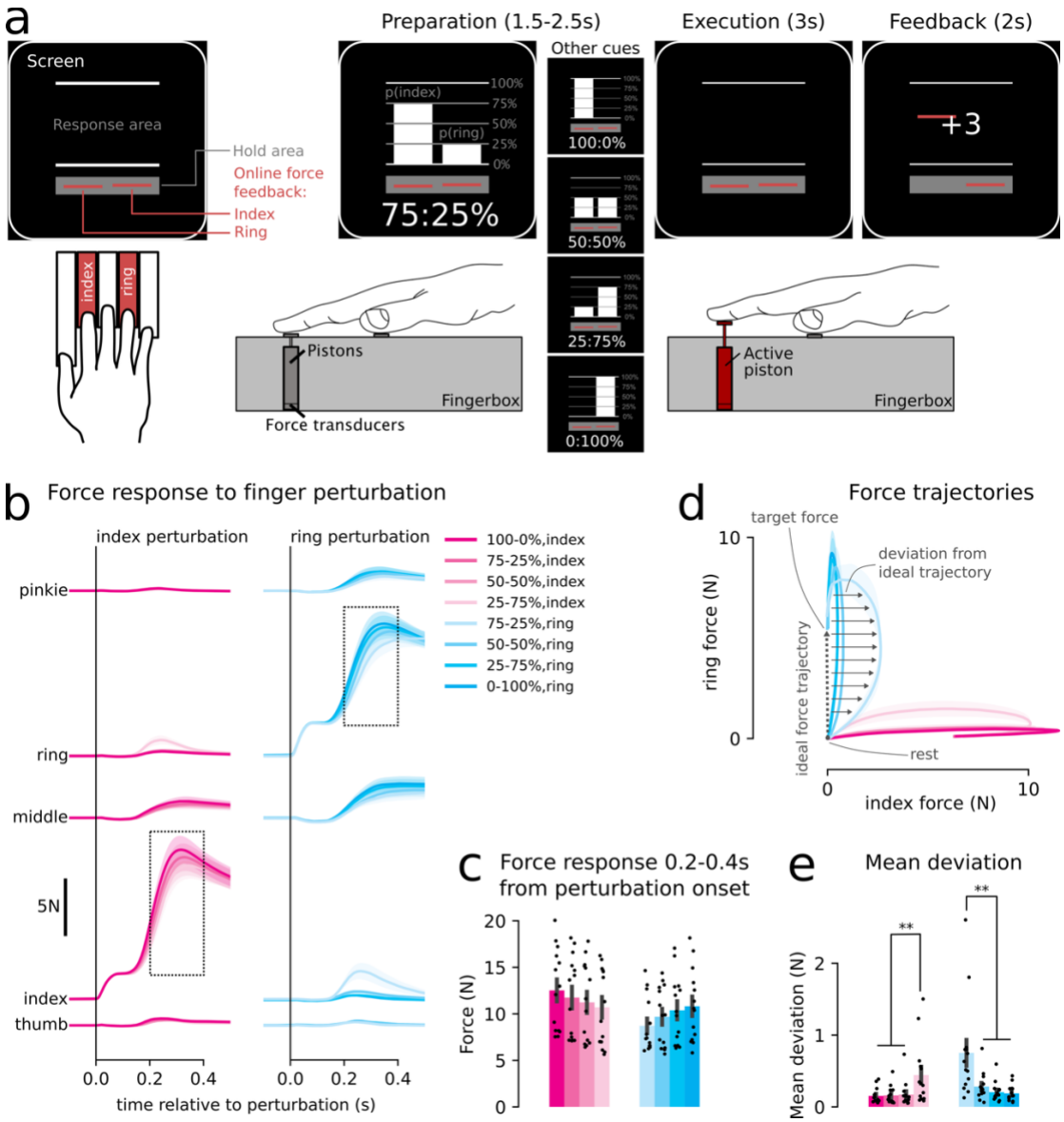


Figure 1. **(a)** Participants placed their right hand on a 5-finger keyboard. During preparation, a visual cue indicated the probability that either the index or ring finger would be perturbed. Participants were instructed to maintain the force on both fingers (indicated by red line cursors) within a 0.1-0.6N range as indicated by the hold area (grey rectangle). During execution, participants had to counter the perturbation applied to the index or ring finger by pushing back the pneumatic piston as quickly as possible. In the feedback epoch, participants received a score (-1, 0, +1, or +3) depending on their performance (see Methods). **(b)** Force response on all 5 fingers for index and ring finger perturbations, depending on the cued probability. The initial bump reflects the mechanical perturbation applied to the index (left column) or ring (right column) finger. The black arrows mark the start of the active force responses. **(c)** Mean force responses between 0.2-0.4s (see dashed rectangles in b) from the perturbation. Error bars denote ±SEM across participants. **(d)** Mean force trajectories for index and ring finger perturbation between 0-0.5s from the perturbation. To assess corrections in finger selection, we calculated, in each trial, the mean deviation from the ideal straight force trajectory (dashed arrow). **(e)** The mean deviation was significantly larger when the stimulated finger was the one cued with lower probability. Black dots indicate individual participants. Asterisks denote statistical significance (*P<0.05, **P<0.01).

Representation of expectation and uncertainty during preparation

In Experiment 1, participants performed the task while being scanned with 7T fMRI. Because the haemodynamic response is slow, we included both go- and no-go trials in the design. During no-go trials, the cue was shown but no perturbation occurred, allowing us to estimate the BOLD response during preparation independent of execution.

BOLD activity aligned with cue presentation and averaged across participants within each region-of-interest (ROI; see Fig. 2a) showed a clear separation between go and no-go trials (Fig. 2b). To characterise brain activation, we fitted a general linear model (GLM) with separate regressors for response preparation and execution (see Methods). During response preparation, we observed a significant BOLD activation relative to resting baseline in all ROI except S1 (Fig. 2c; Table 1, first row).

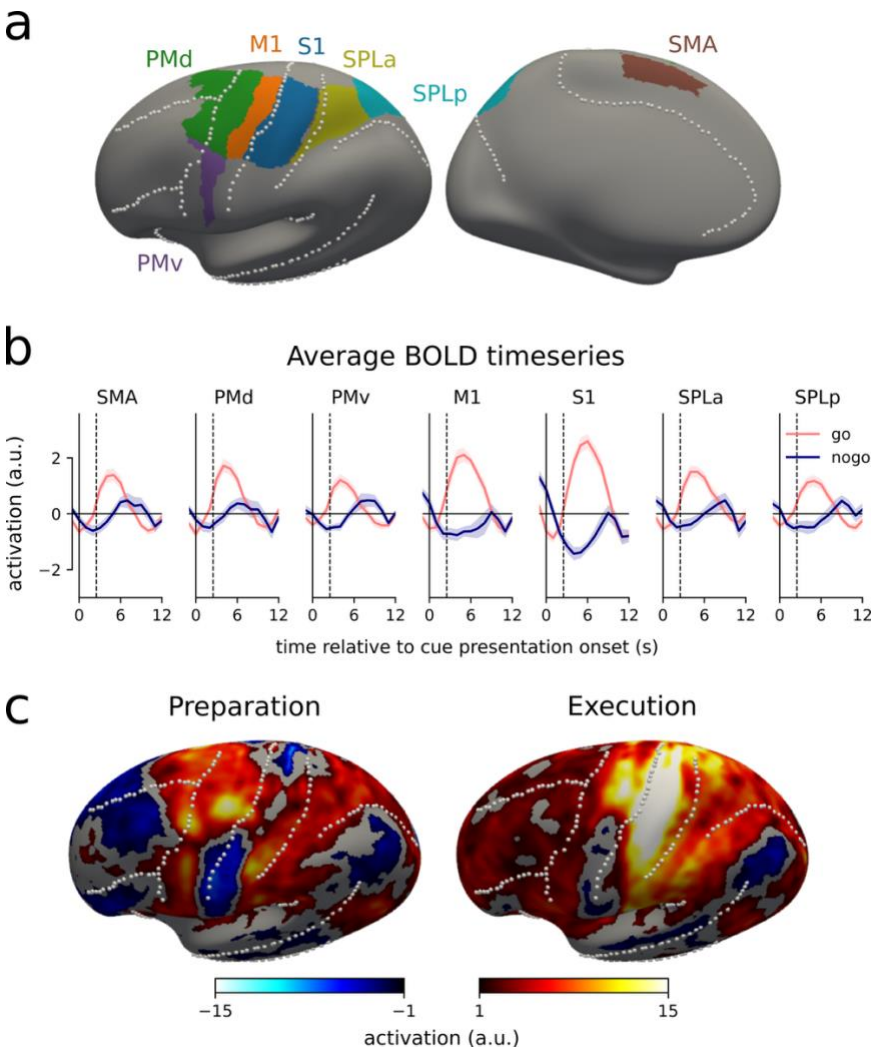


Figure 2. **(a)** ROI as defined on the group inflated surface of the left hemisphere. Dotted lines indicate major sulci. **(b)** Average BOLD time series for each ROI for go and no-go trials, aligned to cue onset (continuous vertical line). The dashed vertical line marks 2.5s after cue onset, i.e., the longest jitter allowed for cue presentation. Error bars indicate ±SEM across participants. **(c)** Activation relative to resting baseline (i.e., average contrast estimates, a.u.) during preparation and execution in the left hemisphere. Results for the right hemisphere are shown in Supplementary Materials 1.

Yet, univariate activation offers only a superficial view of the neural processes that occur in a brain region. In different conditions, activations and deactivations across voxels can yield similar regional averages, while converging to distinct neural states. We used representational similarity analysis (RSA)[24–26] to assess multi-voxel activity patterns and characterise the neural representation of the task in each ROI. We first evaluated whether the preparatory activity patterns differed significantly across the 5 probability cues. To this end, we calculated the cross-validated Mahalanobis (crossnobis) dissimilarities[27] between the activity elicited by different probability cues in each ROI. In a region without cue representation, the average dissimilarity estimate should be zero. In contrast, all ROI showed above-chance encoding (Table 1, second row).

We then asked how the probability cues were represented in each ROI. A-priori we hypothesized two possible representational geometries (Fig. 3a). The *expectation* representation encodes the upcoming perturbation (index vs. ring). Therefore, the neural activity patterns for the 100:0% should be maximally different from the 0:100% condition, with the other patterns being weighted averages of these extremes.

This corresponds to activity patterns that are linearly ordered according to the expected perturbation (Fig. 3b, horizontal axis). In contrast, in the *uncertainty* representation (Fig. 3b, vertical axis) the neural activity patterns for 100:0% and 0:100% are identical (certain perturbation) but maximally different from the 50:50% condition (undetermined perturbation).

