

Bayesian adaptive design for neural recruitment curves

Vishweshwar Tyagi^{1*}, Saihari S. Dukkupati¹, Daniel Twili²,
Jonas M. Mikhaeil², Andrew Gelman², Jason B. Carmel^{1†},
James R. McIntosh^{1*†}

^{1*}Department of Neurology, Columbia University, New York, NY, 10032.

²Department of Statistics, Columbia University, New York, NY, 10027.

*Corresponding author(s). E-mail(s): vt2353@cumc.columbia.edu;
jrm2263@cumc.columbia.edu;

[†]Joint senior authors.

Abstract

Nonlinear response models remain ubiquitous across disciplines with recruitment curves describing muscle response to neural stimulation. Conventional acquisition methods typically rely on static uniform sampling, requiring many stimuli, while adaptive threshold-hunting designs such as ML-PEST forfeit accurate characterization of the overall curve. We present a Bayesian adaptive design that uses posterior uncertainty over curve parameters to propose stimuli that maximize information gain across the full recruitment curve. We additionally implement a focused acquisition method that can target arbitrary subset of curve parameters, which we use to evaluate and interpret our design. In simulations, our design outperforms uniform sampling in recovering threshold while matching ML-PEST, and outperforms ML-PEST in predictive curve accuracy while matching uniform. It likewise matches our threshold-focused acquisition, suggesting our design doesn't trade off threshold estimation for recovering the curve. We further show data collected using our design yields greater power when detecting a change in threshold in intervention studies. Our design proposes the next stimulus within 5 seconds, making it feasible for transcranial magnetic stimulation (TMS) experiments. Finally, we validate our design in TMS experiments and provide a library for Python, called `onMEP`, for online acquisition of neural recruitment curves.

Keywords: optimal experimental design, expected information gain, Bayesian inference, motor threshold, transcranial magnetic stimulation, spinal cord stimulation

1 Introduction

Neural recruitment curves characterize nonlinear muscle response to neural stimulation. They are derived using electromagnetic and electrical stimulation techniques, such as transcranial magnetic stimulation (TMS), spinal cord stimulation (SCS), and nerve stimulation, which are used to assess and modulate the neural systems that control movement [1–14]. When targeted to the sensorimotor system, these techniques produce motor-evoked potentials (MEPs) (Fig. 1a), which are quantified by their size, typically in terms of peak-to-peak amplitude or area under the curve (Fig. 1b, gray and blue dots). The recruitment curve (Fig. 1b, black curve) models the MEP size as a non-linear, non-decreasing function of stimulation intensity and exhibits a characteristic S-shape, with a steep increase above the threshold leading up to a plateau phase at high intensities (Fig. 1b, lower and side panels). Whereas parameters such as the slope and saturation may have specific neurophysiological interpretations, the threshold remains a critical parameter for inferring changes in corticospinal excitability [15–19], while the overall recruitment curve characterizes the total corticospinal output over the stimulation range [20–27].

In this paper, we consider the design of experiments for acquiring recruitment curves. Specifically, we consider designs that sequentially determine the stimulation intensity at which to acquire the next MEP sample. Collecting data using TMS and SCS can be a costly and burdensome process for both the participant and the experimenter [12, 19, 28–33], and our requirements for an efficient design are therefore high accuracy in recovering both the threshold and the overall curve using a limited number of stimuli, typically fewer than 40.

Conventional methods for recruitment curve acquisition include static designs that use uniform sampling, which requires many stimuli for accurate threshold estimation and results in samples located in non-informative regions of the curve, such as well below the threshold (Fig. 1b, left). Ideally, we would reallocate these samples to more informative regions along the rising portion of the curve, such as near the threshold and in the upper region transitioning to saturation. Moreover, such designs lack a continuous feedback loop during acquisition and require the experimenter to predetermine the number of stimuli to be tested. Consequently, experiments can neither terminate early when the curve has been estimated with sufficient certainty nor continue longer when additional samples are needed in the presence of high measurement noise. Other designs exist, most notably the adaptive threshold-hunting design known as maximum likelihood parameter estimation by sequential testing (ML-PEST) [34–39]. However, ML-PEST is designed specifically for estimating the resting motor threshold and consequently forfeits accurate characterization of the overall curve.

In contrast, Bayesian experimental design provides a principled information-theoretic framework for dynamically adapting experimental decisions as data is acquired [40–43]. We introduce a Bayesian adaptive design that uses posterior uncertainty over curve parameters to propose stimuli that maximize the expected information gain (EIG) across the recruitment curve (Fig. 1b, right). It concentrates samples in informative regions of the curve, focusing on the threshold in the early phase and subsequently sampling the transition regions to estimate the slope and saturation (Fig. 1c). We use a generative hierarchical framework to simulate high-fidelity

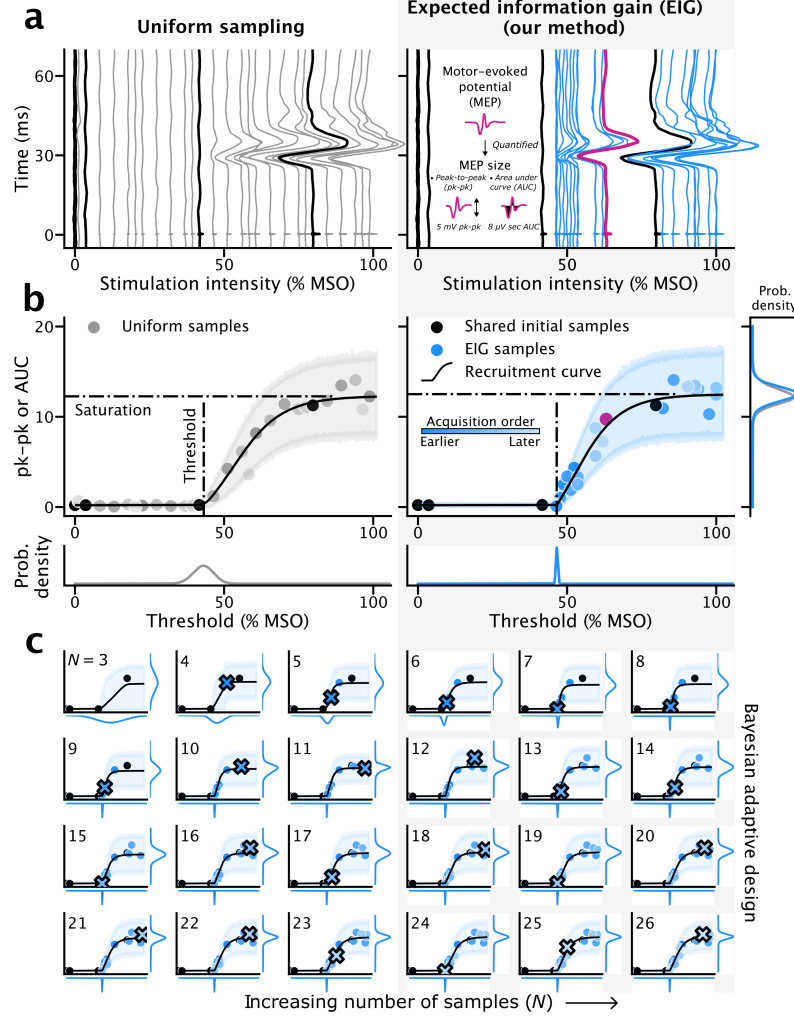


Figure 1. Bayesian adaptive acquisition of recruitment curves using posterior uncertainty over curve parameters to maximize information gain. (a) Example MEPs from a real participant’s APB muscle in TMS experiment, recorded at uniform intensities (left), and at intensities proposed by the expected information gain (EIG) design (right). Both uniform and EIG are block randomized and start with the same initial samples (black dots)—eight samples at zero intensity (not counted) and three uniform samples. The abscissa is specified as percent maximum stimulator output (% MSO) for TMS but may instead use units of current, such as μ A or mA, for spinal cord stimulation. Inset: schematic quantification of MEPs using either peak-to-peak (pk-pk) amplitude or area under the curve (AUC). (b) Recruitment curves estimated from uniform and EIG samples at the end of the experiment, with 26 samples each. Samples are colored by acquisition order from earlier (dark) to later (light). EIG yields a narrower threshold posterior (bottom), indicating greater certainty in threshold estimate, and allocates samples to more informative regions of the curve. Side panel: posterior distribution of saturation. (c) Starting with three initial samples, the EIG design updates the posterior over curve parameters and uses it to propose the intensity for the next sample (cross mark). It samples near the threshold early and later samples the transition regions to estimate the slope and saturation. Accordingly, the threshold posterior (bottom) narrows early, while the saturation posterior (right) narrows later.

synthetic data that closely matches empirical TMS observations for model comparison [44]. In simulations, we compare the number of stimuli needed to reach acceptable error in recovering the motor threshold and the recruitment curve. We additionally implement a focused version of our method that can target information gain in an arbitrary subset of curve parameters, which we use to evaluate and interpret our design. To ensure practical feasibility, we optimize our design to reduce the time to choose the next stimulus and evaluate its speed and accuracy when acquiring either a single target muscle or multiple muscles simultaneously. For a use case, we show data collected using our design yields greater statistical power and requires fewer participants to detect a change in the threshold in intervention studies. Finally, we validate our design in human TMS experiments and provide a library for Python, called `onMEP`, for online acquisition of neural recruitment curves.

2 Results

2.1 Accurate estimation on sparse data

We previously introduced a Bayesian model for neural recruitment curves [44], which we extended into a Bayesian adaptive design that sequentially selects stimulation intensities based on the data acquired during experiment (Methods 4.1). The design selects the next intensity by maximizing the expected information gain (EIG), which measures how much a new observation is expected to reduce uncertainty about the recruitment curve. We additionally implemented focused EIG that can target information gain in an arbitrary subset of curve parameters, allowing the design to target a specific parameter, such as the threshold, or the full curve, depending on the choice of subset. To enable a comparative analysis of acquisition methods, we used a hierarchical model estimated on 14 uninjured TMS participants to simulate synthetic data that closely matched empirical observations [44]. We evaluated EIG for its accuracy in recovering the curve parameters and compared it against three methods—the conventional methods of uniform sampling (U) and ML-PEST (ML), as well as against EIG focused on the threshold (Focused Information Gain or FIG). In particular, we used each design to acquire data for a given recruitment curve and subsequently used our Bayesian model to estimate the curve parameters from the acquired data. By keeping the estimation model fixed, we could discriminate between the designs based on their acquired data (Methods 4.3).