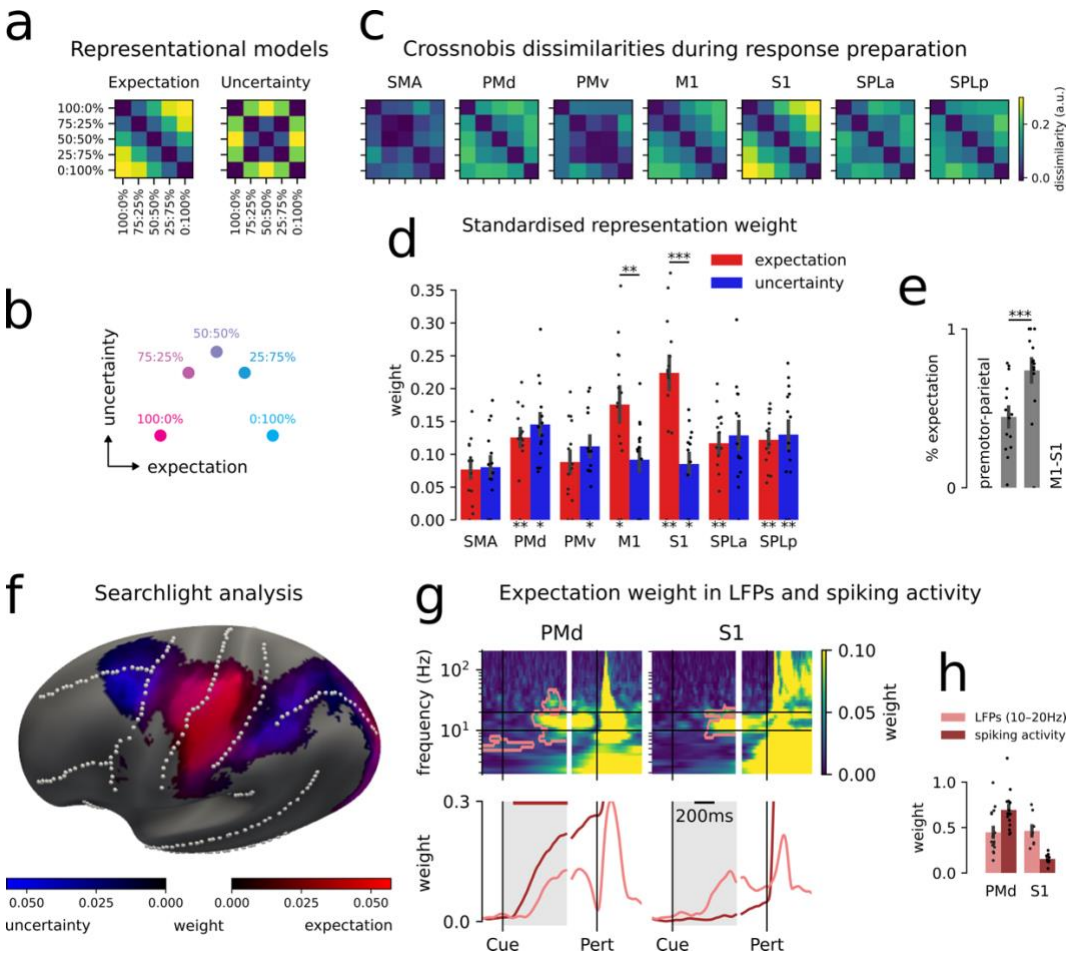


Figure 3. Activity patterns during preparation. **(a)** Hypothesised RDMs of the expectation and uncertainty representations. **(b)** Representational geometry of an area with a mixture of expectation and uncertainty encoding. **(c)** Average crossnobis dissimilarities for each ROI. **(d)** Standardised weight of expectation and uncertainty in each ROI. The asterisks below the bars denote the statistical significance (*P<0.05, **P<0.01, ***P<0.001) of the log-Bayes factor against 0, indicating that removing the component significantly reduces the overall model performance (see Methods). Error bars indicate ±SEM across subjects. The horizontal bars denote significant weight differences between expectation and uncertainty. Black dots show individual participants. **(e)** Weight of expectation normalized by the sum of expectation and uncertainty in M1-S1 and premotor-parietal regions. **(f)** Weight of expectation and uncertainty component in a continuous searchlight analysis conducted on the surface of the left hemisphere. **(g)** Weight of expectation in the LFPs (upper and lower panels, pink lines) and spiking activity (lower panels, red lines) recorded from PMd and S1 in our previous non-human primate dataset[9], aligned to cue presentation (Cue) and perturbation onset (Pert). In the upper panels, pink contours denote time-frequency clusters where the log-Bayes factor for expectation was significantly higher than 0. **(h)** Average expectation weight in the grey-shaded time interval in panel g. Note that, for non-human primate data, black dots refer to individual recording sessions rather than different monkeys.

Visual inspection of the representational dissimilarity matrices (RDMs; Fig. 3c) suggested that M1 and S1 reflect the expectation, whereas premotor and parietal areas are more similar to the uncertainty representation. We used pattern component modelling (PCM) to quantify these observations and express the information content in each region as a weighted combination of the expectation and uncertainty representations. Both representations contributed to preparatory activity patterns. Indeed, removing either

representation worsened the model performance, as indicated by significantly positive log-Bayes factor (see Methods) in most ROI for both expectation and uncertainty (Table 1, third and fourth rows). While this suggests that preparatory activity across cortical motor areas represents both expectation and uncertainty, the strength of the two representations varied markedly.

In M1 and S1, the standardised weight of the expectation representation was significantly higher than for the uncertainty (Fig. 3d; M1, t_{13}=2.934, P=0.006; S1, t_{13}=4.056, P<0.001). By contrast, in premotor and parietal areas uncertainty was slightly stronger compared to the expectation representation, although this difference was not statistically reliable (all t_{13}<1.108, P>0.144). To directly demonstrate the different information in the two groups of regions, we assessed the weight of the expectation relative to the summed weight for both representations (Fig. 3e). The proportional expectation weight was significantly larger in M1 and S1 compared to premotor and parietal areas (t_{13}=5.188, P<0.001). The different weight of expectation and uncertainty in M1-S1 and premotor-parietal areas is also evident in a continuous searchlight analysis (Fig. 3f).

Table 1. ROI-based statistics for preparation. T-values are one-sided t-tests against 0 with uncorrected P-values provided. For a family-wise error rate of 0.05, the region passes the Bonferroni correction for an uncorrected P-value<0.007.

	Statistics	SMA	PMd	PMv	M1	S1	SPLa	SPLp
	Activity>rest	t_{13}=3.050, P=0.005	t_{13}=3.886, P=0.001	t_{13}=3.535, P=0.002	t_{13}=2.812, P=0.007	t_{13}=0.726, P=0.240	t_{13}=2.593, P=0.011	t_{13}=1.895, P=0.040
	Encoding	t_{13}=3.501, P=0.002	t_{13}=4.598, P<0.001	t_{13}=5.381, P<0.001	t_{13}=4.900, P<0.001	t_{13}=7.019, P<0.001	t_{13}=4.419, P<0.001	t_{13}=6.070, P<0.001
log-Bayes factor	Expectation	t_{13}=1.082, P=0.149	t_{13}=3.697, P=0.001	t_{13}=1.618, P=0.065	t_{13}=2.447, P=0.015	t_{13}=3.397, P=0.002	t_{13}=3.035, P=0.004	t_{13}=3.251, P=0.003
	Uncertainty	t_{13}=1.570, P=0.070	t_{13}=2.108, P=0.027	t_{13}=2.124, P=0.027	t_{13}=1.421, P=0.089	t_{13}=2.165, P=0.025	t_{13}=1.675, P=0.059	t_{13}=2.985, P=0.005

Information about expectation reaches S1 as synaptic input

In contrast to our current fMRI results, our previous findings in non-human primates did not show a strong expectation signal in S1[9]. To confirm this, we re-visited our electrophysiological recordings from monkeys performing the arm perturbation task[22] and fitted the relative weight of expectation and uncertainty to the spiking activity across cortical motor regions (see Supplementary Materials 3 for results about uncertainty). Unlike in PMd (and M1, see Fig. S3a), the expectation representation was nearly absent in S1 (red line in Fig. 3g, lower panel).

While this discrepancy could be due to differences between paradigms or species, it could also depend on the physiological underpinnings of the different recording modalities. Spiking activity reflects the neuronal output, whereas the BOLD signal is more influenced by the synaptic input to neural populations[18,19]. To test the idea that expectations are represented in the synaptic input to S1 without influencing the spiking activity, we fitted the relative weight of expectation and uncertainty to the LFPs recorded simultaneously with spiking activity in our previous electrophysiological dataset[22] and reflecting synchronized synaptic activity[21] (see Supplementary Materials 2 for power modulations across frequency bands). We found that, in the LFPs recorded from both PMd and S1, expectations were significantly represented (P<0.05; cluster-based permutations) in a frequency band from 10-20Hz (i.e., low beta-band; Fig. 3g; see Supplementary Materials 3 for results about uncertainty). In these areas, the expectation weight in the LFPs and spiking activity showed a significant interaction between recording modality and ROI (Fig. 3h; $F_{1,48}$=25.721, P<0.001). Thus, a parsimonious explanation for the apparent discrepancy between fMRI and electrophysiological findings is

that the information about expectations reaches S1 as synaptic input but does not influence the neuronal spiking.

Expectations pre-activate the representation of the future sensory input

We then asked how the expectation signal is related to the activity patterns elicited by incoming sensory information. At one extreme, expectations may simply pre-activate the sensory representation of the most likely finger. Alternatively, the activity that generates the expectation signal could be orthogonal to the actual sensory input in voxel space.

To test this, we calculated the difference between the activity patterns in M1 and S1 for index (100:0% and 75:25%) and ring finger (0:100% and 25:75%) expectation during preparation, as well as between the activity patterns for index and ring perturbation during execution. If expectations and sensory input activate the same cortical representation, preparatory activity would be a scaled version of execution; consequently, the difference between the activity patterns for index and ring expectation during preparation should be perfectly correlated with the difference between index and ring perturbation during execution (i.e., correlation=1). This hypothesis is impossible to test using Pearson's correlation, because measurement noise reduces the correlation estimates to lower than 1 even if the true underlying patterns are perfectly correspondent. For this reason, we used PCM to obtain maximum-likelihood correlation estimates (MLE) unbiased by measurement noise (see Methods).

In M1 and S1, the unbiased correlation estimates were positive but lower than 1 (Fig. 4; M1, ρ_{MLE}=0.633, 95% CI [0.516, 0.713]; S1, ρ_{MLE}=0.611, 95% CI [0.468, 0.728]). Therefore, the expectation of a perturbation on a specific finger pre-activates the voxels that also receive the incoming sensory information caused by the perturbation. At the same time, a correlation of ~0.6 implies that roughly two thirds of execution variance is not captured by preparatory activity, suggesting that in these two epochs cortical activity explores partly different subspaces, consistent with well-established previous findings[28,29].