EIG demonstrated improved accuracy in recovering both the threshold and the curve when acquiring a single target muscle (Fig. 2a,b, left panels). Beginning with the same three initial samples, the threshold error was identical across designs. By 16 stimuli, EIG achieved lower threshold error than uniform ($e_{\text{EIG}} - e_{\text{U}} = -1.87 \pm 0.03$) and ML (-0.67 ± 0.06). FIG had the lowest error, although it was comparable to EIG ($e_{\text{EIG}} - e_{\text{FIG}} = 0.13 \pm 0.02 \in (-0.5, 0.5)\%$ MSO). The improvement over uniform was largest at this relatively low number of stimuli, with the gap narrowing at 64 stimuli, albeit still not comparable ($e_{\text{EIG}} - e_{\text{U}} = -0.56 \pm 0.02 \notin (-0.5, 0.5)\%$ MSO). EIG likewise improved recovery of the overall curve and achieved predictive error comparable to uniform at 32 stimuli ($0.36 \pm 0.05 \in (-0.5, 0.5)$), whereas ML (-11.41 ± 0.17) performed the worst, followed by FIG (-8.17 ± 0.14), illustrating the

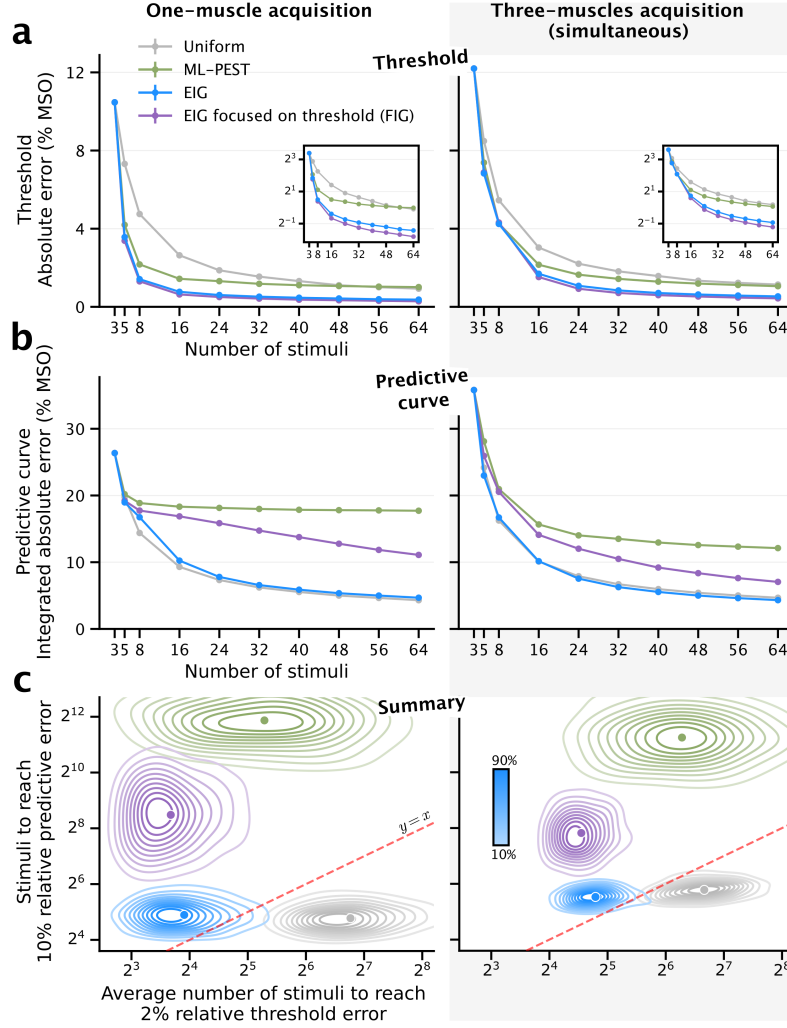


Figure 2. EIG simultaneously improves threshold and recruitment curve estimation on sparse data. (a) Mean absolute error in threshold estimation as a function of number of stimuli when acquiring a single target muscle (left) and three muscles simultaneously, with error averaged across muscles (right). Error bars represent standard error of the mean. Inset: zoom to show error on logarithmic scale. (b) Mean integrated absolute error for the estimated predictive curve. (c) Joint distribution of the average number of stimuli needed to reach relative threshold error of 2% and relative predictive error of 10%, estimated from experiments with 16 participants. Contours span the (10, 90)% iso-proportion interval, and midpoints indicate the mean. The dashed diagonal line indicates equal numbers of stimuli needed to reach both error criteria simultaneously.

trade-off between acquiring for a single parameter and learning the full recruitment curve.

For most practical use cases, EIG minimized the number of stimuli needed to jointly reach acceptable accuracy in threshold and curve estimation (Fig. 2c, left panel). We

estimated the number of stimuli required on average to reach a relative threshold error of 2% and relative predictive error of 10% by sampling experiments with 16 participants, a sample size representative of typical studies (Methods 4.3). EIG matched ML and FIG in the number of stimuli required for threshold recovery, as indicated by their overlap along the threshold axis, while matching uniform in curve recovery, indicated by their overlap along the predictive axis. Additionally, during early acquisition, stimuli proposed by EIG matched those of FIG, with their acquisition functions having similar shapes and peaking around the same stimulus values close to threshold (Supplementary Fig. S1). Thus, EIG is comparable to designs focused on threshold while retaining accurate recovery of the overall curve, and choosing a different design would require more stimuli to reach either criterion. Note in the limiting case of a large number of participants, we expect these distributions to shrink towards their mean values (the midpoints), consistent with the comparison shown above (Fig. 2a,b).

2.2 Extension to multiple muscles

In TMS and SCS experiments, MEPs are often recorded from multiple muscles simultaneously, as opposed to a single target muscle [9, 12, 14, 45–50]. In these cases, an efficient design must select the next stimulus by optimizing acquisition over multiple muscles. EIG naturally combines information across different muscles into a single acquisition function to jointly maximize information gain (see Methods 4.2), even when the muscles differ in response scales, as can occur with differences in cortical representation of muscles in TMS or electrode placement in SCS. Since ML-PEST has no analogous acquisition function, we instead paired it with a round-robin strategy across muscles.

The relative advantages of EIG were preserved when acquiring three muscles simultaneously, with EIG achieving lower average error across muscles for both threshold and curve estimation (Fig. 2a–c, right panels). As expected, the number of stimuli required to reach acceptable threshold error increased across all designs because acquisition now spanned multiple muscles. In contrast, simultaneous acquisition improved curve estimation for the focused designs. This is because when muscles have sufficiently separate thresholds, samples allocated near the threshold of one muscle can fall within the recruitment region of another, thereby providing additional information about its curve as opposed to single muscle acquisition. This also benefits EIG during early acquisition, when samples are concentrated around the threshold, as evident from EIG and uniform achieving comparable predictive error at a relatively low number of 16 stimuli ($e_{\text{EIG}} - e_{\text{U}} = 0.05 \pm 0.06 \in (-0.5, 0.5)$). Nonetheless, EIG continued to minimize the number of stimuli required to jointly reach acceptable error in threshold and curve estimation (Fig. 2c, right panel).

2.3 Online acquisition feasibility

EIG consists of two computational steps between consecutive stimuli—the update step, which estimates the posterior distribution over the curve parameters after collecting a new observation, and the proposal step, which computes the acquisition function over candidate stimuli to choose the next stimulus. The update step uses Markov

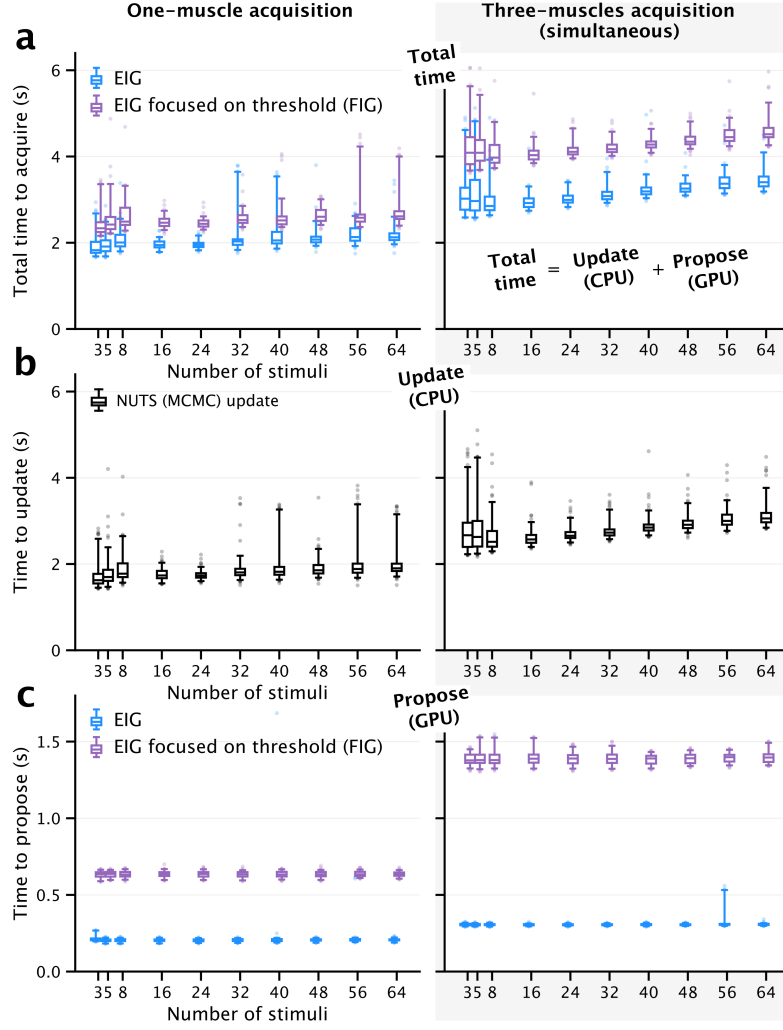


Figure 3. EIG proposes the next stimulus within 5 seconds, making it feasible for online TMS experiments. (a) Total computation time between consecutive stimuli for EIG and EIG focused on threshold when acquiring a single target muscle (left) and three muscles simultaneously (right). Boxes indicate the median time and span the interquartile range, whiskers span the (2.5, 97.5) percentiles, and dots indicate outliers. (b) Computation time for the posterior update step using the No-U-Turn sampler (NUTS), which accounts for most of the total computation time and gradually increases with the number of stimuli. (c) Computation time for the proposal step for EIG and EIG focused on threshold, which remains constant with the number of stimuli.

chain Monte Carlo (MCMC) methods, such as the No-U-Turn sampler (NUTS) [51], whereas the acquisition function can be computed across candidate stimuli in parallel. To account for this, we implemented posterior updating on the CPU and acquisition

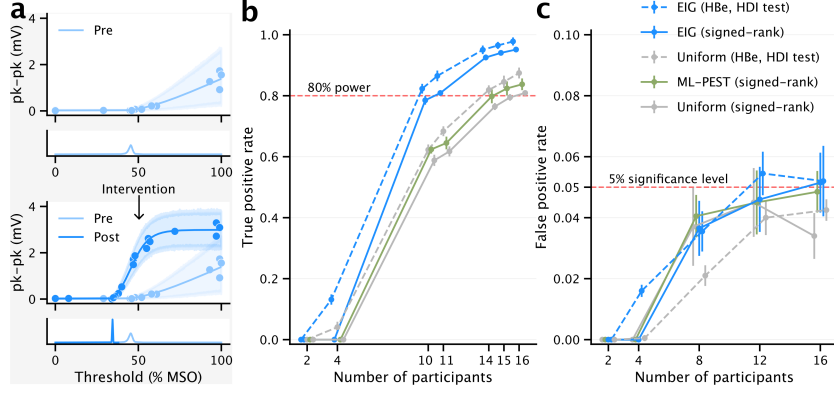


Figure 4. EIG increases statistical power and requires fewer participants when detecting a shift in threshold. (a) Example pre- and post-intervention recruitment curves with the intervention resulting in a negative shift in threshold. Bottom panel: threshold posterior. (b) True positive rate as a function of the number of participants. Here the threshold differences are simulated from a normal($\mu = -3, \sigma = 2.5$) distribution, where the alternative hypothesis (non-zero shift) is true. (c) Comparison of false positive rates against the set significance level of 5%. The differences are simulated from normal($0, 2.5$), where the null hypothesis (zero shift) is true. All simulations consisted of 16 samples each for the pre- and post-intervention phases, not counting eight samples at zero intensity. Error bars represent the standard error of the mean rate.

evaluation on the GPU within our Python library `onMEP`. For multiple-muscle acquisition, the update and proposal steps are further parallelized across muscles, such that, with sufficient compute, the required time should approach that of a single muscle.

EIG proposes the next stimulus within 5 seconds, making it feasible for online TMS experiments (Fig. 3). For one-muscle acquisition, the total computation time for EIG remained below 3 seconds, with posterior updating accounting for most of the computation time (Fig. 3a–c, left panels). The update time increases gradually as observations accumulate, whereas proposal time remains constant because the posterior dimension and candidate stimuli remain fixed. For simultaneous three-muscle acquisition, parallelization across muscles limits the total time to below 4 seconds despite tripling the number of recruitment curves being acquired (Fig. 3a–c, right panels). Thus, both EIG and focused EIG remain below 5 seconds throughout acquisition in either case, with focused EIG requiring more time due to its costlier acquisition function. These computation times fall well within the typical interstimulus interval of at least 5 seconds used in TMS experiments [31, 52–54].

2.4 Detecting a shift in threshold

Changes in threshold are used to assess changes in corticospinal excitability in intervention studies [15–19]. We evaluated how the acquisition designs perform in this case by simulating two hypothetical scenarios—an intervention resulting in a negative shift in threshold, indicating increased excitability (Fig. 4a), and an intervention resulting in no shift. We used each design to acquire data for the pre- and post-intervention recruitment curves and subsequently tested for a shift in threshold using a paired two-sided signed-rank test on the point threshold estimates at the 5% significance level.