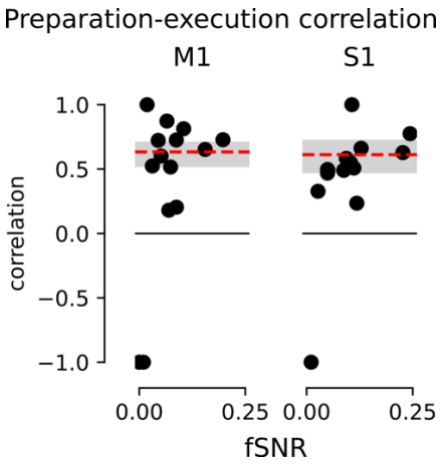
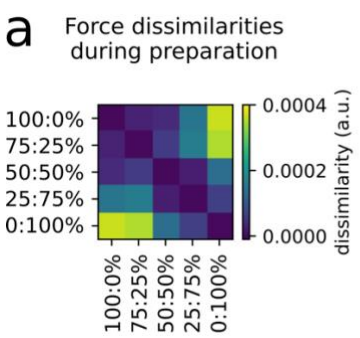


Figure 4. Maximum-likelihood correlation estimates between activity for expectation and sensory input. The correlation estimates for each participant are plotted against the functional signal-to-noise ratio (fSNR; see Methods). The dashed red line corresponds to the group correlation estimate. The shaded grey area denotes the 95% central CI established by participant-wise bootstrap.

The expectation representation is not caused by overt motor output during preparation

While the expectation representation in M1 and S1 was strong, we needed to rule out that it was not driven by subtle finger pre-activation. To discourage unwanted modulation of finger force before perturbation onset, we required that participants kept the two cursors indicating the force produced by the index and ring finger in a range between 0.1-0.6N (see Fig. 1a). Participants complied well with this instruction, and the force difference between the cued and non-cued fingers before the perturbation was minimal (0.050±0.025N).

However, crossnobis dissimilarities between the force patterns produced by the 5 fingers during preparation suggested that sometimes this subtle finger pre-activation reflected the cued probability (Fig. 5a).



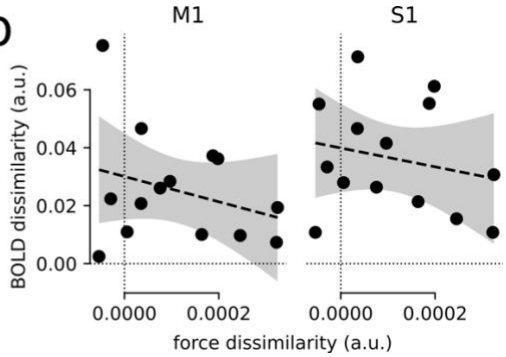


Figure 5. **(a)** Crossnobis dissimilarities between force patterns during response preparation. Note that the dissimilarities closely reflect the expectation model in Fig. 3a. **(b)** Linear fit between average crossnobis dissimilarity observed in force and in BOLD activity patterns during preparation. The black dots denote individual participants. The shaded region represents the 95% confidence interval (CI) for the regression line.

To test whether these pre-activation patterns could explain the expectation representation in M1 or S1, we performed a linear regression analysis between the average crossnobis dissimilarity observed in force (see Fig. 5a) and in M1 and S1 activity patterns (see Fig. 3b) during preparation for each participant. We found no systematic relationship between these two variables. Most importantly, the intercept of the regression (i.e., our estimate of the neural difference for participants that did not show any expectation-driven pre-activation) was significantly larger than 0 both for M1 and S1 (Fig. 5b; M1, intercept=0.030, P=0.001; S1, intercept=0.040, P<0.001). Therefore, the expectation representation during preparation genuinely reflected probabilistic predictions, rather than pre-activation of the finger cued with higher probability.

Representation of sensory input and surprise during execution

The mechanical perturbation applied to the fingertip and the ensuing participant's response elicited a strong increase in the BOLD signal across cortical motor areas (see Fig. 2b,c). To examine the representational geometry during execution, we computed the crossnobis dissimilarities between activity patterns elicited by index or ring finger stimulation. For each finger, we also split the data by the cued probability. We hypothesised three possible representational geometries for execution activity (Fig. 6a). First, execution activity may only reflect the *sensory input* (i.e., the stimulated finger). Second, the *expectation* representation identified in preparatory activity may be sustained into the execution epoch independently of the sensory input representation (see Methods). Finally, the brain may represent how *surprising* the perturbation was relative to the expectation, i.e., a representation of the absolute prediction error between expectations and sensory input. Using PCM, we estimated the relative weights for the sensory input, expectation and surprise.

Table 2. ROI-based statistics for execution. See Table 1 for statistical tests and conventions.

	Statistics	SMA	PMd	PMv	M1	S1	SPLa	SPLp
	Encoding	t_{13}=3.741, P=0.002	t_{13}=4.633, P<0.001	t_{13}=4.872, P<0.001	t_{13}=4.856, P<0.001	t_{13}=4.091, P=0.001	t_{13}=4.644, P<0.001	t_{13}=5.413, P<0.001
log-Bayes factor	Sensory input	t_{13}=-0.437, P=0.665	t_{13}=2.812, P=0.007	t_{13}=1.346, P=0.101	t_{13}=3.751, P=0.001	t_{13}=4.937, P<0.001	t_{13}=2.111, P=0.027	t_{13}=-0.751, P=0.767
	Expectation	$t_{13} \ll 0$, P=1.000	t_{13}=-4.487, P=1.000	$t_{13} \ll 0$, P=1.000	t_{13}=-0.860, P=0.797	t_{13}=2.429, P=0.015	t_{13}=-1.262, P=0.885	$t_{13} \ll 0$, P=1.000
	Surprise	t_{13}=3.482, P=0.002	t_{13}=4.817, P<0.001	t_{13}=3.224, P=0.003	t_{13}=2.415, P=0.016	t_{13}=2.906, P=0.006	t_{13}=3.695, P=0.001	t_{13}=2.692, P=0.009

The empirical RDMs suggested a strong representation of the sensory input, especially in M1 and S1 (Fig. 6b). Indeed, the log-Bayes factor (see Methods) for sensory input was significantly positive in PMd, M1, S1 and SPLa (Table 2, second row). Furthermore, all ROIs showed a significant representation of surprise (Table 2, fourth row). In contrast, no expectation representation independent of the sensory input was found during the execution period in any ROI except S1 (Table 2, third row).

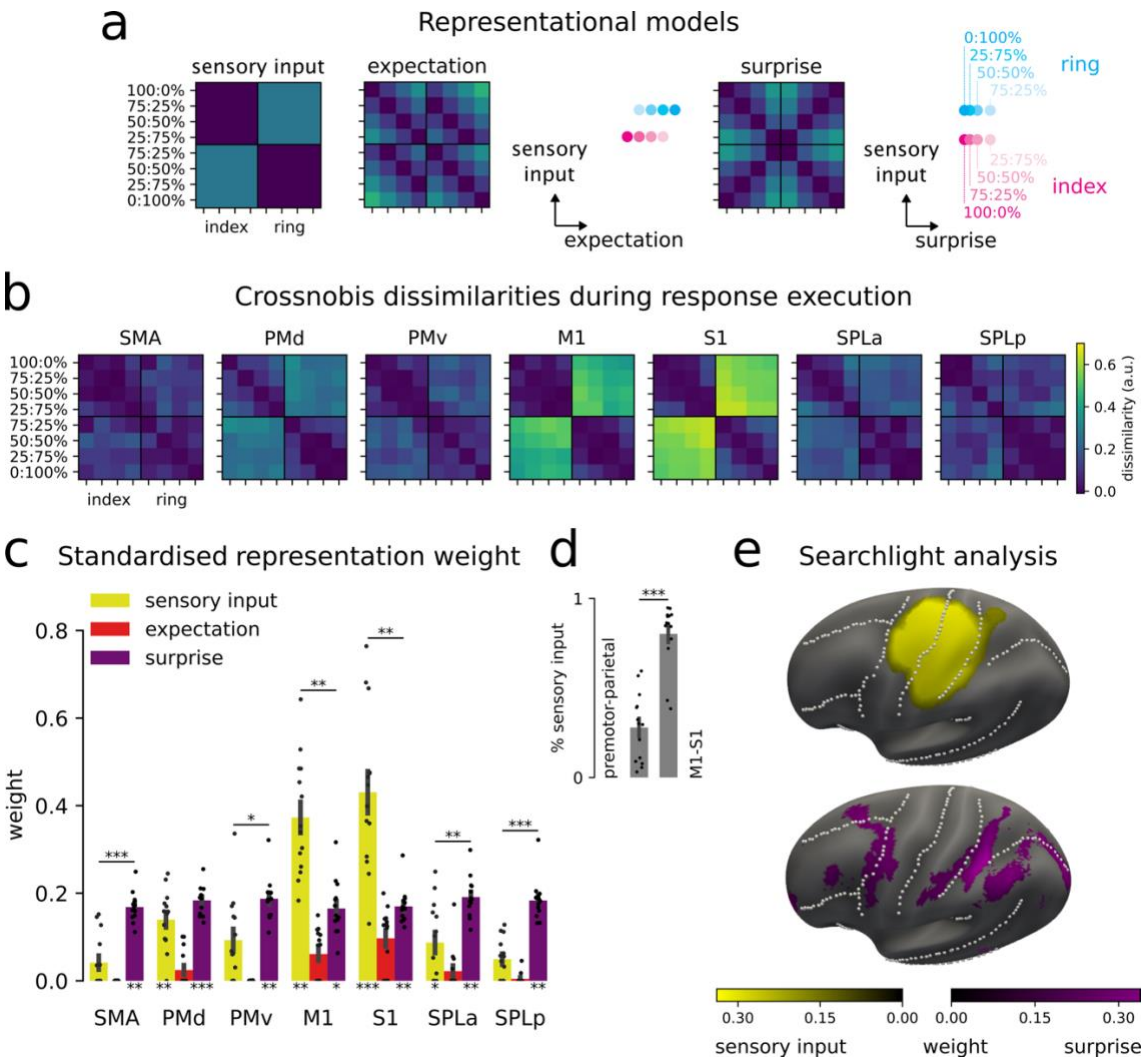


Figure 6. **(a)** Hypothesised RDMs and corresponding representational geometries for response execution. **(b)** Crossnobis dissimilarity during response execution. Note that the representational geometry in M1 and S1, beside a strong effect of the stimulated finger, also reflects the negative correlation model in Fig. 7c,d. For comparison, see EMG activity in the LLR and Vol time windows in Fig. 7e, which instead reflects the positive correlation model. **(c)** Standardised representation weight of sensory input, expectation and surprise within each ROI. **(d)** Relative weight of sensory input vs. sensory input+surprise in primary sensorimotor and premotor-parietal regions. The asterisks below the bars denote that the log-Bayes factor of the corresponding representation was significantly larger than 0 (*P<0.05, **P<0.01, ***P<0.001). The horizontal bar with asterisks denotes the significant difference between the two groups of ROI. **(e)** Standardised representation weight of sensory input and surprise projected on the inflated surface of the left hemisphere. Black dots denote individual participants.