We additionally used hierarchical Bayesian estimation (HBe), which directly models participant-level differences in threshold between the pre- and post-intervention phases and summarizes these differences across participants using a population-level location parameter [44], with its 95% highest density interval (HDI) used for hypothesis testing [44, 55–58]. For HBe, the decision rule rejects the null hypothesis (zero shift) in favor of the alternative (non-zero shift) if the 95% HDI excludes zero. This allowed us to evaluate whether the non-uniform samples acquired by EIG would affect HBe inference either by reducing statistical power or inflating the false positive rate.

EIG paired with Bayesian estimation required the fewest participants to detect a shift in threshold (Fig. 4b,c). Using the signed-rank test, uniform sampling required 16 participants to reach 80% power (80.9 ± 0.9), compared with 15 for ML-PEST (82.4 ± 2.3 ; 79.8 ± 2.3 with 14 participants), and only 11 for EIG (80.9 ± 0.8), a 26.7% reduction over ML-PEST. HBe further increased statistical power with uniform sampling requiring 14 participants (81.9 ± 1.5), one less than ML-PEST paired with the signed-rank test. Finally, EIG paired with HBe required only 10 participants (82.3 ± 1.6), albeit with a smaller improvement over the signed-rank test, likely because EIG samples already enable accurate threshold estimation at the participant level. Furthermore, the false positive rates did not exceed the 5% significance level for any method, validating that the differences in statistical power were not because of an inflated false positive rate.

2.5 Validation in TMS experiments

Figure 5a shows data acquired using EIG from eight neurotypical participants without evidence of neurological injury enrolled in an ongoing TMS study. The early adaptive samples (dark blue) are concentrated around the threshold, with later samples (light blue) extending across the recruitment curve. For participants P7 and P8, with high thresholds exceeding 80% MSO, EIG could still locate their thresholds despite all initial samples (black dots) falling within the baseline response, and subsequently allocate samples along the rising portion of their recruitment curves, even with the limited stimulation range available beyond 80% MSO. This is desirable when tuning stimulation parameters, as it can quickly identify configurations for which the threshold lies beyond the available stimulation range and may therefore require adjustment. In contrast, uniform sampling would result in most samples within the baseline response, providing little information about either their thresholds or the rising portion of the curves.

To determine a sample size for a study comparing the acquisition methods, we performed a prospective power analysis by bootstrapping data from the eight participants (Fig. 5b–e). Because data were collected only using EIG, we used our model to estimate a generative distribution for each participant and sampled from it to evaluate each acquisition method (Methods 4.3). Consistent with our earlier results, EIG improved both threshold and recruitment curve estimation (Fig. 5b,c). Based on our simulations, we defined two primary null hypotheses—EIG not differing from uniform sampling in the number of stimuli needed for threshold estimation (Fig. 5d, gray curve), and not differing from ML-PEST in stimuli needed for curve estimation (green curve). The study was considered successful if both null hypotheses were rejected simultaneously

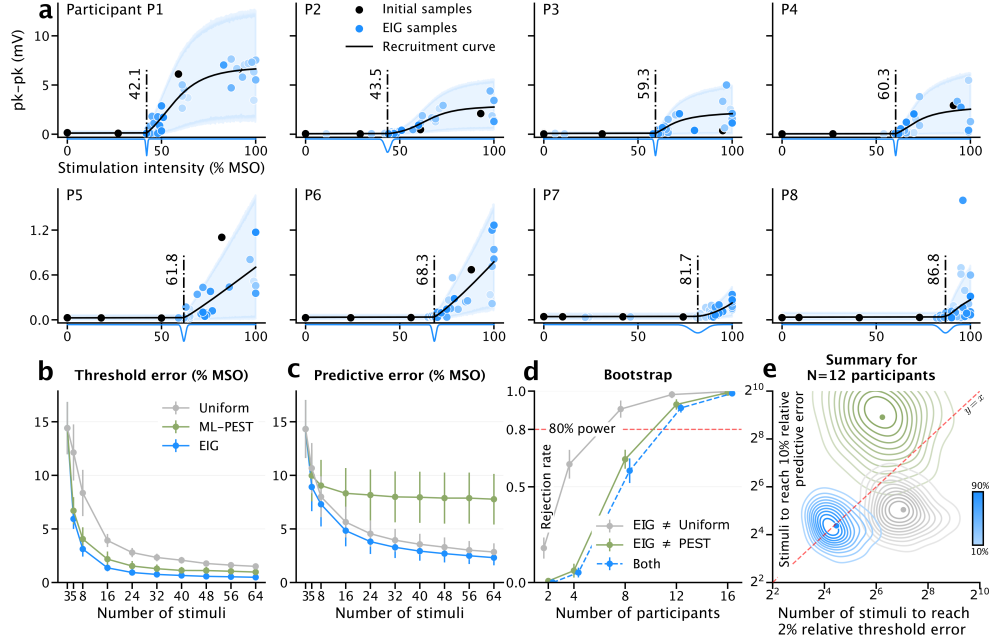


Figure 5. Validation of EIG in real TMS experiments. (a) Recruitment curves acquired using EIG from APB muscle of eight participants enrolled in TMS study, with 40 stimuli each, not counting eight samples at zero intensity. Samples are colored by acquisition order from earlier (dark) to later (light). Bottom: threshold posterior. Participants are ordered by increasing threshold. (b) Mean absolute error in threshold estimation. Error bars represent the standard error across participants. (c) Mean integrated absolute error in recruitment curve estimation. (d) Rejection rates for the null hypotheses of no difference between EIG and uniform sampling in the number of stimuli needed to reach 2% relative threshold error (gray curve), between EIG and ML-PEST in stimuli needed to reach 10% relative predictive error (green curve), and for both hypotheses simultaneously (dashed blue curve). (e) Joint distribution of the number of stimuli needed to reach acceptable threshold and predictive errors with $N=12$ participants. Contours span the (10, 90)% iso-proportion interval, and midpoints indicate the mean. The dashed diagonal line indicates equal numbers of stimuli needed to reach both error criteria simultaneously.

(dashed blue curve), which reached 80% power with 12 participants. At this sample size, the joint distributions of the number of stimuli needed to reach acceptable errors were consistent with our simulations, with EIG achieving the minimum (Fig. 5e).

3 Discussion

We introduce a Bayesian adaptive design that uses posterior uncertainty over recruitment curve parameters to maximize expected information gain (EIG). It improves threshold estimation over uniform sampling while matching ML-PEST, and improves predictive accuracy over ML-PEST while matching uniform. These advantages extend to simultaneous acquisition over multiple muscles. We show EIG performs comparably to EIG focused on threshold; hence, it can be adopted without compromising on threshold estimation while retaining predictive accuracy. For a use case, we show data collected using EIG yields greater statistical power when detecting a change in

threshold in intervention studies. We further optimize our design to operate within the 5-second interstimulus interval for TMS and demonstrate its feasibility in acquiring real data. Finally, our Python library `onMEP` for online acquisition of neural recruitment curves makes our approach broadly accessible to researchers for diverse applications.

EIG improves recruitment curve acquisition without trading off threshold estimation. Our primary motivation in developing the design was to improve upon uniform sampling, rather than methods that target a single curve parameter. In general, recruitment curves provide a more robust characterization of corticospinal output than a single parameter or MEPs recorded at a single intensity [15, 29]. They use data from multiple trials across a stimulation range and yield, in addition to threshold, other curve parameters including slope and saturation, which are subsequently analyzed offline. As such, uniform sampling is a natural choice for characterizing the overall curve. However, it is inefficient because it allocates samples well below threshold, where they provide little information about the curve. EIG instead reallocates these samples to more informative regions along the rising portion of the curve (Fig. 1, 5a). However, this improvement in acquisition efficiency should not come at the expense of threshold estimation, since threshold remains a critical parameter for assessing changes in corticospinal excitability [15–19]. We therefore compared EIG with ML-PEST, EIG focused on threshold (Fig. 2, 5b–e), and variance reduction on threshold (Supplementary Fig. S2). EIG performed comparably to these threshold focused designs while matching uniform sampling in predictive accuracy, suggesting it does not trade off threshold estimation for recovering the overall curve. These improvements hold for both single and multiple-muscle acquisition. Thus, EIG combines accurate threshold and curve estimation, such that choosing a different design would require more stimuli to reach either criterion.

Our adaptive design can only be as accurate as the underlying model describing neural recruitment, which we previously validated on empirical TMS and SCS data [44]. Existing adaptive approaches have used Fisher information with a four-parameter logistic function [30]. These approaches rely on numerical optimization over non-convex search spaces to estimate curve parameters, which is susceptible to suboptimal solutions with small samples and requires repeated random reinitializations to avoid local minima, and provide only point estimates. In contrast, our approach uses the posterior distribution over curve parameters to quantify estimation uncertainty, and EIG chooses the next stimulus by averaging information gain over this uncertainty rather than relying on a point estimate. Furthermore, using a recruitment curve function that does not appropriately represent the data can produce systematic errors in parameter estimates and compromise subsequent analyses. The threshold is neither an explicit parameter nor can it be derived from the logistic functions since they are smooth and cannot capture sharp deflections from the baseline response. Instead, our rectified-logistic function provides a consistent parametric representation of threshold as a deflection of MEP size from the baseline response. It outperformed traditional four- and five-parameter logistic functions in predictive performance, as determined through cross-validation on empirical TMS and SCS data (see Fig. 3 of [44]). Lastly, we do not establish whether EIG is preferable to Fisher information as an acquisition

objective; rather, EIG provides a natural information-theoretic criterion within our Bayesian framework.

While we demonstrate feasibility in human TMS experiments, a study comparing EIG with uniform sampling and ML-PEST remains necessary. However, the present experiments establish the necessary experimental setup and provide a prospective sample size estimate for such a study. A prospective comparison also introduces experimental variability absent from our simulations. In particular, the threshold can drift during an experimental session, potentially confounding comparisons between acquisition designs [52, 59]. Such drift may arise from changes in participant state over long sessions, from inadvertent effects of neurostimulation, or from anticipation of the next stimulation event. To mitigate the latter, we plan to use a minimum interstimulus interval of 5 seconds followed by an exponentially distributed delay, whose memoryless property prevents anticipation of the next stimulus based on elapsed waiting time. Additional work is needed to characterize the within session changes in threshold and distinguish them from differences attributable to the acquisition design.

Although we do not prescribe a stopping criterion, the posterior distribution available after each update provides a natural basis for adaptive stopping. For instance, acquisition could terminate once the 95% HDI of the threshold reaches an acceptable width or its width stabilizes across consecutive stimuli. The appropriate criterion will depend on the intended application and the precision required for the parameter of interest. We therefore leave the systematic evaluation of stopping criteria to future work rather than prescribing one criterion for all experiments.

Finally, although we validate our method experimentally for TMS, its deployment in SCS requires additional development. However, the principal limitations are computational rather than methodological. SCS spans a larger stimulation range compared to 101 integral values over $[0, 100]\%$ MSO for TMS, thereby increasing the candidate stimuli over which the acquisition function must be evaluated and consequently the proposal time. Moreover, SCS may also use interstimulus intervals shorter than the approximately 5 seconds currently required by our implementation. Faster posterior updates using importance sampling or sequential Monte Carlo could reduce this computational time [60, 61]. Nonetheless, the reduction in the total number of stimuli offsets a modest increase in the interstimulus interval, shortening the overall session in SCS.

More broadly, our method is inherently myopic since it optimizes a utility function for one stimulus at a time. Whether a multistep look-ahead approach would improve acquisition remains an open question. Policy-based approaches instead optimize decisions over a sequence of experiments. For instance, deep adaptive design learns an amortized policy by directly optimizing a sequential information objective [62]. Such approaches could improve acquisition beyond myopic EIG while shifting much of the computational cost from online acquisition to offline training, making them feasible for SCS. Further work is needed to determine whether these computational advantages can be achieved without compromising estimation accuracy. Nonetheless, our generative hierarchical framework provides a basis for training and testing such methods [44].