Sensory input and surprise representations also showed different strength across ROI. The sensory input was prominent in M1 and S1, with a significantly larger weight than surprise (Fig. 6c; M1, t_{13}=3.422, P=0.002; S1, t_{13}=3.347, P=0.003). On the other hand, the surprise representation was significantly stronger than the sensory input in most premotor and parietal areas (SMA, t_{13}=4.816, P<0.001; PMv, t_{13}=1.819, P=0.046; SPLa, t_{13}=3.034, P=0.010; SPLp, t_{13}=4.872, P<0.001). The proportion of sensory input weight

relative to the summed weight of sensory input and surprise was significantly larger in M1-S1 compared to premotor-parietal areas (Fig. 6d; t_{13}=12.162, P<0.001), providing direct support to the distinct representation of these two features across different ROI. The differential distribution of sensory input and surprise representations can also be seen in a continuous searchlight analysis (Fig. 6e).

Integration of expectation and sensory input across motor hierarchy

While the expectation was not represented during execution independently of the perturbation, the modulation of the force response in different conditions (see Fig. 1b,c) suggests that participants integrated the expectations with the incoming sensory information. How does this integration occur? One possibility is that execution activity results from the weighted sum of expectation and sensory input, in a Bayesian-like additive process. In this case, a high probability cued on the index finger would make response patterns more index-like, and vice versa for the ring finger. Accordingly, the dissimilarity between execution activity patterns would be larger between conditions in which expectations and sensory input are congruent and smaller between conditions in which they are incongruent (Fig. 7a); that is, the neural dimensions representing the expectations and the sensory input would be positively correlated (Fig. 7b). In addition, the brain could calculate a weighted difference between expectations and sensory input. This would result in a signed prediction error signal, with larger dissimilarities between conditions in which expectation is opposite to the sensory input (Fig. 7c), yielding a negative correlation between the two neural dimensions (Fig. 7d).

Participants' behaviour (see Fig. 1b,c) suggested that, in the motor output, expectations and sensory input are combined through the additive process. To characterise the response in more detail, in Experiment 2 we recorded the electromyographic (EMG) activity of hand muscles while participants performed the task seated at a desk outside of the scanner. We then used PCM to obtain unbiased estimates of the correlation between expectation and sensory input in EMG activity elicited after the perturbation (see Methods). There was no clear modulation in a 100ms time window before the perturbation (Fig. 7e,f; no reliable fSNR, see Methods), nor in the short-latency reflex (SLR; 25-50ms from perturbation; ρ_{MLE}=-0.036, 95% CI [-1.000, 0.719]). This is expected because SLRs are entirely mediated by a spinal circuit[30] and typically show limited modulation based on contextual influences[31,32]. On the other hand, we found a positive correlation between expectation and sensory input in the long-latency reflex (LLR; 50-100ms; ρ_{MLE}=0.646, 95% CI [0.370, 0.850]) and into the voluntary response (Vol; 100-500ms; ρ_{MLE}=0.774, 95% CI [0.673, 0.896]), consistent with the notion these response components receive cortical contributions via the transcortical feedback[33] loop and can therefore be subject to more sophisticated modulations[34].

The comparison of the estimated crossnobis dissimilarities in EMG and cortical activity in M1 and S1 suggested that expectations and sensory input are combined through distinct processes across different levels of the motor system (see Fig. 6b and 7e). To confirm this observation, we obtained unbiased estimates of the correlation between expectation and sensory input also for the neural activity patterns in M1 and S1 during execution. In contrast to the EMG patterns, execution activity in M1 and S1 exhibited a negative correlation between expectation and sensory input (Fig. 7g; M1, ρ_{MLE}=-0.823, 95% CI [-1.000, -0.685]; S1, ρ_{MLE}=-0.613, 95% CI [-0.933, -0.536]). Therefore, the BOLD activity in M1 and S1 does not reflect the muscular output but a signed prediction error signal between expected and actual sensory input.

We again leveraged our electrophysiological dataset in non-human primates[22] to determine whether the signed prediction error signal in M1 and S1 spreads to spiking activity or is confined to the synaptic input. The spiking activity of neurons in M1 and S1 showed a positive correlation between expectation and perturbation direction (Fig. 7i; ρ_{MLE}=0.407, 95% CI [0.280, 0.503]) from 0.04-0.24s after the perturbation, similar to EMG activity in the human dataset, and consistent with the additive integration process. In the same time window, the LFPs showed a negative correlation (Fig. 7h) spanning the 13-25Hz (i.e., beta-band;

ρ_{MLE} =-0.763, 95% CI [-1.000, -0.359]) and 25-100Hz (i.e., gamma-band; ρ_{MLE}=-0.947, 95% CI [-1.000, -0.658]) frequency bands.

Together, these findings indicate that the synaptic activity in M1 and S1, both in monkeys and humans, reflects the signed prediction error between the expected and the actual perturbation. In contrast, the spiking activity in both areas corresponds to a signal that additively combines expectations and perturbation and can be used to drive the muscles.

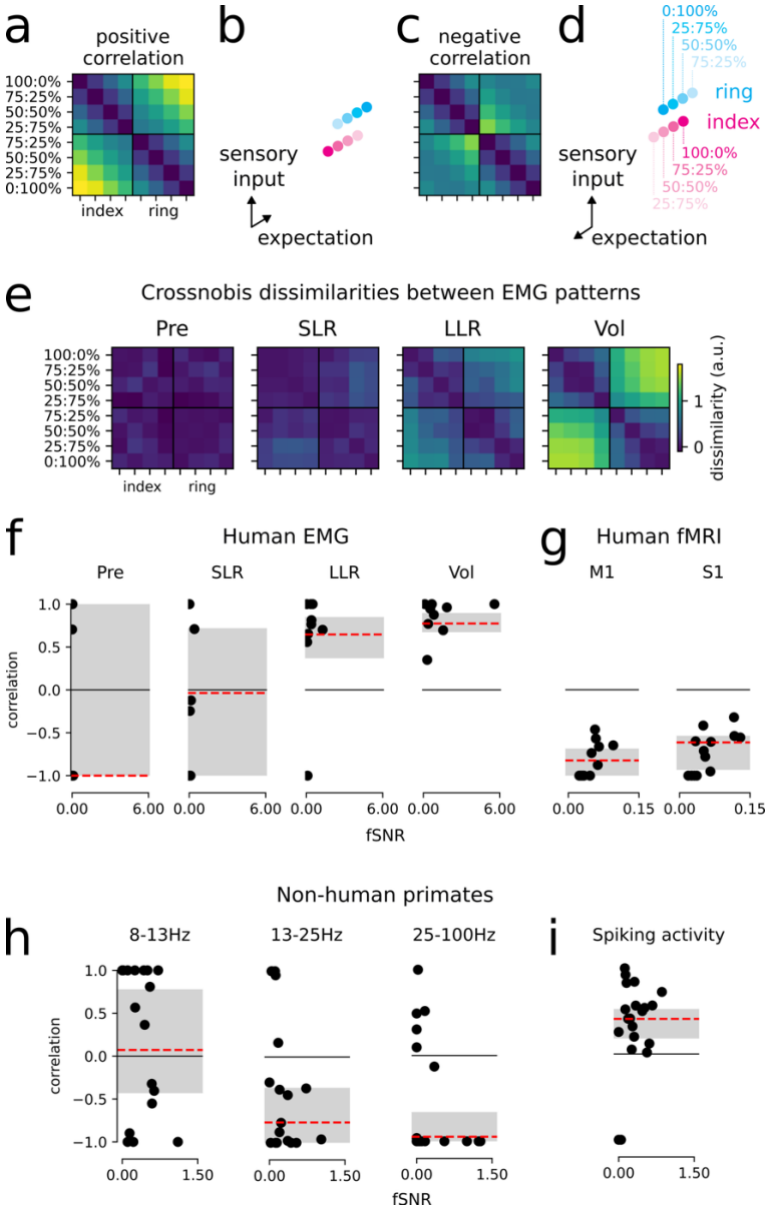


Figure 7. **(a-d)** Hypothesised RDMs for positive (a) and negative (c) correlation between expectation and sensory input and corresponding representational geometries (b,d). **(e)** Average crossnobis dissimilarities between EMG patterns across participants. **(f-i)** Maximum-likelihood correlation estimates between expectation and sensory input in (f) EMG activity, (g) neural activity patterns (i.e., beta coefficients) from human participants, and (h) LFPs and (i) spiking activity in M1-S1 in non-human primates. The black dots indicate individual participants in panels f and g, and different recording sessions in panels h and i.

To summarise our findings in humans, we projected the BOLD activity patterns from each condition across preparation and execution onto the dimensions in voxel space that explained most variance across conditions. The first principal component (PC1) reflected the main difference between preparation and execution. Along PC2, the preparatory activity patterns from M1 and S1 (Fig. 8a, magenta-cyan gradient dots) were linearly ordered according to the expected finger (see Fig. 3c). These low-dimensional projections also reveal the positive correlation between preparatory and execution activity (see Fig. 4): Expecting a perturbation on the index finger (magenta, 100:0%) pushed preparatory activity patterns to the

left, the same direction as the later index-finger perturbation. Both preparing (cyan, 0:100%) and responding to a ring-finger perturbation pushed the neural activity patterns in the opposite direction.

After the perturbation, the neural activity patterns were more index- or ring-like when the finger was cued with lower probability (Fig. 8a, magenta and cyan dots and arrows), consistent with the negative correlation between expectation and sensory input in the BOLD signal from M1 and S1 (see Fig. 7g). In contrast, EMG trajectories projected onto the first two PCs show an opposite geometry (Fig. 8b). Consistent with the positive correlation between expectation and sensory input (Fig. 7f), the EMG patterns elicited by perturbations applied to the finger cued with higher probability were further apart from each other (Fig. 8b, dark trajectories), compared to those elicited by perturbations to the finger cued with lower probability (light trajectories).

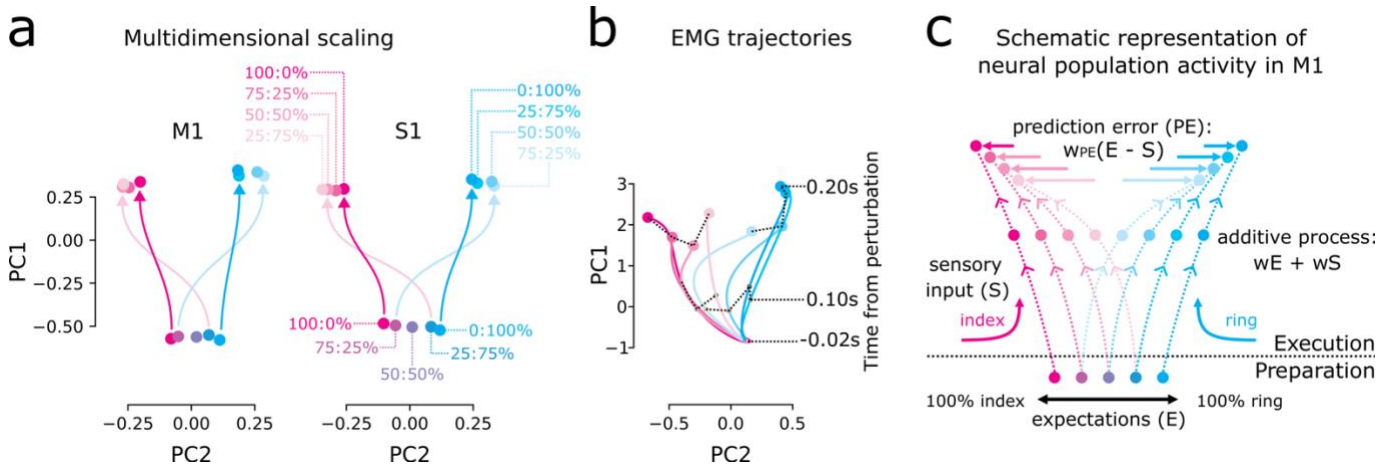


Figure 8. **(a)** Projections of M1 and S1 BOLD activity patterns during preparation and execution onto the first two principal components. The dots along the magenta-cyan gradient correspond to the five probability cues. The magenta and cyan dots reflect the activity patterns for index and ring perturbation, respectively. The arrows denote the transition from preparation to execution. **(b)** Trajectories of average EMG patterns projected onto the first two principal components (PCs). **(c)** Schematic representation of the integration of expectation and sensory input in M1 spiking activity. During preparation, neural population activity reflects the expectations about the upcoming perturbations (see Fig. S3a). During execution, the expectation is combined with the incoming sensory input through a simple additive mechanism (see Fig. 7e). Additionally, M1 receives a prediction error signal that may push the execution activity towards a state corresponding to the actual perturbation (horizontal magenta and cyan arrows).