In conclusion, the proposed design reduces the experimental burden of acquiring neural recruitment curves. By allocating samples to informative regions of the curve, it reduces the number of stimuli needed to accurately characterize both the threshold and the overall curve, thereby shortening experimental sessions and reducing inadvertent neuromodulation while enabling neuromuscular parameters to be characterized across multiple muscles simultaneously. Future work will explore extensions to spatial problems, such as optimizing coil placement in TMS, as well as amortized approaches to facilitate deployment in SCS.

4 Methods

4.1 Bayesian adaptive design

4.1.1 Preliminaries

Let $\mathbb{D} = \langle (x_i, y_i) \rangle_{i=1}^n$ denote the sequence of n observations collected so far during an experiment, where $x_i \in [0, \infty)$ is the stimulation intensity and $y_i \in (0, \infty)$ is the observed motor-evoked potential (MEP) size at that intensity. Once we solve the sequential design problem going from n to $n+1$ observations, the same is repeated for each subsequent observation; as such, we omit the dependence on n from the notation. Let $p(\theta | \mathbb{D})$ denote the joint posterior over model parameters θ , which is obtained using Markov chain Monte Carlo (MCMC) methods, such as the No-U-Turn sampler (NUTS) [51]. In the context of sequential design, the current posterior serves as the prior for choosing the next stimulation intensity x . Hence, to avoid confusion, we refer to $p(\theta | \mathbb{D})$ as the prior. Once the next observation (x, y) is available, the updated prior becomes the posterior $p(\theta | \mathbb{D}, x, y)$.

The model is specified under a likelihood $p(y | \theta, x)$. The prior $p(\theta | \mathbb{D})$ is our knowledge of the model parameters based on the data collected so far. The evidence $p(y | \mathbb{D}, x)$ is the predictive distribution of the MEP size y at intensity x , obtained by averaging the likelihood over the prior. Equation (1) summarizes how these relate through Bayes' rule where we assume the model is completely specified by the model parameters, so that $p(y | \mathbb{D}, \theta, x) = p(y | \theta, x)$, and the prior is independent of the design (stimulation intensity) choice, so that $p(\theta | \mathbb{D}, x) = p(\theta | \mathbb{D})$.

$$\underbrace{p(\theta | \mathbb{D}, x, y)}_{\text{posterior}} = \frac{\underbrace{p(y | \theta, x)}_{\text{likelihood}} \underbrace{p(\theta | \mathbb{D})}_{\text{prior}}}{\underbrace{p(y | \mathbb{D}, x)}_{\text{evidence}}} \quad (1)$$

A Monte Carlo (MC) estimator of the evidence is given as,

$$\begin{aligned} p(y | \mathbb{D}, x) &= \int_{\theta} p(y | \theta, x) p(\theta | \mathbb{D}) d\theta \\ &= \mathbb{E}_{p(\theta | \mathbb{D})} p(y | \theta, x) \end{aligned}$$

$$\approx \frac{1}{S_1} \sum_{s=1}^{S_1} p(y | \theta^{(s)}, x) \quad \theta^{(s)} \sim p(\theta | \mathbb{D}) \quad (2)$$

Thus, the sequential design problem is to determine how to optimally choose the next stimulation intensity x based on the data collected so far. The appropriate design criterion depends on the inferential objective—which aspects of the recruitment curve we wish to learn. In the subsequent sections, we present several such objectives.

4.1.2 Expected information gain (EIG)

The information gain (IG) from a new observation (x, y) is defined as the Kullback–Leibler (KL) divergence [63] between the posterior $p(\theta | \mathbb{D}, x, y)$ and the prior $p(\theta | \mathbb{D})$,

$$\begin{aligned} \text{IG}(x, y) &= \text{KL}[p(\theta | \mathbb{D}, x, y) \| p(\theta | \mathbb{D})] \\ &= \int_{\theta} \log \left(\frac{p(\theta | \mathbb{D}, x, y)}{p(\theta | \mathbb{D})} \right) p(\theta | \mathbb{D}, x, y) d\theta \end{aligned}$$

Thus, the information gain measures how much the posterior differs from the prior after receiving a new observation (x, y) . However, when choosing the next intensity x , the corresponding MEP size y has not yet been observed. Instead, the evidence $p(y | \mathbb{D}, x)$ provides a predictive distribution over the possible MEP sizes y at the candidate intensity x . Therefore, we average the information gain over the evidence to obtain the expected information gain (EIG) [40, 42, 43],

$$\begin{aligned} \text{EIG}(x) &= \mathbb{E}_{p(y | \mathbb{D}, x)} \text{IG}(x, y) \\ &= \int_y \int_{\theta} \log \left(\frac{p(\theta | \mathbb{D}, x, y)}{p(\theta | \mathbb{D})} \right) p(\theta | \mathbb{D}, x, y) p(y | \mathbb{D}, x) d\theta dy \\ &= \int_y \int_{\theta} \log \left(\frac{p(y | \theta, x)}{p(y | \mathbb{D}, x)} \right) p(y, \theta | \mathbb{D}, x) d\theta dy \\ &= \mathbb{E}_{p(y, \theta | \mathbb{D}, x)} \log \left(\frac{p(y | \theta, x)}{p(y | \mathbb{D}, x)} \right) \\ &\approx \frac{1}{S_2} \sum_{s=1}^{S_2} \log \left(\frac{p(y^{(s)} | \theta^{(s)}, x)}{p(y^{(s)} | \mathbb{D}, x)} \right) \quad (3) \\ &\quad \theta^{(s)} \sim p(\theta | \mathbb{D}), \quad y^{(s)} \sim p(y | \theta^{(s)}, x) \end{aligned}$$

Equation (3) gives the nested MC (NMC) estimator of the EIG [42, 64]. It is nested because the denominator $p(y^{(s)} | \mathbb{D}, x)$ must itself be approximated for each $y^{(s)}$ using (2), resulting in $\mathcal{O}(S_1 S_2)$ likelihood evaluations for each candidate intensity x .

4.1.3 Focused EIG

We partition the model parameters as $\theta = \phi \sqcup \eta$, i.e., $\theta = \phi \cup \eta$ and $\phi \cap \eta = \emptyset$. Here, we consider the information gain only about the parameters of interest ϕ ,

$$\begin{aligned} \text{IG}_\phi(x, y) &= \text{KL}[p(\phi | \mathbb{D}, x, y) \| p(\phi | \mathbb{D})] \\ &= \int_\phi \log\left(\frac{p(\phi | \mathbb{D}, x, y)}{p(\phi | \mathbb{D})}\right) p(\phi | \mathbb{D}, x, y) d\phi \end{aligned}$$

As before, we average the information gain over the evidence to obtain the focused expected information gain (EIG_ϕ) [65],

$$\begin{aligned} \text{EIG}_\phi(x) &= \mathbb{E}_{p(y|\mathbb{D}, x)} \text{IG}_\phi(x, y) \\ &= \int_y \int_\phi \log\left(\frac{p(\phi | \mathbb{D}, x, y)}{p(\phi | \mathbb{D})}\right) p(\phi | \mathbb{D}, x, y) p(y | \mathbb{D}, x) d\phi dy \\ &= \int_y \int_\phi \log\left(\frac{p(y | \phi, \mathbb{D}, x)}{p(y | \mathbb{D}, x)}\right) p(y, \phi | \mathbb{D}, x) d\phi dy \\ &= \int_y \int_\phi \int_\eta \log\left(\frac{p(y | \phi, \mathbb{D}, x)}{p(y | \mathbb{D}, x)}\right) p(y, \phi, \eta | \mathbb{D}, x) d\eta d\phi dy \\ &= \mathbb{E}_{p(y, \theta | \mathbb{D}, x)} \log\left(\frac{p(y | \phi, \mathbb{D}, x)}{p(y | \mathbb{D}, x)}\right) \\ &\approx \frac{1}{S_2} \sum_{s=1}^{S_2} \log\left(\frac{p(y^{(s)} | \phi^{(s)}, \mathbb{D}, x)}{p(y^{(s)} | \mathbb{D}, x)}\right) \\ &\quad \theta^{(s)} = (\phi^{(s)}, \eta^{(s)}) \sim p(\theta | \mathbb{D}), \quad y^{(s)} \sim p(y | \theta^{(s)}, x) \end{aligned} \tag{4}$$

The conditional evidence $p(y^{(s)} | \phi^{(s)}, \mathbb{D}, x)$ is given as,

$$\begin{aligned} p(y^{(s)} | \phi^{(s)}, \mathbb{D}, x) &= \int_\eta p(y^{(s)} | \phi^{(s)}, \eta, x) p(\eta | \phi^{(s)}, \mathbb{D}) d\eta \\ &= \mathbb{E}_{p(\eta | \phi^{(s)}, \mathbb{D})} p(y^{(s)} | \phi^{(s)}, \eta, x) \\ &\approx \sum_{r=1}^{S_3} \tilde{k}_r(\phi^{(s)}) p(y^{(s)} | \phi^{(s)}, \eta^{(r)}, x) \end{aligned} \tag{5}$$

$$\theta^{(r)} = (\phi^{(r)}, \eta^{(r)}) \sim p(\theta | \mathbb{D}), \quad k_r(\phi^{(s)}) = K(\phi^{(r)} - \phi^{(s)}), \quad \tilde{k}_r(\phi^{(s)}) = \frac{k_r(\phi^{(s)})}{\sum_{\ell=1}^{S_3} k_\ell(\phi^{(s)})}$$

where $K : \mathbb{R}^{|\phi|} \rightarrow [0, \infty)$ is a kernel in ϕ , and (5) is the Nadaraya–Watson kernel regression estimator of the conditional expectation [66, 67]. Intuitively, it averages over nuisance parameters $\eta^{(r)}$ whose corresponding parameters of interest $\phi^{(r)}$ are close to $\phi^{(s)}$, as determined by the kernel weights $\tilde{k}_r(\phi^{(s)})$. These weighted samples approximate the conditional posterior $p(\eta | \phi^{(s)}, \mathbb{D})$. Thus, the numerator in (4) requires $\mathcal{O}(S_2 S_3)$ likelihood evaluations, resulting in $\mathcal{O}(S_2(S_1 + S_3))$ likelihood evaluations overall.

4.1.4 Expected variance reduction (EVR)

A natural objective is to reduce the variance of a function of the model parameters, $f : \mathbb{R}^{|\theta|} \rightarrow \mathbb{R}$, $\theta \mapsto f(\theta)$. For instance, one may define $f(\theta) := a$, where a denotes the motor threshold, or $f(\theta) := \mathcal{F}(x = 1.2 \cdot a | \Omega)$, $\Omega \subset \theta$, the expected MEP size at 120% of the motor threshold under the model (11)–(12).

Consider the reduction in the variance of $f(\theta)$ from the prior to the posterior after receiving a new observation (x, y) ,

$$\Delta(x, y) = \text{Var}_{p(\theta|\mathbb{D})} f(\theta) - \text{Var}_{p(\theta|\mathbb{D}, x, y)} f(\theta)$$

As before, we average this under the evidence to obtain the expected variance reduction (EVR) [68].

$$\begin{aligned} \text{EVR}(x) &= \mathbb{E}_{p(y|\mathbb{D}, x)} \Delta(x, y) \\ &= \text{Var}_{p(\theta|\mathbb{D})} f(\theta) - \mathbb{E}_{p(y|\mathbb{D}, x)} \text{Var}_{p(\theta|\mathbb{D}, x, y)} f(\theta) \end{aligned}$$

The variance under the prior is given as,

$$\begin{aligned} \text{Var}_{p(\theta|\mathbb{D})} f(\theta) &= \mathbb{E}_{p(\theta|\mathbb{D})} f(\theta)^2 - (\mathbb{E}_{p(\theta|\mathbb{D})} f(\theta))^2 \\ &\approx \frac{1}{S_1} \sum_{s=1}^{S_1} f(\theta^{(s)})^2 - \left(\frac{1}{S_1} \sum_{s=1}^{S_1} f(\theta^{(s)}) \right)^2 \quad \theta^{(s)} \sim p(\theta | \mathbb{D}) \end{aligned} \quad (6)$$

The variance under the posterior is given as,

$$\begin{aligned} \text{Var}_{p(\theta|\mathbb{D}, x, y)} f(\theta) &= \mathbb{E}_{p(\theta|\mathbb{D}, x, y)} f(\theta)^2 - (\mathbb{E}_{p(\theta|\mathbb{D}, x, y)} f(\theta))^2 \\ &\approx \sum_{s=1}^{S_1} \tilde{\omega}_s(y) f(\theta^{(s)})^2 - \left(\sum_{s=1}^{S_1} \tilde{\omega}_s(y) f(\theta^{(s)}) \right)^2 \\ \theta^{(s)} &\sim p(\theta | \mathbb{D}), \quad \omega_s(y) = p(y | \theta^{(s)}, x), \quad \tilde{\omega}_s(y) = \frac{\omega_s(y)}{\sum_{\ell=1}^{S_1} \omega_\ell(y)} \end{aligned}$$

where $\tilde{\omega}_s(y)$ are the self-normalized importance sampling weights obtained when using the prior $p(\theta | \mathbb{D})$ as the proposal distribution for the posterior $p(\theta | \mathbb{D}, x, y)$ [68]. This

follows from (1), since $p(\theta | \mathbb{D}, x, y) \propto p(y | \theta, x) p(\theta | \mathbb{D})$, so dividing by the proposal distribution $p(\theta | \mathbb{D})$ gives importance weights proportional to the likelihood $p(y | \theta, x)$. Finally, the expectation under the evidence is given as,

$$\begin{aligned}
\mathbb{E}_{p(y|\mathbb{D},x)} \text{Var}_{p(\theta|\mathbb{D},x,y)} f(\theta) &= \int_y \text{Var}_{p(\theta|\mathbb{D},x,y)} [f(\theta)] p(y | \mathbb{D}, x) dy \\
&= \int_y \text{Var}_{p(\theta|\mathbb{D},x,y)} [f(\theta)] \int_{\theta' \in \mathbb{R}^{|\theta|}} p(y, \theta' | \mathbb{D}, x) d\theta' dy \\
&= \int_y \int_{\theta' \in \mathbb{R}^{|\theta|}} \text{Var}_{p(\theta|\mathbb{D},x,y)} [f(\theta)] p(y, \theta' | \mathbb{D}, x) d\theta' dy \\
&= \mathbb{E}_{p(y, \theta' | \mathbb{D}, x)} \text{Var}_{p(\theta|\mathbb{D},x,y)} f(\theta) \\
&\approx \frac{1}{S_2} \sum_{r=1}^{S_2} \left[\sum_{s=1}^{S_1} \tilde{\omega}_s(y^{(r)}) f(\theta^{(s)})^2 - \left(\sum_{s=1}^{S_1} \tilde{\omega}_s(y^{(r)}) f(\theta^{(s)}) \right)^2 \right] \\
&\quad \theta^{(r)} \sim p(\theta | \mathbb{D}), \quad y^{(r)} \sim p(y | \theta^{(r)}, x) \tag{7}
\end{aligned}$$

thereby requiring $\mathcal{O}(S_1 S_2)$ likelihood evaluations. Supplementary Fig. S2 shows EVR on threshold ($f(\theta) := a$) matches EIG focused on threshold ($\phi = \{a\}$) in accuracy while having faster proposal time.

4.2 Extension to multiple muscles

More generally, the observation (x, y) consists of MEP sizes $y = (y_m)_{m=1}^M \in (0, \infty)^M$ recorded simultaneously from M muscles at stimulation intensity $x \in [0, \infty)$. Let $\mathbb{D}_m = \langle (x_i, y_{i,m}) \rangle_{i=1}^n$ denote the data for muscle m , so the collected data partitions as $\mathbb{D} = \bigsqcup_m \mathbb{D}_m$. Similarly, let θ_m denote the model parameters for muscle m , so the model parameters partition as $\theta = \bigsqcup_m \theta_m$. Under the following assumptions, we will show the acquisition functions for EIG, focused EIG, and EVR decompose into sums of their corresponding muscle-specific acquisition functions, and hence can be computed in parallel. We assume the priors are independent across muscles $p(\theta | \mathbb{D}) = \prod_m p(\theta_m | \mathbb{D}_m)$, and the observations are conditionally independent across muscles given the model parameters, such that $p(y | \theta, x) = \prod_m p(y_m | \theta_m, x)$. It follows that the evidence factorizes as,

$$\begin{aligned}
p(y | \mathbb{D}, x) &= \int_{\theta} p(y | \theta, x) p(\theta | \mathbb{D}) d\theta \\
&= \int_{\theta} \prod_m [p(y_m | \theta_m, x) p(\theta_m | \mathbb{D}_m)] d\theta
\end{aligned}$$

$$\begin{aligned}
&= \prod_m \int_{\theta_m} p(y_m | \theta_m, x) p(\theta_m | \mathbb{D}_m) d\theta_m \\
&= \prod_m p(y_m | \mathbb{D}_m, x)
\end{aligned} \tag{8}$$

Using (8) together with (1), the posterior also factorizes $p(\theta | \mathbb{D}, x, y) = \prod_m p(\theta_m | \mathbb{D}_m, x, y_m)$. Now consider,

$$\text{EIG}(x) = \mathbb{E}_{p(y, \theta | \mathbb{D}, x)} \log \left(\frac{p(y | \theta, x)}{p(y | \mathbb{D}, x)} \right) \tag{9}$$

$$\begin{aligned}
&= \mathbb{E}_{p(y, \theta | \mathbb{D}, x)} \log \left(\prod_m \frac{p(y_m | \theta_m, x)}{p(y_m | \mathbb{D}_m, x)} \right) \\
&= \sum_m \mathbb{E}_{p(y, \theta | \mathbb{D}, x)} \log \left(\frac{p(y_m | \theta_m, x)}{p(y_m | \mathbb{D}_m, x)} \right) \\
&= \sum_m \underbrace{\mathbb{E}_{p(y_m, \theta_m | \mathbb{D}_m, x)} \log \left(\frac{p(y_m | \theta_m, x)}{p(y_m | \mathbb{D}_m, x)} \right)}_{\text{EIG}_m(x)} \\
&= \sum_m \text{EIG}_m(x)
\end{aligned} \tag{10}$$

where $\text{EIG}_m(x)$ is the EIG for muscle m and can be approximated using the MC estimator in (3).

Now assume the muscle-specific parameters partition as $\theta_m = \phi_m \sqcup \eta_m$, and define $\phi = \sqcup_m \phi_m$ the set of parameters of interest, and $\eta = \sqcup_m \eta_m$ the remaining nuisance parameters. It is easy to see that the conditional posterior factorizes as $p(\eta | \phi, \mathbb{D}) = \prod_m p(\eta_m | \phi_m, \mathbb{D}_m)$. Consequently, the conditional evidence also factorizes as $p(y | \phi, \mathbb{D}, x) = \prod_m p(y_m | \phi_m, \mathbb{D}_m, x)$. An argument similar to (9)–(10) shows the focused EIG also decomposes across muscles and can be approximated by the sum of muscle-specific focused EIG estimates obtained using (4).

Finally, let $f(\theta) = (f_1(\theta_1), \dots, f_M(\theta_M))^\top$ denote a function of the model parameters, where $f_m : \mathbb{R}^{|\theta_m|} \rightarrow \mathbb{R}$. For instance, one may define $f_m(\theta_m) := a_m$, where a_m denotes the muscle-specific threshold. We define the reduction in variance as

$$\begin{aligned}
\Delta(x, y) &= \text{tr}(\text{Cov}_{p(\theta | \mathbb{D})} f(\theta)) - \text{tr}(\text{Cov}_{p(\theta | \mathbb{D}, x, y)} f(\theta)) \\
&= \sum_m \text{Var}_{p(\theta | \mathbb{D})} f_m(\theta_m) - \sum_m \text{Var}_{p(\theta | \mathbb{D}, x, y)} f_m(\theta_m) \\
&= \sum_m \underbrace{[\text{Var}_{p(\theta_m | \mathbb{D}_m)} f_m(\theta_m) - \text{Var}_{p(\theta_m | \mathbb{D}_m, x, y_m)} f_m(\theta_m)]}_{\Delta_m(x, y_m)}
\end{aligned}$$

$$= \sum_m \Delta_m(x, y_m)$$

Averaging under the evidence gives,

$$\begin{aligned} \text{EVR}(x) &= \mathbb{E}_{p(y|\mathbb{D}, x)} \Delta(x, y) \\ &= \sum_m \mathbb{E}_{p(y|\mathbb{D}, x)} \Delta_m(x, y_m) \\ &= \sum_m \underbrace{\mathbb{E}_{p(y_m|\mathbb{D}_m, x)} \Delta_m(x, y_m)}_{\text{EVR}_m(x)} \\ &= \sum_m \text{EVR}_m(x) \end{aligned}$$

where $\text{EVR}_m(x)$ is the EVR for muscle m and can be approximated using the MC estimators (6), (7). Supplementary Fig. S2 shows EVR on threshold applied to the acquisition of multiple muscles.

4.3 Model evaluation

We model the MEP size y as a rectified-logistic function $\mathcal{F} : \mathbb{R} \rightarrow \mathbb{R}^+$ of stimulation intensity x under a gamma (shape–rate) likelihood (11)–(12) [44]. Supplementary Fig. S3a shows the effect of varying the function parameters. Here, parameter a is the threshold, b controls the growth rate, g is the baseline response of the recording device, h controls the distance between the baseline and saturation ($g+h$), and v controls the location of the inflection point, or point of maximum gradient, whether near the baseline g or saturation ($g+h$). The parameters c_1 and c_2 together control the noise around the curve. Supplementary Fig. S3b,c specifies the model used in the update step to estimate the curve parameters from the data acquired in Results 2.1–2.4 (Fig. 2–4). For the real TMS experiments in Results 2.5 (Fig. 5a), the model was replaced with a mixture extension to account for outliers (Supplementary Fig. S4). In this case, we turn off the mixture component ($z = 0$) when evaluating the EIG acquisition function.

$$y | x, \theta \sim \text{gamma}(\mu\beta, \beta) \quad (11)$$

$$\mu = \mathcal{F}(x | \Omega)$$

$$\beta = \frac{1}{c_1} + \frac{1}{c_2\mu}$$

$$\mathcal{F}(x | \Omega) = g + \max \left\{ 0, -v + \frac{h+v}{1 + \left(\frac{h}{v}\right) e^{-b(x-a)}} \right\} \quad (12)$$

$$\Omega = \{a, b, g, h, v\} \quad \theta = \Omega \cup \{c_1, c_2\} \subset (0, \infty)$$

Given any acquired dataset $\mathbb{D}$, we use the model to estimate the parameters with MCMC and obtain current posterior samples $\{\tilde{\theta}^{(s)}\}_{s=1}^S \sim p(\theta | \mathbb{D})$. In simulations, the true model parameters θ are known. We evaluate the L_1 loss for threshold estimation (e_a , absolute error) and curve estimation (e_y , integrated absolute error) as

$$\begin{aligned} e_a &= \mathbb{E}_{\tilde{\theta} \sim p(\theta | \mathbb{D})} |a - \tilde{a}| \\ &\approx \frac{1}{S} \sum_{s=1}^S |a - \tilde{a}^{(s)}| \\ e_y &= \frac{1}{h} \mathbb{E}_{\tilde{\theta} \sim p(\theta | \mathbb{D})} \int_{x=0}^{100} |\mathcal{F}(x | \Omega_0) - \mathcal{F}(x | \tilde{\Omega}_0)| dx \\ &\approx \frac{1}{Sh} \sum_{s=1}^S \int_{x=0}^{100} |\mathcal{F}(x | \Omega_0) - \mathcal{F}(x | \tilde{\Omega}_0^{(s)})| dx \end{aligned}$$

where $\Omega_0 = (\theta \setminus \{c_1, c_2\})|_{g=0}$ and $\tilde{\Omega}_0 = (\tilde{\theta} \setminus \{\tilde{c}_1, \tilde{c}_2\})|_{\tilde{g}=0}$, i.e., we remove the baseline response before comparing the curves. Since the integral does not have a closed form, we evaluate it numerically using the trapezoidal rule on a grid of 501 points over $[0, 100]\%$ MSO. We repeat this evaluation across 2,000 simulations (participants), and Fig. 2a,b shows the mean error and standard error of the mean across those repetitions, with e_a labeled as the threshold error and e_y the predictive error.