Discussion

Our fMRI results in humans demonstrate that neural activity patterns in PMd and M1 scale linearly with sensory expectations about upcoming finger perturbations (Fig. 8c, preparation epoch). These results are consistent with our recent findings in non-human primates showing that spiking activity in PMd and M1 linearly represents expectations about upcoming elbow perturbations[9]. In the fMRI data, we found a similar expectation representation in S1, which was absent in the spiking activity recorded in the monkeys. This discrepancy likely occurs because the BOLD signal is more related to the synaptic input to neural populations and less to spiking activity itself[18,19]. Indeed, the LFPs recorded from non-human primates also showed a strong expectation signal in S1. Together, these results suggest that both M1 and S1 receive information about expectations in their synaptic input. Yet, unlike in M1, in S1 this input information does not influence the local spiking activity. This is consistent with the notion that the spiking of S1 neurons is tightly linked to the actual sensory input[35], without an expansive representation of other latent dimensions as in M1[28,29,36]. Interestingly, previous studies have shown similar expectation signals in the BOLD signal recorded from S1 during the preparation of self-initiated movements[37–40]. This suggests that the expectation

signal elicited by probabilistic information taps into a more general mechanism that injects information about an upcoming movement in the synaptic input to S1 before its onset.

Whether this information serves a specific function remains an open question. It is possible that this expectation signal could modulate or improve subsequent sensory processing in S1[41]. On the other hand, our previous work shows that S1 spiking activity is not modulated by expectations during execution (different from M1)[9]. Therefore, it remains possible that the expectation signal in the synaptic input to S1 is epiphenomenal, and reflects neural processes in other motor regions (e.g., M1, PMd)[42,43].

In both M1 and S1, the neural activity patterns during preparation and execution were positively correlated, which indicates that the neural populations pre-activated during preparation are those that receive sensory information from the finger cued with higher probability (Fig. 8c). Once the perturbation began, the expectation was combined additively with the incoming sensory input. In the finger task, this was visible from the stronger force response when the perturbation was applied to the finger cued with higher probability. The same finger dominated the initial force response even when the perturbation violated the expectations. This is consistent with our previous finding that feedback responses are initially triggered by an unspecific signal that marks the occurrence of the perturbation, without specifying on which finger (or in which direction) it is applied[9]. In the presence of the pre-activation of the most likely finger, this unspecific signal drives the initial feedback response depending on the expectations. In line with this idea, the spiking activity of M1 neurons recorded in monkeys, as well as the initial EMG response in human participants, reflected a weighted sum of expectation and sensory input.

In contrast, the synaptic input to M1 and S1, as indexed by BOLD or LFPs, reflected the signed prediction error (i.e., the difference) between expectation and sensory input. This is surprising because, from a control perspective, the nervous system would not need to compute a prediction error. The weighted sum of expectation and sensory input would be sufficient to guide feedback responses and can be computed without the explicit representation of their difference. In this scenario, expectation and sensory input should be positively correlated both in the spiking activity and in the synaptic input. In contrast, the correlation was negative in both BOLD signal and LFPs.

There are two potential, mutually non-exclusive uses for this signed prediction error signal. First, the mismatch between expectations and perturbation could be added to the response, driving the system toward the correct response more rapidly than the sensory input alone (Fig. 8c, horizontal magenta and cyan arrows). This is consistent with predictive coding theories of motor control, which propose that motor commands are generated based on the difference between expectations and incoming sensory information[14–16,44]. Second, the signed prediction error could be used for updating the expectations for subsequent trials[45]. This updating signal could be weighted by the precision of the expectations, reflected in the uncertainty representation we observed in premotor and parietal regions.

Where could the signed prediction error be computed? The absence of prediction error signals in the spiking activity of M1 and S1 makes it unlikely that this subtractive operation is performed locally. More likely candidates are premotor and parietal regions. These regions receive information about both expectations and sensory input, although not as strong as M1 and S1. The fMRI data show that this information is combined to calculate the unsigned prediction error, i.e., the surprise representation. Assuming independent and spatially intermingled neural populations encoding the signed prediction error separately for each finger, pooling their activity in the BOLD signal would rectify the signal and produce a response that reflects only the magnitude of the mismatch. At the same time, the original, unrectified prediction error signal transmitted to M1 and S1 could contribute to the representational geometry we found in the synaptic input to these regions.

In conclusion, we show that the synaptic input to M1 and S1 is similarly modulated in humans and monkeys responding to sudden mechanical perturbations applied to the fingers or the arm, respectively. These results provide new insights into the neural machinery that governs rapid feedback responses. Expectations are not only represented in the expansive latent dimensions of M1[28,29] without causing overt muscle activity[9], but are also transmitted as synaptic input to S1. As perturbations occur, feedback responses may not only benefit from the additive combination of this expectation signal with the incoming sensory information but also be further refined by the prediction error signal acting as a corrective drive in the synaptic input to M1.

Methods

Participants

We recruited 14 participants for Experiment 1 (6 females; age 18-34 years, mean 21.35 years, SD 3.77 years) and 10 participants for Experiment 2 (2 females; age 21-32 years, mean 25.70 years, SD 4.16 years). All participants were right-handed and did not report any neurological condition. The experimental procedures were approved by the Research Ethics Committee at Western University (HSREB 107061 for Experiment 1 and HSREB 108479 for Experiment 2). Participants provided written informed consent and were compensated for their participation.

Apparatus

We used a custom-made MRI-compatible keyboard device to deliver mechanical perturbations independently to the right index or ring fingertip and record the force response generated by the stimulated finger (Fig. 1a). The keys were equipped with force transducers that measured the isometric force generated by each finger (Honeywell FS series; dynamic range 0–16N; sampling frequency 500Hz). The fingers were comfortably restrained by a padded clamp adjusted to each participant's hand size. The mechanical perturbation (~3.5N) was delivered using pneumatic pistons (diameter 3mm) embedded underneath each key and operated by compressed air (~70psi).

Task

We instructed participants to counter the finger perturbation as quickly as possible. At the beginning of each trial, a probability cue consisting of two vertical bars was shown on a computer screen, indicating the probability that either index or ring finger would receive the perturbation (Fig. 1a). The probabilities were 0:100%, 25:75%, 50:50%, 75:25%, or 100:0% (index:ring) and were shown for 1.5–2.5s. During this preparation epoch participants received continuous force feedback from two horizontal cursors projected on the screen. The cursors moved upward when the corresponding key was pressed. To limit unwanted anticipatory finger presses, we asked participants to keep both cursors within a force range between 0.1-0.6N, symbolized by a grey rectangular hold area (Fig. 1a), while the probability cue was on the screen.

In Experiment 1, we used both go and no-go trials, while Experiment 2 included go trials only. In go trials, the probability cue disappeared at the end of the preparation epoch. Then, the piston underneath one of the two fingers was activated, applying an upward force of ~3.5N for 3s (i.e., response execution epoch). The padded clamp above the finger limited the upward movement of the finger to less than 5mm.

The stimulated finger was randomly drawn from the cued probability distribution. Participants were instructed to respond as quickly as possible by pressing the piston down until it was deactivated. In no-go trials, the probability cue remained visible for 5.5s (i.e., the longest jitter for response preparation, 2.5s, plus the equivalent of execution duration in go-trials, 3s).

The force feedback was frozen at perturbation onset (go trials) or after the probability cue was presented for 2.5s (no-go trials). At the end of the trial, both cursors showed the average force exerted by each finger during the execution phase. This delayed feedback helped participants adjust the amount of force applied in response to the perturbation. In go trials, the cursor of the stimulated finger should re-appear between two horizontal lines displayed on the screen, corresponding to a force range of 3.5–8.5 N; the cursor of the non-stimulated finger should remain within the grey hold area. In the training session (see Procedures), participants learned to produce the correct force output within a few trials. In no-go trials, both cursors should re-appear within the grey hold area.

In Experiment 1, participants also received a score based on how quickly they responded to the perturbation at the end of each trial. The displayed reaction time was the interval between perturbation onset and the time when the force generated by any of the two fingers exceeded 3.5N. The score was assigned using a staircase system: +3 points for reaction time below the 25th percentile of the previous block, +1 between the 25th and 75th percentile, and 0 above the 75th percentile. The thresholds used in the first block were 0.25s and 0.50s for all participants. Participants received a negative score (-1) if they exerted >1N isometric force with any finger before receiving the perturbation. No scoring was used in Experiment 2.

Procedures

In Experiment 1, participants completed a single fMRI session, consisting of 10 functional runs of 30 trials each and 1 anatomical scan. We used an event-related design in which we randomly interleaved the five probability cues (0:100%, 25:75%, 50:50%, 75:25%, 100:0%, index:ring) and the three stimulation outcomes (index, ring, or no perturbation for no-go trials). Because we always drew the stimulated finger from the displayed probability distribution, the 100:0% and 0:100% cues were always followed (in go trials) by index and ring stimulation, respectively, resulting in 13 trial types overall. Each functional run included 6 trials for each probability cue. Of these, 4 were go trials and 2 no-go trials, totalling 30 trials per run. Three 12.5-s periods of rest were randomly included in each functional run to allow for the estimation of baseline activation.

The day before the fMRI experiment, participants completed a short training session of 3-5 functional-like runs to familiarise themselves with the task. The training session was carried out in a mock fMRI scanner to make participants accustomed with the posture in which they would perform the task in the fMRI session.

Experiment 2 included go trials only, resulting in 8 trial types. Because the task was performed while sitting at a desk, it was not necessary for participants to get accustomed to an unfamiliar posture. The training session was therefore replaced by a brief familiarisation with the task and equipment before starting the experiment.