Both errors are expressed in % MSO, or the unit of stimulation intensity. To define an acceptable error, we further define the corresponding relative errors as

$$\begin{aligned} r_a &= \frac{e_a}{a} \times 100 \\ r_y &= \frac{e_y}{(100 - a)} \times 100 \end{aligned}$$

i.e., we normalize the predictive error by the activation interval $(100 - a)$ within the range of integration, where we assume the true threshold $a \in (0, 100)$. For each participant, we follow an acquisition method and calculate its relative threshold error r_a and relative predictive error r_y for number of stimuli from $n = 3$ to $n = 64$. To compare acquisition designs by the number of stimuli needed to reach an acceptable error, we estimate where the relative error trajectory crosses that level. In particular, we model the error trajectory (say r_q , $q \in \{a, y\}$) as a piecewise linear function of the number of stimuli in log-log space, which corresponds to a piecewise power decay model in the original space. For an acceptable relative error $k_q \in (0, \infty)$, the likelihood is

$$\log_2 r_q | n, \theta_q \sim \text{normal}(\mu_q(n), \omega) \quad \theta_q = \{t, \lambda_{\text{late}}, \rho, \tau, \omega\}$$

$$\begin{aligned}\mu_q(n) &= \log_2 k_q - \lambda_{\text{early}} (\log_2 n - t) \\ &\quad - (\lambda_{\text{late}} - \lambda_{\text{early}}) [\max(0, \log_2 n - \tau) - \max(0, t - \tau)] \\ \lambda_{\text{early}} &= \frac{\lambda_{\text{late}}}{\rho} \quad \lambda_{\text{late}} > 0 \quad 0 < \rho < 1\end{aligned}$$

The parameters λ_{early} and λ_{late} are the rates of error reduction before and after the change point $n = 2^\tau$, respectively. The constraint $\lambda_{\text{early}} > \lambda_{\text{late}}$ allows the rate of error reduction to decrease as more stimuli are acquired. By construction, $\mu_q(2^t) = \log_2 k_q$, so the model reaches a median error of $r_q = k_q$ at $n = 2^t \in (0, \infty)$ stimuli. Thus, t determines the number of stimuli (in log space) needed to reach the acceptable error and is the parameter of interest for comparing the acquisition designs.

We used the likelihood above within a hierarchical Bayesian estimation (HBe) model [44] to compare the number of stimuli needed by each design to reach acceptable relative errors of 2% ($k_a = 2$) and 10% ($k_y = 10$) in threshold and curve estimation, respectively. Supplementary Fig. S5 specifies the HBe model and Supplementary Fig. S6 shows a few resulting estimates. Specifically, we sampled error trajectories of 16 participants without replacement from the 2,000 simulations and modeled the corresponding values of t , with μ_0 describing the population-level location for EIG, and μ_Δ the population-level location of paired differences between EIG and each comparison design. This was repeated 500 times and Fig. 2c shows the median posterior estimates of μ_0 for EIG and $(\mu_0 + \mu_\Delta)$ for each comparison design across those repetitions.

The simulations and evaluation of true and false positive rates for Fig. 4 are covered in [44] (see Methods 2.3). Supplementary Fig. S7 specifies the HBe model used to summarize the location of threshold differences across participants. The null hypothesis of no difference in threshold between the pre- and post-intervention was rejected if the 95% highest density interval (HDI) of the location parameter μ_Δ excluded zero.

For Fig. 5b–e, we repeated the analysis in Fig. 2 using the eight participants from the real TMS experiments. Because the ground truth parameters are not available, we instead used the posterior estimated from the 40 samples acquired for each participant by the end of their experimental session. In particular, we sampled 200 parameter sets from each participant’s posterior draws as possible ground truths and evaluated each acquisition design on them. For Fig. 5b,c, we first took the median error over the 200 sets for each participant and then reported the mean and standard error across the eight participants.

For the prospective power analysis in Fig. 5d, we used the same HBe model (Supplementary Fig. S5) but varied the participant sample sizes $n \in \{2, 4, 8, 12, 16\}$. For each sample size, we sampled n participants with replacement from the eight participants and paired them with n posterior-draw indices sampled without replacement from $\{1, \dots, 200\}$ to choose the corresponding parameter set from each participant’s posterior as ground truth. The null hypothesis of no difference between EIG and a comparison design in the number of stimuli required to reach an acceptable error was rejected if the 95% HDI of the location parameter μ_Δ excluded zero. The outcome was a success if the null hypotheses comparing EIG against uniform sampling for threshold estimation and against ML-PEST for curve estimation were rejected simultaneously.

This was repeated 500 times and Fig. 5d shows the rejection rate across those repetitions, which reached 80% power with 12 participants. Figure 5e additionally shows the median posterior estimates of μ_0 for EIG and $(\mu_0 + \mu_\Delta)$ for each comparison design across these repetitions with 12 participants.

4.4 Implementation

All acquisition methods were initialized with eight samples at zero intensity, followed by three stimuli from uniform sampling. We do not count the eight samples at zero intensity in the number of stimuli because they do not correspond to actual stimulation events, require no additional effort from the participant or experimenter, and are readily available from the recording device [30]. EIG does not require initial samples; however, all acquisition methods used the same initial dataset to ensure they begin with the same error. The update step included four chains, each using 400 warmup samples and 400 post-warmup samples with the NUTS sampler, for a total of $400 \times 4 = 1600$ posterior samples. These samples were also used to evaluate estimation error after each update step (Methods 4.3). Throughout acquisition, we used all posterior samples, with $S_1 = S_2 = S_3 = 1600$ in (2)–(7). The acquisition step was batched across candidate stimuli to avoid GPU memory oversubscription. The implementation of EIG follows directly from the estimators described above, whereas focused EIG and the comparison designs require a few additional details we note below. Detailed implementation of each design is available in our Python library `onMEP`.

4.4.1 Focused EIG

Focused EIG used a full covariance Gaussian kernel for the kernel regression estimator in (5), with bandwidth determined by the multivariate Silverman rule [66, 67, 69]. We describe below how we construct the kernel matrix and obtain the normalized weights in (5). Let $F = |\phi|$ denote the cardinality of the focused set. We used $S_2 = S_3 (= S)$ in (4) and (5). Let $U \in \mathbb{R}^{F \times S}$ denote the focused posterior samples mapped to unconstrained space. We define $\mathbf{1} \in \mathbb{R}^{S \times 1}$ the column vector of ones. The sample mean and centered samples are

$$\bar{U} = \frac{1}{S} U \mathbf{1} \in \mathbb{R}^{F \times 1} \quad Z = U - \bar{U} \mathbf{1}^T \in \mathbb{R}^{F \times S}$$

The covariance matrix $C = \frac{1}{S} Z Z^T \in \mathbb{R}^{F \times F}$ determines the shape and orientation of the kernel. The Silverman rule gives the bandwidth matrix

$$h = \left(\frac{S(F+2)}{4} \right)^{-2/(F+4)} \quad H = hC \in \mathbb{R}^{F \times F}$$

Let $U^{(i)} \in \mathbb{R}^{F \times 1}$ denote the i th sample (column) of U . The matrix of squared Mahalanobis distances $D = (D_{ij}) \in \mathbb{R}^{S \times S}$ is

$$D_{ij} = \left(U^{(i)} - U^{(j)} \right)^T H^{-1} \left(U^{(i)} - U^{(j)} \right)$$

The Gaussian kernel matrix $K = (K_{ij}) \in \mathbb{R}^{S \times S}$, and the normalized weights $\tilde{K} = (\tilde{K}_{ij}) \in \mathbb{R}^{S \times S}$ are then

$$K_{ij} = \exp\left(-\frac{1}{2}D_{ij}\right) \in (0, 1] \quad \tilde{K}_{ij} = \frac{K_{ij}}{\sum_{\ell=1}^S K_{i\ell}} \in (0, 1]$$

Thus, $\tilde{K}_{sr}$ gives the weight $\tilde{k}_r(\phi^{(s)})$ in (5). The weights are precomputed once per posterior update and reused across candidate stimulation intensities.

4.4.2 Uniform sampling and ML-PEST

Uniform sampling was implemented using nested quasi-random sequences in SciPy [70] so that stimuli remained roughly equispaced for any number of acquired samples from $n = 3$ to $n = 64$ (Supplementary Fig. S8) [71, 72]. We used an existing implementation of ML-PEST for single-muscle acquisition (Supplementary Fig. S9) [39]. We restricted the baseline response in our simulations to remain below the 50 μV cutoff required by ML-PEST. For multiple-muscle acquisition, we extended ML-PEST using a round-robin strategy across muscles. This allowed ML-PEST to use all samples recorded simultaneously from multiple muscles at a given intensity, as opposed to applying it to each muscle iteratively, which would discard samples from the remaining muscles and artificially inflate the number of stimuli needed to reduce the average error across muscles. To our knowledge, such an extension of ML-PEST does not exist, and we make our implementation available in the Python library onMEP.

4.4.3 Notation

Where they appear, $\text{normal}(\mu, \sigma)$ is the normal distribution (location-scale), $\text{gamma}(\alpha, \beta)$ the gamma distribution (shape-rate), $\text{truncated normal}(\mu, \sigma)$ the truncated-normal on $(0, \infty)$ (location-scale), and $\text{normal}^+(\sigma)$ the half-normal (scale). $\text{Bernoulli}(p)$ denotes the Bernoulli distribution with success probability p , and $\text{Uniform}(a, b)$ the uniform distribution on $[a, b]$. For convenience, we used the terms statistical power and true positive rate interchangeably when the alternative hypothesis was known to be true.

4.5 Statistics & reproducibility

All models were fit using the No-U-Turn sampler (NUTS) implemented in NumPyro [73, 74]. The Bayesian model used in the update step and the hierarchical Bayesian models used for Bayesian estimation were implemented in hbMEP [44]. All acquisition functions were implemented in JAX. The timing benchmarks (Fig. 3) were performed on a machine with an Intel Core i9-13900K CPU for the update step and an NVIDIA RTX 4090 GPU for the proposal step. Code to reproduce the presented analyses is available on GitHub (see Code availability 5.6).

4.6 Experimental Validation

4.6.1 Participants

All procedures were reviewed and approved by the Institutional Review Board (IRB) of Columbia University Irving Medical Center (Protocols AAAV6853 and AAIV2760). The studies were pre-registered at clinicaltrials.gov (NCT07561372, NCT07539025). Written informed consent was obtained prior to study enrollment, and experimental procedures were conducted in compliance with all applicable institutional and governmental regulations. Individuals over the age of 18 without neurological injury were eligible for recruitment. Exclusion criteria included a history of seizures, stroke, traumatic brain injury, neurosurgery, or other neurological diseases; prior adverse reactions to transcranial magnetic stimulation (TMS) or non-invasive neurostimulation; cranial metallic implants or implanted electronic medical devices; current or suspected pregnancy; and the use of medications known to alter cortical excitability or lower seizure threshold. One participant was left-handed, while the remaining participants were right-handed, as determined by the Edinburgh Handedness questionnaire.

4.6.2 Transcranial magnetic stimulation

Transcranial magnetic stimulation (TMS) was delivered using a MagPro X100 stimulator equipped with a C-B70 butterfly coil (MagVenture). The stimulator was operated in biphasic mode (1.0 ms pulse width), and the coil was positioned over the motor cortex hotspot to elicit an optimal response in the target muscle. Coil position and orientation were continuously tracked using a stereotactic neuronavigation system (Rogue Research). Stimulator triggering was controlled via a custom MATLAB interface (The MathWorks) paired with a PCIe-6363 data acquisition board and a BNC-2090A terminal block (National Instruments). Inter-stimulus intervals (ISIs) were set to 5 seconds. Participants were seated with their palms facing down on a table and instructed not to move either hand. They were asked to maintain fixation on a cross positioned in the center of their field of view.