EMG recordings

In Experiment 2, we used an 11-channel surface EMG montage (Delsys, Trigno Research+ System, Trigno Duo Sensors) to record the activity of the extensor digitorum communis (EDC), extensor digiti minimi (EDM) and extensor indicis (EI), extensors of the thumb (extensor pollicis brevis and longus), the flexor digitorum superficialis (FDS), abductor pollicis brevis (APB), and abductor digiti minimi (ADM), and first dorsal interosseous (FDI). Raw EMG signals were acquired at 2148Hz. The skin was cleaned with alcohol before placing the electrodes for reducing impedance and improving the signal. We defined the ideal location of each electrode by asking the participant to perform slight isometric presses with each finger either in flexion

or extension direction. The electrode was placed where muscle activation was maximal, as indexed by palpation and through continuous monitoring of EMG activity.

Behavioural analysis

Visual inspection of the forces aligned to the perturbation suggested that the response scaled with the cued probabilistic predictions (see Fig. 1b). To quantify this observation, we averaged the isometric force produced by the stimulated finger (index or ring) in go trials between 0.2-0.4s after perturbation onset (see dashed rectangles in Fig. 1b) and tested the within-participant effect of the cued probability using a repeated-measures ANOVA. This analysis was performed separately for index and ring finger perturbation.

We also hypothesised that, following the 25:75% or 75:25% probability cues, if the stimulated finger was cued with lower probability (25%), participants may initially respond with the finger cued with higher probability (75%) and then correct their response as sensory evidence accumulated. To quantify these corrections, we considered the two-dimensional force trajectory (f_t) generated by the index and ring finger in each trial between 0–0.5s after perturbation onset and defined the ideal straight response direction (c) as the vector connecting the extreme points of this trajectory. We then computed, at each time point, the Euclidean distance of the force vector from this ideal straight line and averaged this distance over time (see Fig. 1d):

$$mean\ deviation = \frac{1}{T}\sum_{t=1}^{T}\left\| f_t - \frac{c^T f_t}{\|c\|^2}\cdot c \right\| \quad (1)$$

Where t corresponds to the perturbation onset and $T = t + 0.5s$. If the force response unfolded along the ideal straight line, the mean deviation would be 0. On the other hand, if the response was initiated with the non-stimulated finger and later corrected to the stimulated finger, the force trajectory would deviate from the ideal straight line, thus increasing the mean deviation. For each stimulated finger, we then performed a paired-sample t-test between conditions in which the perturbation was delivered on the finger cued with lower probability, and all the other conditions.

EMG preprocessing

The raw EMG signals were rectified, time-aligned to the perturbation and baseline-corrected by subtracting the mean EMG activity in the 0.1s preceding the perturbation. We defined four time windows relative to the perturbation, i.e., pre-perturbation (Pre, -0.1–0s), short-latency stretch reflex (SLR, 0.025–0.05s), long-latency stretch reflex (LLR, 0.05–0.1s) and voluntary response (Vol, 0.1–0.5s). Multivariate analysis and model fitting (see “Multivariate dissimilarity analysis” below) were performed separately on the mean EMG activity within each time window in each participant and condition.

Principal component analysis of EMG recordings

For principal component analysis (PCA; see Fig. 8b), we first concatenated trials from each participant and then standardised the EMG timeseries of each channel to zero mean and unit variance (i.e., z-score normalisation). We then extracted the two orthogonal dimensions in EMG channel space capturing the largest variance. The resulting timeseries of the two principal components were stratified by condition to produce the low-dimensional EMG trajectories shown in Fig. 8b.

Imaging data acquisition

In Experiment 1, we used a 7T Siemens Magnetom scanner with a 32-channel head coil to acquire high-field fMRI data. The anatomical T1-weighted scan was acquired at the end of the scanning session using a

Magnetization-Prepared Rapid Gradient Echo sequence (MPRAGE) with voxel size of 0.75 x 0.75 x 0.75mm isotropic (field of view = 208 x 157 x 110mm [A-P; R-L; F-H], encoding direction coronal). For functional scans (336 volumes) we used the following sequence parameters: GRAPPA 3, multiband acceleration factor 2, repetition time (TR) = 1.0s, echo time (TE) = 20ms, flip angle (FA) = 30°, slice number: 57, voxel size: 1.8 x 1.8 x 1.8mm isotropic. To estimate and correct for magnetic field inhomogeneities, we also acquired a gradient echo field map with the following parameters: transversal orientation, field of view: 210 x 210 x 160mm, 64 slices, 2.5mm thickness, TR = 475ms, TE = 4.08ms, FA = 35°.

Preprocessing of fMRI data and general linear model

Preprocessing of functional images was performed with SPM12 (https://www.fil.ion.ucl.ac.uk/spm/) and custom MATLAB code and involved the following steps: (1) correction of geometric distortions using field maps[46]; (2) rigid-body motion realignment of all images to the first image of the first functional run; and (3) co-registration to the anatomical scan. No smoothing or normalisation to a standard template was applied at this stage.

We then analysed the pre-processed images using a general linear model (GLM)[47], with separate regressors for response preparation and execution. Preparation was modelled with five regressors, one for each probability cue. Execution was modelled with eight regressors, capturing the activation elicited by index or ring perturbation (and the ensuing force response) following each cue. This resulted in 13 regressors per run, plus an intercept. Each regressor consisted of a delta function convolved with a two-gamma haemodynamic response function (HRF). We used a gridsearch approach to adjust the time to peak (4, 5, 6, 7, 8, and 9s) and the time to undershoot (10, 12, 14, 16, 18 and 20s) of the HRF and obtain the best fit to the BOLD timeseries. For preparation regressors, the delta function was placed at cue onset (in both go and no-go trials); for execution regressors, it was placed at perturbation onset (go trials only). Before GLM estimation, the BOLD time series were high-pass filtered with a standard cutoff frequency of 128s. The 1st-level GLM analysis resulted in activation images consisting of the fitted beta coefficients across voxels for each of the 13 trial types, for each run and participant.

Surface reconstruction and regions of interest definition

We used Freesurfer[48,49] to reconstruct the white–grey matter and pial surfaces from each participant's anatomical image. Each individual surface was then inflated to a sphere and nonlinearly aligned to the Freesurfer average atlas by matching cortical folding patterns, using sulcal depth and surface curvature to guide the alignment of gyri and sulci. Next, we resampled both hemisphere of each participant into a symmetric fs32k template, which represented each hemisphere using 32k vertices. In this way, by selecting corresponding vertices in each participant, it is possible to compare similar cortical regions.

We used a searchlight approach to assess the information represented across the entire cortical surface. For each vertex of the fs32k template, we defined a circular region of cortical grey matter (a "searchlight") with a 20mm diameter. We then fitted the relative weight of each representation of interest (see "Pattern component modelling") to the beta coefficients estimated in the 1st-level GLM within each searchlight. Then, we assigned the resulting weights to the centre vertex.

We used a probabilistic cytoarchitectonic atlas projected onto the group surface[49] to define eight anatomical ROI encompassing primary sensorimotor regions. M1 was defined by including all nodes belonging to Brodmann area 4 (BA4) within 2cm from the hand knob[50]. S1 was defined by selecting the nodes belonging to BA1, 2 and 3 within 2cm of the hand knob. We divided BA6 into a medial part (supplementary motor area, SMA), a lateral dorsal part (dorsal premotor cortex, PMd), and a ventral part (ventral premotor cortex, PMv). Finally, we separated the anterior and posterior parts of the superior parietal

lobule (SPLa and SPLp) approximately at the midpoint of the intraparietal sulcus. To avoid contamination of activity between different ROI across sulci, we excluded voxels with more than 10% of their volume lying in a neighbouring ROI.

Univariate analysis of fMRI data

To evaluate cortical activation during preparation and execution, we performed univariate contrasts of brain activity as compared to rest. We projected the individual beta coefficients to the group surface via the individual reconstructed surfaces. For visualization purposes (see Fig. 2c), we averaged the beta coefficients estimated across conditions, runs and participants. For statistical testing, we conducted a one-sample t-test of the average activation values in each ROI against 0 across participants.

Multivariate dissimilarity analysis

We used representational similarity analysis (RSA)[24–26] to assess task-related representations in neural (i.e., beta coefficients from the 1st-level GLM) and EMG activity patterns. Within each ROI or searchlight, we first performed a multivariate spatial pre-whitening on the neural activity patterns using the residuals from the 1st-level GLM:

$$\beta_{prewhitened} = \beta\Sigma^{-\frac{1}{2}} \quad (2)$$

Where β is the N (conditions x runs) by P (voxels) matrix of the estimated beta coefficients and Σ is the P-by-P noise covariance between voxels estimated from the residuals of the 1st-level GLM. The noise covariance Σ was regularised using the Ledoit-Wolf procedure[51] to ensure invertibility. Because voxels often show different (and correlated) levels of noise, the weighting of activation patterns by the inverse noise covariance makes dissimilarity estimates more reliable[27].

We used a similar pre-whitening procedure for EMG responses in the Pre, SLR, LLR and Vol time windows. For each time window and acquisition run, we first calculated the residuals of the EMG pattern observed in each condition relative to the mean across all conditions. We then used the variance of the residuals ($\sigma^{2(w)}$) estimated separately for each channel to perform a univariate pre-whitening:

$$EMG^{(w)}_{prewhitened} = \frac{EMG^{(w)}}{\sqrt{\sigma^{2(w)}}}, w \in \{Pre, SLR, LLR, Vol\} \quad (3)$$

Where $EMG^{(w)}$ is the N-(conditions x runs)-by-P (channels) matrix of the EMG activity in the w time window.

For both fMRI and EMG data, we then calculated the cross-validated squared Mahalanobis (crossnobis) dissimilarity d between conditions i and j as follows:

$$d_{i,j} = \frac{1}{M}\sum_{m}^{M}\left(x_i - x_j\right)_m^T\left(x_i - x_j\right)_{\sim m} \quad (4)$$

Where M is the number of runs, x_i and x_j are the pre-whitened (neural or EMG) activity patterns for conditions i and j, either from run m or averaged across all the other runs ($\sim m$). Cross-validation makes the dissimilarity estimates unbiased by measurement noise[27]. Because measurement noise pulls activity patterns in random directions, without cross-validation the expected dissimilarity between two activity patterns would be always larger than 0, even when they are identical and only differ by their noise. With cross-validation, the expected dissimilarity between identical patterns is 0, which means that we could test the average dissimilarity against 0 using a one-sided t-test to determine whether activity patterns carried

reliable information. Note that, especially when two patterns are very similar, cross-validated dissimilarity estimates can be negative[24,25,52].

Correlation between behavioural and BOLD dissimilarities

To ensure that the expectation representation in M1 and S1 during response preparation was not driven by subtle finger pre-activation, we first calculated the crossnobis dissimilarities between the mean force patterns for each probability cue across the five fingers measured in the 1.5s time interval before the perturbation. Then, we performed a linear regression analysis between the mean force dissimilarity and the mean neural dissimilarity observed during preparation in the same participant. If the dissimilarities between neural activity patterns for different probability cues are solely driven by finger pre-activation, then the intercept of the regression should not be significantly larger than 0; that is, participants that did not show finger pre-activation should also not show any BOLD dissimilarity. To test this, we performed a one-sided t-test of the intercept estimate against 0.

Pattern component modelling

We used Pattern component modelling (PCM)[53], a complementary framework to RSA, to characterise the nature of the representational geometry in neural and muscle activity patterns. Rather than estimating and evaluating the dissimilarities between activity patterns, PCM is a probabilistic framework that evaluates the marginal likelihood that the observed activity patterns have a multivariate Gaussian distribution of mean 0 and a covariance matrix G. Because we assumed a mean of 0, we did not subtract out the mean of each condition across voxels, thus G is more accurately called the second moment matrix of the distribution. The second moment matrix can be directly translated to squared Euclidean distances (D) through the equation:

$$D_{i,j} = G_{i,i} + G_{jj} - 2 * G_{i,j} \quad (5)$$

Where i and j are two different conditions (e.g., two different probability cues or stimulated fingers in the current experimental design). This allows us to visualize the representational models as RDMs (see Fig. 3a and 5a), while still using the more powerful[24] approach of PCM for model evaluation. Note that the squared Euclidean distance is mathematically equivalent to crossnobis dissimilarity when the activity patterns are pre-whitened[27].

Representational models for preparation

For the preparation epoch we considered two different representations of the probability cue . The expectation representation predicted that the activity of each voxel scaled linearly with a feature vector (f) corresponding to the difference in probability between index and ring:

$$f_{expectation} = [-1, -0.5, 0, 0.5, 1] \quad (6)$$

The uncertainty representation predicted that the activity patterns scaled with the variance of a Bernoulli distribution, defined as:

$$f_{uncertainty} = p(1-p) = [0, 0.1875, 0.25, 0.1875, 0] \quad (7)$$

Where p is the probability of a certain outcome (e.g., index stimulation) and $1-p$ is the probability of the opposite outcome (e.g., ring stimulation).

For each representational model, the predicted second moment matrix G was defined as the outer product of f:

$$G = ff^T \quad (8)$$

For visualization, we then calculated the predicted dissimilarity matrices of each model from the corresponding second moment matrices according to Eq. 5 (see Fig. 3a and 5a).

Representational models for execution

For execution, we designed three different representational models that predicted the neural or EMG response in 8 conditions, including 4 different probability levels (i.e., 25% to 100%) for each finger. The sensory input representation predicted that the data simply based on the identity of the stimulated finger and was defined as:

$$f_{input} = [-1, -1, -1, -1, 1, 1, 1, 1] \ (9)$$

With -1 and 1 indicating index and ring finger, respectively.

We also assessed the expectation representation, in which the activity of each voxel scaled with the difference between index and ring finger probability:

$$f_{expectation} = [-0.5, 0, 0.5, 1, -1, -0.5, 0, 0.5] \ (10)$$

The surprise representation predicted that the activity patterns scaled linearly with the Shannon surprise, defined as the negative log-probability of the observed event given the cue:

$$f_{surprise} = -log_2(p) \ (11)$$

Where $p = [1, 0.75, 0.5, 0.25, 0.25, 0.5, 0.75, 1]$, i.e., the probability cued on the stimulated finger. Note that the Shannon surprise reflects the absolute prediction error between finger and cue.

Model fitting and evaluation

In PCM, the second moment matrix (G) of the observed activity patterns is modelled as the linear combination of different representational models (G_h):

$$G = \sum_h \exp(\theta_h) G_h \ (12)$$

Where $\exp(\theta_h)$ is the weight parameter of the h^{th} representational model, and G_h the predicted second moment matrix. Because weights cannot be negative, using the exponential allows for unconstrained optimisation of the model parameters. To make the weights comparable, we normalised the trace of the predicted second moment matrices for each G_h to 1. In this way, the weight can be used to estimate the amount of variance explained by each representation included in the model. Importantly, this approach relies on the assumption that each representation spans an independent neural dimension. Therefore, the weight of each representation reflects how strongly the corresponding information is encoded in a certain brain region *independently* from all other representations in the model.

For the fMRI data, we fitted the relative weights of each representational model to the pre-whitened beta coefficients, separately for each task epoch (preparation and execution) and within each ROI or searchlight. We performed one-sided dependent-sample t-tests to compare the relative weight of different representations within the same ROI. Then, to directly compare the information encoded in different premotor-parietal vs. M1-S1, we averaged the weight of each representation within each ROI group and performed a one-sided dependent-sample t-test on the weight of a representation of interest (e.g., expectation in Fig. 3e and sensory input in Fig. 6d) relative to the sum of all representations of interest (e.g., expectation+uncertainty in Fig. 3e and sensory input+surprise in Fig. 6d).

To assess the expectation (and uncertainty) representation in the LFPs and spiking activity recorded from non-human primates in PMd and S1, we aligned the recordings to cue presentation and perturbation onset and then fitted the relative weight of each representation at each time point (and frequency band, for

the LFPs). Then, to assess whether the expectation encoding differed across each areas (i.e., PMd and S1) and recording modality (spiking activity and LFPs), we averaged the expectation weight over 0.64s (and between 10-20Hz for LFPs) after cue presentation (see grey-shaded time interval in Fig. 3g, lower panels) and performed a 2-by-2 repeated-measure ANOVA.

To establish whether a predicted representation contributed significantly to explain activity, we first fitted all the possible combinations of the candidate representations (e.g., expectation alone, uncertainty alone and expectation+uncertainty). Then, we calculated the log-Bayes factor (BF_F) of each representation, defined as the difference between the marginal log-likelihoods of the activity patterns under the models that included the representation and those that did not:

$$BF_F = \log \sum_{M:F=1} p(data|M) - \log \sum_{M:F=0} p(data|M) \quad (13)$$

Where F is the representation of interest (i.e., expectation, uncertainty, sensory input, or surprise) and M is a model including a combination of representations; $F = 1$ indicates that the representation of interest is included in M. The maximal likelihood of the observed activity under model M is given by $p(data|M)$ and was estimated using the Akaike Information Criterion (AIC), which corrects the maximal likelihoods for the model complexity[54].

For fMRI data, we tested the log-Bayes factor of each representational model against 0 using a one-sample t-test across participants in each ROI. A positive log-Bayes factor indicates that the representational component helped to explain the activity patterns in the context of all the other components.

For the LFPs and spiking activity, we used cluster-based permutations to establish the significance of the log-Bayes factor in the in different time (and frequency) bins, while controlling for multiple comparisons[55]. First, we calculated a one-sample t-statistic against 0 across sessions and defined clusters of contiguous significant time(-frequency) bins. We then generated a null distribution by randomly inverting the sign of the log-Bayes factor for a random subset of sessions and recomputing the t-statistic over 1,000 permutations. In each permutation, we thresholded the t-values at the uncorrected $p < 0.05$ level and recorded the cluster of contiguous significant bins with the largest weighed size, defined as the number of bins in the cluster multiplied by the sum of the absolute t-values within the cluster. The observed clusters were considered significant if their weighted size exceeded the 95th percentile of the null distribution.

Correlation model

In Experiment 1, we used a PCM correlation model to achieve correlation estimates between neural activity patterns during response preparation and execution unbiased by measurement noise. First, in each run, we contrasted the neural activity patterns elicited during preparation by sensory expectation cued on the index (i.e., 100:0% and 75:25% probability cues) vs. ring (0:100% and 25:75%) finger. A similar contrast was obtained for the execution epoch between the activity elicited by sensory input to the index vs. ring finger. Then, we estimated the correlation between contrasts for preparation and execution. The correlation model uses the repeated measures of the two contrasts (x and y) across runs to estimate their signal variances (σ_x^2 and σ_y^2) as well as the variance of the noise (σ_ϵ^2). The predicted second moment matrix ($\hat{G}$) of the true activity pattern then is:

$$\hat{G} = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} \quad (14)$$

We derived the maximum-likelihood estimate for the correlation between the true activity patterns both for each individual participant and for the group (see black dots and dashed horizontal red line, respectively, in

Fig. 4 and 7f-i). To obtain confidence intervals for the group estimate, we conducted a bootstrap procedure, resampling the participants with replacement at each iteration. We performed 1000 iterations and calculated the 95% central confidence interval of the correlation estimate (see dashed grey areas in Fig. 4 and 7f-i).

We used a similar approach to estimate the correlation between expectation and sensory input in BOLD, EMG, LFPs and spiking activity during execution. In this case, we calculated the contrasts between execution activity following the 75-25% vs. 25-75% probability cues and between index vs. ring finger (or flexion vs. extension, for LFPs and spiking activity in monkeys) perturbation. Then, we obtained maximum-likelihood estimates of the correlation between the two contrasts using PCM. For the EMG data, we estimated the correlation using the pre-whitened (see Eq. 3) mean activity patterns in four separate time windows, corresponding to the 100ms before perturbation onset (Pre), 25-50ms after perturbation (SLR), 50-100ms after perturbation (LLR) and 100-500ms after perturbation (Vol). For the same correlation in LFPs and spiking activity, we used the pre-whitened (as done for EMG, see Eq. 3) mean patterns across electrodes (LFPs) or units (spiking activity) between 40-240ms after the perturbation. For the LFPs, separate correlations were estimated in the alpha (8-13Hz), beta (13-25Hz) and gamma (25-100Hz) frequency bands.

To diagnose the reliability of our correlation estimates, we plotted the correlation estimates against the estimated functional signal-to-noise ratio (fSNR) of the two contrasts, defined as:

$$fSNR = \frac{\sqrt{\sigma_x^2 * \sigma_y^2}}{\sigma_\epsilon^2} \quad (15)$$

At low fSNR, maximum-likelihood correlation estimates become unstable and may fall at the parameter bounds (i.e., correlation=±1). For example, in the Pre epoch in EMG recordings (see Fig. 7f), the fSNR was close to 0 and the preparation-execution correlation estimate was -1 for the majority of the participants. This scenario flags the correlation estimate as unreliable.

Low-dimensional projections of BOLD activity patterns

We used a multidimensional scaling approach to obtain the low-dimensional projections of BOLD activity show in Fig. 8a. This corresponds to performing an eigen-decomposition of the second moment matrix of the observed activity patterns, yielding orthogonal dimensions in voxel space that capture most of the variance across conditions. In this way, the projection of the activity patterns onto the first two principal components preserves the dominant representational geometry of the second moment matrix and can be used for visualisation.

Electrophysiological recordings in non-human primates

The spiking data from the non-human primate electrophysiological datasets are publicly available[22] and described in detail in a recently published paper[9]. In brief, monkeys countered mechanical perturbations delivered with a KINARM robot exoskeleton (BKIN Technologies)[56] that could rotate the elbow either in flexion or extension direction. Before the perturbation, the monkeys received probabilistic information about the upcoming perturbation (0:100%, 25:75%, 50:50%, 75:25%, 100:0%, flexion:extension). Electrophysiological recordings were carried out using high-density Neuropixels probes (1.0 - 1 cm, 1.0 NHP - 1 cm, and 1.0 NHP - 4.5 cm), pre-processed using a custom pipeline (https://github.com/JonathanAMichaels/PixelProcessingPipeline), and spike sorted with Kilosort 2.0[57]. The LFPs were read from the Neuropixels LF stream, which was recorded at 2,500Hz. We downsampled the initial 384 channels to a total of 32 channels and then performed a time-frequency analysis using the

FieldTrip toolbox[58]. We defined fifty frequencies of interest logarithmically (log10) spaced from 1 Hz to 400 Hz. The time bins of interest were sampled at a resolution of 0.01 seconds. After the LFPs were demeaned and a bandpass filtered (1-400 Hz, 3rd order), we calculated the power spectrum in each time bin using the multi-taper convolution method with a Hanning taper. The time windows for the convolution were dynamically adjusted relative to the frequencies of interest to cover 5 cycles at each frequency. For the subsequent analysis, we pooled the recording sessions from both Monkeys (PMd, 17 sessions; M1, 9 sessions; S1, 9 sessions) totalling ~35000 trials overall.