4.6.3 Electromyographic (EMG) recordings

To capture both on-target and off-target effects, motor responses were recorded bilaterally from multiple upper extremity muscles. A common ground electrode was positioned over the lateral epicondyle of the humerus on the targeted arm. Surface EMG signals were amplified using a Digitimer D360 amplifier and hardware band-pass filtered between 20 and 1900 Hz. The analog signals were digitized at a sampling rate of 20 kHz using a PCIe-6363 data acquisition board and a BNC-2090A terminal block (National Instruments). Raw analog signals were stored for subsequent offline analysis, while motor-evoked potentials (MEPs) were concurrently processed in real time in MATLAB R2025b (MathWorks Inc., Natick, MA, USA). Peak-to-peak MEP size between 15 ms and 75 ms was extracted and transmitted in real time from MATLAB via the Lab Streaming Layer (LSL) API to onMEP. Prior to data collection, signal quality was validated to ensure that baseline peak-to-peak noise was below 50 μ V.

4.6.4 Motor hotspot and maximum intensity determination

TMS was applied to the primary motor cortex (M1) contralateral to the dominant hand. To determine the motor hotspot, the M1 representation was systematically mapped to identify the coil position that produced the highest MEP sizes in the resting contralateral abductor pollicis brevis (APB) muscle at a constant suprathreshold intensity. Once identified, this location was saved within the neuronavigation system to guide precise coil placement throughout the experiment. Subsequently, the participant’s tolerable maximum stimulation intensity was determined by incrementally increasing the stimulator output until the individual’s comfort threshold was reached. All participants were able to tolerate the full range of stimulation [0,100]% MSO.

4.6.5 Experimental design

In a pilot with a single participant (Protocol AAV6853), we sought to demonstrate the efficacy of EIG in a head-to-head comparison against uniform sampling. A randomized block trial design was used to evaluate the performance of the algorithms implemented as part of onMEP. The algorithms under test formed a block, and their order was randomized for each trial so that both the experimenter and the participant were blinded to the upcoming condition. This dataset is used in Fig. 1.

As part of an ongoing study evaluating non-invasive paired brain and peripheral stimulation (Protocol AAV2760), eight participants additionally had their recruitment curves acquired using onMEP running EIG exclusively. This dataset is used in Fig. 5a.

5 Additional information

5.1 Competing interests

Jason B. Carmel is a founder and stockholder in BackStop Neural and a scientific advisor and stockholder in SharperSense. The other authors have nothing to disclose.

5.2 Funding

This work was supported by the National Institute of Neurological Disorders and Stroke, National Institutes of Health (awards R01NS124224, 1R03NS141040).

5.3 Acknowledgements

We thank Columbia University’s Department of Systems Biology IT (NIH: S10OD032433, S10OD021764) for providing computational resources.

5.4 CRediT authorship contribution statement

VT: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - Original Draft, Writing - Review & Editing. SSD: Investigation, Data curation, Writing - Review & Editing. DT: Methodology, Software, Validation, Writing - Review & Editing. JMM: Methodology, Validation, Writing - Review & Editing. AG:

Methodology, Validation, Writing - Review & Editing. JBC: Methodology, Resources, Writing - Review & Editing, Supervision, Funding acquisition. JRM: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

5.5 Data availability

The datasets used in this study are available from the corresponding authors upon reasonable request.

5.6 Code availability

The `onMEP` Python package for online acquisition of MEP recruitment curves implements the acquisition methods. The package is available at <https://github.com/hbmep/onmep>. Code to reproduce the presented analyses is available at <https://github.com/hbmep/onmep-manuscript>.

References

- [1] Stefan, K., Kunesch, E., Cohen, L.G., Benecke, R., Classen, J.: Induction of plasticity in the human motor cortex by paired associative stimulation. *Brain* **123**(3), 572–584 (2000) <https://doi.org/10.1093/brain/123.3.572>
- [2] Chen, R., Cros, D., Curra, A., Di Lazzaro, V., Lefaucheur, J.-P., Magistris, M.R., Mills, K., Rösler, K.M., Triggs, W.J., Ugawa, Y., Ziemann, U.: The clinical diagnostic utility of transcranial magnetic stimulation: Report of an IFCN committee. *Clinical Neurophysiology* **119**(3), 504–532 (2008) <https://doi.org/10.1016/j.clinph.2007.10.014>
- [3] Picht, T., Schmidt, S., Brandt, S., Frey, D., Hannula, H., Neuvonen, T., Karhu, J., Vajkoczy, P., Suess, O.: Preoperative functional mapping for rolandic brain tumor surgery: comparison of navigated transcranial magnetic stimulation to direct cortical stimulation. *Neurosurgery* **69**(3), 581–588 (2011) <https://doi.org/10.1227/NEU.0b013e3182181b89>
- [4] Millet, G.Y., Martin, V., Martin, A., Vergès, S.: Electrical stimulation for testing neuromuscular function: from sport to pathology. *European Journal of Applied Physiology* **111**(10), 2489–2500 (2011) <https://doi.org/10.1007/s00421-011-1996-y>
- [5] Bunday, K.L., Perez, M.A.: Motor Recovery after Spinal Cord Injury Enhanced by Strengthening Corticospinal Synaptic Transmission. *Current Biology* **22**(24), 2355–2361 (2012) <https://doi.org/10.1016/j.cub.2012.10.046>
- [6] MacDonald, D.B., Skinner, S., Shils, J., Yingling, C.: Intraoperative motor evoked potential monitoring – A position statement by the American Society of Neurophysiological Monitoring. *Clinical Neurophysiology* **124**(12), 2291–2316 (2013) <https://doi.org/10.1016/j.clinph.2013.07.025>
- [7] Sayenko, D.G., Atkinson, D.A., Dy, C.J., Gurley, K.M., Smith, V.L., Angeli, C., Harkema, S.J., Edgerton, V.R., Gerasimenko, Y.P.: Spinal segment-specific transcutaneous stimulation differentially shapes activation pattern among motor pools in humans. *Journal of Applied Physiology* **118**(11), 1364–1374 (2015) <https://doi.org/10.1152/jappphysiol.01128.2014>
- [8] Hofstoetter, U.S., Perret, I., Bayart, A., Lackner, P., Binder, H., Freundl, B., Minassian, K.: Spinal motor mapping by epidural stimulation of lumbosacral posterior roots in humans. *iScience* **24**(1), 101930 (2021) <https://doi.org/10.1016/j.isci.2020.101930>
- [9] Greiner, N., Barra, B., Schiavone, G., Lorach, H., James, N., Conti, S., Kaeser, M., Fallegger, F., Borgognon, S., Lacour, S., Bloch, J., Courtine, G., Capogrosso, M.: Recruitment of upper-limb motoneurons with epidural electrical stimulation of the cervical spinal cord. *Nature Communications* **12**(1), 435 (2021) <https://doi.org/10.1038/s41467-021-22888-8>

- [10] Pal, A., Park, H., Ramamurthy, A., Asan, A.S., Bethea, T., Johnkutty, M., Carmel, J.B.: Spinal cord associative plasticity improves forelimb sensorimotor function after cervical injury. *Brain*, 235 (2022) <https://doi.org/10.1093/brain/awac235>
- [11] Arora, T., Desai, N., Kirshblum, S., Chen, R.: Utility of transcranial magnetic stimulation in the assessment of spinal cord injury: Current status and future directions. *Frontiers in Rehabilitation Sciences* **3** (2022) <https://doi.org/10.3389/fresc.2022.1005111>
- [12] McIntosh, J.R., Joiner, E.F., Goldberg, J.L., Murray, L.M., Yasin, B., Mendiratta, A., Karceski, S.C., Thuet, E., Modik, O., Shelkov, E., Lombardi, J.M., Sardar, Z.M., Lehman, R.A., Mandigo, C., Riew, K.D., Harel, N.Y., Virk, M.S., Carmel, J.B.: Intraoperative electrical stimulation of the human dorsal spinal cord reveals a map of arm and hand muscle responses. *Journal of Neurophysiology* **129**(1), 66–82 (2023) <https://doi.org/10.1152/jn.00235.2022>
- [13] Balbinot, G., Li, G., Kalsi-Ryan, S., Abel, R., Maier, D., Kalke, Y.-B., Weidner, N., Rupp, R., Schubert, M., Curt, A., Zariffa, J.: Segmental motor recovery after cervical spinal cord injury relates to density and integrity of corticospinal tract projections. *Nature Communications* **14**(1), 723 (2023) <https://doi.org/10.1038/s41467-023-36390-7>
- [14] Koponen, L.M., Martinez, M., Wood, E., Murphy, D.L.K., Goetz, S.M., Appelbaum, L.G., Peterchev, A.V.: Transcranial magnetic stimulation input–output curve slope differences suggest variation in recruitment across muscle representations in primary motor cortex. *Frontiers in Human Neuroscience* **18** (2024) <https://doi.org/10.3389/fnhum.2024.1310320>
- [15] Devanne, H., Lavoie, B.A., Capaday, C.: Input-output properties and gain changes in the human corticospinal pathway. *Experimental Brain Research* **114**(2), 329–338 (1997) <https://doi.org/10.1007/PL00005641>
- [16] Farzan, F.: Single-Pulse Transcranial Magnetic Stimulation (TMS) Protocols and Outcome Measures. In: Rotenberg, A., Horvath, J.C., Pascual-Leone, A. (eds.) *Transcranial Magnetic Stimulation*, pp. 69–115. Springer, New York, NY (2014). https://doi.org/10.1007/978-1-4939-0879-0_5
- [17] Schickelanz, N., Schwegler, K., Fastenrath, M., Spalek, K., Milnik, A., Papasotiropoulos, A., Nyffeler, T., Quervain, D.J.-F.: Motor threshold predicts working memory performance in healthy humans. *Annals of Clinical and Translational Neurology* **1**(1), 69–73 (2014) <https://doi.org/10.1002/acn3.22>
- [18] Vucic, S., Stanley Chen, K.-H., Kiernan, M.C., Hallett, M., Benninger, D.H., Di Lazzaro, V., Rossini, P.M., Benussi, A., Berardelli, A., Currà, A., Krieg, S.M.,

- Lefaucheur, J.-P., Long Lo, Y., Macdonell, R.A., Massimini, M., Rosanova, M., Picht, T., Stinear, C.M., Paulus, W., Ugawa, Y., Ziemann, U., Chen, R.: Clinical diagnostic utility of transcranial magnetic stimulation in neurological disorders. Updated report of an IFCN committee. *Clinical Neurophysiology* **150**, 131–175 (2023) <https://doi.org/10.1016/j.clinph.2023.03.010>
- [19] Bestmann, S.: On the use of the motor threshold as a dependent variable in TMS research. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation* **17**(4), 780–781 (2024) <https://doi.org/10.1016/j.brs.2024.06.009>
- [20] Carson, R.G., Nelson, B.D., Buick, A.R., Carroll, T.J., Kennedy, N.C., Cann, R.M.: Characterizing Changes in the Excitability of Corticospinal Projections to Proximal Muscles of the Upper Limb. *Brain Stimulation* **6**(5), 760–768 (2013) <https://doi.org/10.1016/j.brs.2013.01.016>
- [21] Potter-Baker, K.A., Varnerin, N.M., Cunningham, D.A., Roelle, S.M., Sankarabramanian, V., Bonnett, C.E., Machado, A.G., Conforto, A.B., Sakaie, K., Plow, E.B.: Influence of Corticospinal Tracts from Higher Order Motor Cortices on Recruitment Curve Properties in Stroke. *Frontiers in Neuroscience* **10** (2016) <https://doi.org/10.3389/fnins.2016.00079>
- [22] Kraus, D., Naros, G., Guggenberger, R., Leão, M.T., Ziemann, U., Gharabaghi, A.: Recruitment of Additional Corticospinal Pathways in the Human Brain with State-Dependent Paired Associative Stimulation. *Journal of Neuroscience* **38**(6), 1396–1407 (2018) <https://doi.org/10.1523/JNEUROSCI.2893-17.2017>
- [23] Iyer, P.C., Madhavan, S.: Characterization of stimulus response curves obtained with transcranial magnetic stimulation from bilateral tibialis anterior muscles post stroke. *Neuroscience Letters* **713**, 134530 (2019) <https://doi.org/10.1016/j.neulet.2019.134530>
- [24] Kemlin, C., Moulton, E., Leder, S., Houot, M., Meunier, S., Rosso, C., Lamy, J.-C.: Redundancy Among Parameters Describing the Input-Output Relation of Motor Evoked Potentials in Healthy Subjects and Stroke Patients. *Frontiers in Neurology* **10**, 535 (2019) <https://doi.org/10.3389/fneur.2019.00535>
- [25] Filipović, S.R., Kačar, A., Milanović, S., Ljubisavljević, M.R.: Neurophysiological Predictors of Response to Medication in Parkinson’s Disease. *Frontiers in Neurology* **12** (2021) <https://doi.org/10.3389/fneur.2021.763911>
- [26] Eginyan, G., Zhou, X., Williams, A.M.M., Lam, T.: Effects of motor stimulation of the tibial nerve on corticospinal excitability of abductor hallucis and pelvic floor muscles. *Frontiers in Rehabilitation Sciences* **3**, 1089223 (2023) <https://doi.org/10.3389/fresc.2022.1089223>