Acknowledgements

Experiments were funded by a project grant from the Canadian Institutes of Health Research (CIHR, PJT – 175010) to JD and AP, a Discovery grant from the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2016-04890) to JD, and the Canada First Research Excellence Fund (BrainsCAN) to Western University. AP received a salary award from the Canada Research Chairs Program.

References

1. Scott, S. H. A Functional Taxonomy of Bottom-Up Sensory Feedback Processing for Motor Actions. *Trends Neurosci.* **39**, 512–526 (2016).
2. Cracco, R. Q. The initial positive potential of the human scalp-recorded somatosensory evoked response. *Electroencephalogr. Clin. Neurophysiol.* **32**, 623–629 (1972).
3. Inui, K., Wang, X., Tamura, Y., Kaneoke, Y. & Kakigi, R. Serial Processing in the Human Somatosensory System. *Cereb. Cortex* **14**, 851–857 (2004).
4. Rosenbaum, D. A. Human movement initiation: specification of arm, direction, and extent. *J. Exp. Psychol. Gen.* **109**, 444–474 (1980).
5. Churchland, M. M., Yu, B. M., Ryu, S. I., Santhanam, G. & Shenoy, K. V. Neural Variability in Premotor Cortex Provides a Signature of Motor Preparation. *J. Neurosci.* **26**, 3697–3712 (2006).
6. Keele, S. W. Movement control in skilled motor performance. *Psychol. Bull.* **70**, 387–403 (1968).
7. Haith, A. M., Pakpoor, J. & Krakauer, J. W. Independence of Movement Preparation and Movement Initiation. *J. Neurosci.* **36**, 3007–3015 (2016).
8. Klapp, S. T. & Erwin, C. I. Relation between programming time and duration of the response being programmed. *J. Exp. Psychol. Hum. Percept. Perform.* **2**, 591–598 (1976).
9. Michaels, J. A. *et al.* Sensory expectations shape neural population dynamics in motor circuits. *Nature* 1–10 (2025) doi:10.1038/s41586-025-09690-9.
10. Bestmann, S. *et al.* Influence of Uncertainty and Surprise on Human Corticospinal Excitability during Preparation for Action. *Curr. Biol.* **18**, 775–780 (2008).
11. Harrison, L. M., Duggins, A. & Friston, K. J. Encoding uncertainty in the hippocampus. *Neural Netw.* **19**, 535–546 (2006).
12. Wong, A. L. & Haith, A. M. Motor planning flexibly optimizes performance under uncertainty about task goals. *Nat. Commun.* **8**, 14624 (2017).
13. Alhussein, L. & Smith, M. A. Motor planning under uncertainty. *eLife* **10**, e67019 (2021).
14. Adams, R. A., Shipp, S. & Friston, K. J. Predictions not commands: active inference in the motor system. *Brain Struct. Funct.* **218**, 611–643 (2013).
15. Aitchison, L. & Lengyel, M. With or without you: predictive coding and Bayesian inference in the brain. *Curr. Opin. Neurobiol.* **46**, 219–227 (2017).

16. Adams, R. A., Friston, K. J. & Bastos, A. M. Active Inference, Predictive Coding and Cortical Architecture. in *Recent Advances on the Modular Organization of the Cortex* (eds Casanova, M. F. & Opris, I.) 97–121 (Springer Netherlands, Dordrecht, 2015). doi:10.1007/978-94-017-9900-3_7.

17. Rouhani, N. & Niv, Y. Signed and unsigned reward prediction errors dynamically enhance learning and memory. *eLife* **10**, e61077 (2021).

18. Goense, J. B. M. & Logothetis, N. K. Neurophysiology of the BOLD fMRI Signal in Awake Monkeys. *Curr. Biol.* **18**, 631–640 (2008).

19. Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150–157 (2001).

20. Logothetis, N. K. The Underpinnings of the BOLD Functional Magnetic Resonance Imaging Signal. *J. Neurosci.* **23**, 3963–3971 (2003).

21. Buzsáki, G., Anastassiou, C. A. & Koch, C. The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* **13**, 407–420 (2012).

22. Michaels, J. A. & Pruszynski, J. A. Data from: Sensory expectations shape neural population dynamics in motor circuits. 1694614743 bytes Dryad https://doi.org/10.5061/DRYAD.0VT4B8HBR (2025).

23. Arbuckle, S. A., Pruszynski, J. A. & Diedrichsen, J. Mapping the Integration of Sensory Information across Fingers in Human Sensorimotor Cortex. *J. Neurosci.* **42**, 5173–5185 (2022).

24. Diedrichsen, J. & Kriegeskorte, N. Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLOS Comput. Biol.* **13**, e1005508 (2017).

25. Bosch, J. J. F. van den *et al.* A Python Toolbox for Representational Similarity Analysis. 2025.05.22.655542 Preprint at https://doi.org/10.1101/2025.05.22.655542 (2025).

26. Kriegeskorte, N. Representational similarity analysis – connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* https://doi.org/10.3389/neuro.06.004.2008 (2008) doi:10.3389/neuro.06.004.2008.

27. Walther, A. *et al.* Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage* **137**, 188–200 (2016).

28. Kaufman, M. T., Churchland, M. M., Ryu, S. I. & Shenoy, K. V. Cortical activity in the null space: permitting preparation without movement. *Nat. Neurosci.* **17**, 440–448 (2014).

29. Churchland, M. M. & Shenoy, K. V. Preparatory activity and the expansive null-space. *Nat. Rev. Neurosci.* **25**, 213–236 (2024).

30. Liddell, E. G. T. & Sherrington, C. S. Reflexes in response to stretch (myotatic reflexes). *Proc. R. Soc. Lond. Ser. B Contain. Pap. Biol. Character* **96**, 212–242 (1924).

31. Hammond. The influence of prior instruction to the subject on an apparently involuntary neuromuscular response. *J Physiol* **132**, 17–19 (1956).

32. Pruszynski, J. A., Kurtzer, I. & Scott, S. H. Rapid Motor Responses Are Appropriately Tuned to the Metrics of a Visuospatial Task. *J. Neurophysiol.* https://doi.org/10.1152/jn.90262.2008 (2008) doi:10.1152/jn.90262.2008.

33. Cheney, P. D. & Fetz, E. E. Corticomotoneuronal cells contribute to long-latency stretch reflexes in the rhesus monkey. *J. Physiol.* **349**, 249–272 (1984).

34. Reschechtko, S. & Pruszynski, J. A. Stretch reflexes. *Curr. Biol.* **30**, R1025–R1030 (2020).

35. Russo, A. A. *et al.* Motor Cortex Embeds Muscle-like Commands in an Untangled Population Response. *Neuron* **97**, 953-966.e8 (2018).

36. Gallego, J. A., Perich, M. G., Miller, L. E. & Solla, S. A. Neural Manifolds for the Control of Movement. *Neuron* **94**, 978–984 (2017).

37. Umeda, T., Isa, T. & Nishimura, Y. The somatosensory cortex receives information about motor output. *Sci. Adv.* **5**, eaaw5388 (2019).

38. Ariani, G., Pruszynski, J. A. & Diedrichsen, J. Motor planning brings human primary somatosensory cortex into action-specific preparatory states. *eLife* **11**, e69517 (2022).

39. Ariani, G., Shahbazi, M. & Diedrichsen, J. Cortical Areas for Planning Sequences before and during Movement. *J. Neurosci.* **45**, (2025).

40. Gale, D. J., Flanagan, J. R. & Gallivan, J. P. Human Somatosensory Cortex Is Modulated during Motor Planning. *J. Neurosci.* **41**, 5909–5922 (2021).

41. Fries, P. Rhythms for Cognition: Communication through Coherence. *Neuron* **88**, 220–235 (2015).

42. Han, M.-J. *et al.* Whole-brain effective connectivity of the sensorimotor system using 7 T fMRI with electrical microstimulation in non-human primates. *Prog. Neurobiol.* **250**, 102760 (2025).

43. Jones, E. G., Coulter, J. D. & Hendry, S. H. C. Intracortical connectivity of architectonic fields in the somatic sensory, motor and parietal cortex of monkeys. *J. Comp. Neurol.* **181**, 291–347 (1978).

44. Keller, G. B. & Mrsic-Flogel, T. D. Predictive Processing: A Canonical Cortical Computation. *Neuron* **100**, 424–435 (2018).

45. Feulner, B., Perich, M. G., Miller, L. E., Clopath, C. & Gallego, J. A. A neural implementation model of feedback-based motor learning. *Nat. Commun.* **16**, 1805 (2025).

46. Hutton, C. *et al.* Image Distortion Correction in fMRI: A Quantitative Evaluation. *NeuroImage* **16**, 217–240 (2002).

47. Friston, K. J. *et al.* Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* **2**, 189–210 (1994).

48. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage* **9**, 179–194 (1999).

49. Fischl, B. *et al.* Cortical Folding Patterns and Predicting Cytoarchitecture. *Cereb. Cortex* **18**, 1973–1980 (2008).

50. Yousry, T. A. *et al.* Localization of the motor hand area to a knob on the precentral gyrus. A new landmark. *Brain* **120**, 141–157 (1997).

51. Ledoit, O. & Wolf, M. Honey, I Shrunk the Sample Covariance Matrix.

52. Diedrichsen, J., Provost, S. & Zareamoghaddam, H. On the distribution of cross-validated Mahalanobis distances. Preprint at https://doi.org/10.48550/arXiv.1607.01371 (2016).

53. Diedrichsen, J., Yokoi, A. & Arbuckle, S. A. Pattern component modeling: A flexible approach for understanding the representational structure of brain activity patterns. *NeuroImage* **180**, 119–133 (2018).

54. Yokoi, A. & Diedrichsen, J. Neural Organization of Hierarchical Motor Sequence Representations in the Human Neocortex. *Neuron* **103**, 1178-1190.e7 (2019).

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

56. Scott, S. H. Apparatus for measuring and perturbing shoulder and elbow joint positions and torques during reaching. *J. Neurosci. Methods* **89**, 119–127 (1999).

57. Pachitariu, M., Sridhar, S., Pennington, J. & Stringer, C. Spike sorting with Kilosort4. *Nat. Methods* **21**, 914–921 (2024).

827 58. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open Source Software for Advanced
828 Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Comput. Intell. Neurosci.* **2011**, 156869
829 (2011).

830