- [27] Murphy, H.M., Fetter, C.M., Snow, N.J., Chaves, A.R., Downer, M.B., Ploughman, M.: Lower corticospinal excitability and greater fatigue among people with multiple sclerosis experiencing pain. *Multiple Sclerosis Journal - Experimental, Translational and Clinical* **9**(1), 20552173221143398 (2023) <https://doi.org/10.1177/20552173221143398>
- [28] Antal, A., Nitsche, M.A., Kincses, T.Z., Lampe, C., Paulus, W.: No correlation between oving phosphene and motor thresholds: a transcranial magnetic stimulation study. *NeuroReport* **15**(2), 297 (2004) <https://doi.org/10.1097/00001756-200402090-00017>
- [29] Kukke, S.N., Paine, R.W., Chao, C., Campos, A.C., Hallett, M.: Efficient and reliable characterization of the corticospinal system using transcranial magnetic stimulation. *Journal of Clinical Neurophysiology* **31**(3), 246–252 (2014) <https://doi.org/10.1097/WNP.0000000000000057>
- [30] Alavi, S.M.M., Goetz, S.M., Peterchev, A.V.: Optimal Estimation of Neural Recruitment Curves Using Fisher Information: Application to Transcranial Magnetic Stimulation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **27**(6), 1320–1330 (2019) <https://doi.org/10.1109/TNSRE.2019.2914475>
- [31] Hassanzahraee, M., Zoghi, M., Jaberzadeh, S.: Longer Transcranial Magnetic Stimulation Intertrial Interval Increases Size, Reduces Variability, and Improves the Reliability of Motor Evoked Potentials. *Brain Connectivity* **9**(10), 770–776 (2019) <https://doi.org/10.1089/brain.2019.0714>
- [32] Manson, G.A., Calvert, J.S., Ling, J., Tychhon, B., Ali, A., Sayenko, D.G.: The relationship between maximum tolerance and motor activation during transcutaneous spinal stimulation is unaffected by the carrier frequency or vibration. *Physiological Reports* **8**(5), 14397 (2020) <https://doi.org/10.14814/phy2.14397>
- [33] Sharma, P., Rampersaud, H., Shah, P.K.: Repeated epidural stimulation modulates cervical spinal cord excitability in healthy adult rats. *Experimental Brain Research* **243**(1), 22 (2024) <https://doi.org/10.1007/s00221-024-06965-x>
- [34] Taylor, M.M., Creelman, C.D.: PEST: Efficient estimates on probability functions **41**(4), 782–787 (1967) <https://doi.org/10.1121/1.1910407>
- [35] Pentland, A.: Maximum likelihood estimation: The best PEST **28**(4), 377–379 (1980) <https://doi.org/10.3758/BF03204398>
- [36] Harvey, L.O.: Efficient estimation of sensory thresholds with ML-PEST (1997) <https://doi.org/10.1163/156856897X00159>
- [37] Awiszus, F.: Chapter 2 TMS and threshold hunting. In: *Supplements to Clinical Neurophysiology* vol. 56, pp. 13–23. Elsevier, ??? (2003). <https://doi.org/10.>

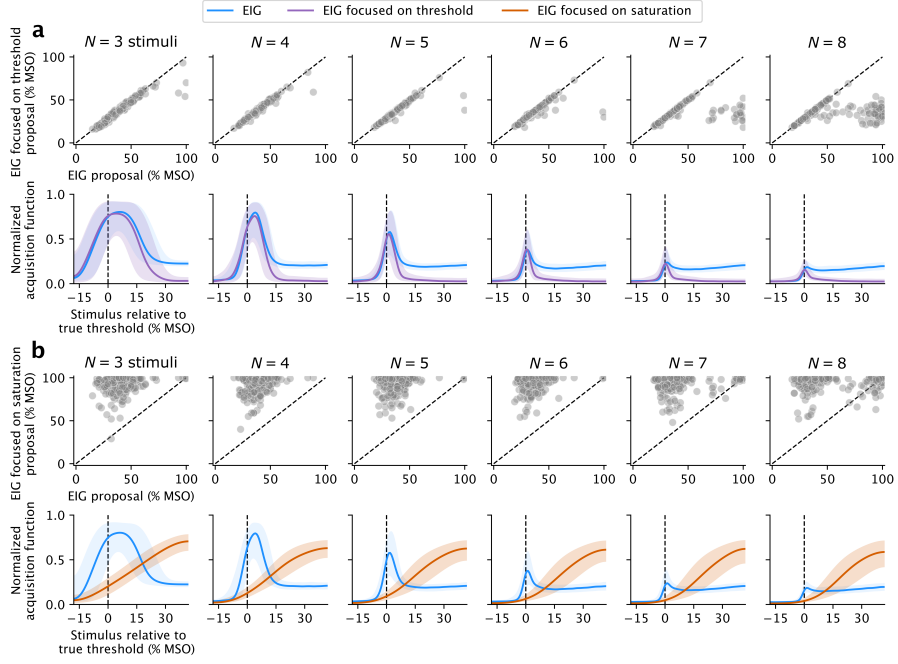
- [38] Mishory, A., Molnar, C., Koola, J., Li, X., Kozel, F.A., Myrick, H., Stroud, Z., Nahas, Z., George, M.S.: The maximum-likelihood strategy for determining transcranial magnetic stimulation motor threshold, using parameter estimation by sequential testing is faster than conventional methods with similar precision **20**(3), 160 (2004)
- [39] Wang, B., Peterchev, A.V., Goetz, S.M.: Three novel methods for determining motor threshold with transcranial magnetic stimulation outperform conventional procedures. *Journal of Neural Engineering* (2023) <https://doi.org/10.1088/1741-2552/acflcc>
- [40] Lindley, D.V.: On a measure of the information provided by an experiment **27**(4), 986–1005 (1956)
- [41] Chaloner, K., Verdinelli, I.: Bayesian experimental design: A review. *Statistical Science* **10**(3), 273–304 (1995) <https://doi.org/10.1214/ss/1177009939>
- [42] Rainforth, T., Cornish, R., Yang, H., Warrington, A., Wood, F.: On Nesting Monte Carlo Estimators. *arXiv* (2018). <https://doi.org/10.48550/arXiv.1709.06181> . <http://arxiv.org/abs/1709.06181>
- [43] Rainforth, T., Foster, A., Ivanova, D.R., Smith, F.B.: Modern Bayesian Experimental Design. *arXiv* (2023). <https://doi.org/10.48550/arXiv.2302.14545> . <http://arxiv.org/abs/2302.14545>
- [44] Tyagi, V., Murray, L.M., Asan, A.S., Mandigo, C., Virk, M.S., Harel, N.Y., Carmel, J.B., McIntosh, J.R.: Hierarchical Bayesian estimation of motor-evoked potential recruitment curves yields accurate and robust estimates. *Brain Stimulation* (2025) <https://doi.org/10.1016/j.brs.2025.09.008>
- [45] Suzuki, M., Kirimoto, H., Onishi, H., Yamada, S., Tamaki, H., Maruyama, A., Yamamoto, J.-i.: Reciprocal changes in input–output curves of motor evoked potentials while learning motor skills. *Brain Research* **1473**, 114–123 (2012) <https://doi.org/10.1016/j.brainres.2012.07.043>
- [46] Nazarova, M., Novikov, P., Ivanina, E., Kozlova, K., Dobrynina, L., Nikulin, V.V.: Mapping of multiple muscles with transcranial magnetic stimulation: absolute and relative test–retest reliability. *Human Brain Mapping* **42**(8), 2508–2528 (2021) <https://doi.org/10.1002/hbm.25383>
- [47] Yuasa, A., Uehara, S., Sawada, Y., Otaka, Y.: Systematic determination of muscle groups and optimal stimulation intensity for simultaneous TMS mapping of multiple muscles in the upper limb. *Physiological Reports* **10**(23), 15527 (2022) <https://doi.org/10.14814/phy2.15527>

- [48] Murray, L.M., McIntosh, J.R., Goldsmith, J.A., Wu, Y.-K., Liu, M., Sanford, S.P., Joiner, E.F., Mandigo, C., Tyagi, V., Virk, M.S., Carmel, J.B., Harel, N.Y.: Timing-dependent synergies between noninvasive motor cortex and spinal cord stimulation in chronic cervical spinal cord injury. *Clinical Neurophysiology* **180**, 2111372 (2025) <https://doi.org/10.1016/j.clinph.2025.2111372>
- [49] Pascual-Leone, A., Tyagi, V., Asan, A.S., Rocha-Flores, P.E., Rodriguez-Lopez, O., Voit, W., McIntosh, J.R., Carmel, J.B.: Electrode position, distance, size, and orientation determine efficacy of cervical epidural stimulation to recruit forelimb muscles in rats. *bioRxiv* (2026). <https://doi.org/10.1101/2025.09.05.674051> . <https://www.biorxiv.org/content/10.1101/2025.09.05.674051v2> Accessed 2026-09-29
- [50] Freitas, R.M., Bhatia, S., Sorensen, E., Verma, N., Carranza, E., Ensel, S., Borda, L., Boos, A., Goldsmith, J., Fisher, L.E., Fields, D.P., Powell, M.P., Gordon, S., Balzer, J., Friedlander, R.M., Wittenberg, G.F., Gerszten, P.C., Krakauer, J.W., Pirondini, E., Weber, D.J., Capogrosso, M.: Spinal cord stimulation for upper limb motor function in people with chronic post-stroke hemiparesis: a feasibility trial. *Nature Medicine* **32**(7), 2579–2590 (2026) <https://doi.org/10.1038/s41591-026-04435-1>
- [51] Hoffman, M.D., Gelman, A.: The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *arXiv* (2011). <https://doi.org/10.48550/arXiv.1111.4246>
- [52] Julkunen, P., Säisänen, L., Hukkanen, T., Danner, N., Könönen, M.: Does second-scale intertrial interval affect motor evoked potentials induced by single-pulse transcranial magnetic stimulation? *Brain Stimulation* **5**(4), 526–532 (2012) <https://doi.org/10.1016/j.brs.2011.07.006>
- [53] Weise, K., Numssen, O., Kalloch, B., Zier, A.L., Thielscher, A., Haueisen, J., Hartwigsen, G., Knösche, T.R.: Precise motor mapping with transcranial magnetic stimulation. *Nature Protocols* **18**(2), 293–318 (2023) <https://doi.org/10.1038/s41596-022-00776-6>
- [54] Hernandez-Pavon, J.C., Veniero, D., Bergmann, T.O., Belardinelli, P., Bortoletto, M., Casarotto, S., Casula, E.P., Farzan, F., Fecchio, M., Julkunen, P., Kallioniemi, E., Lioumis, P., Metsomaa, J., Miniussi, C., Mutanen, T.P., Rocchi, L., Rogasch, N.C., Shafi, M.M., Siebner, H.R., Thut, G., Zrenner, C., Ziemann, U., Ilmoniemi, R.J.: TMS combined with EEG: Recommendations and open issues for data collection and analysis. *Brain Stimulation* **16**(2), 567–593 (2023) <https://doi.org/10.1016/j.brs.2023.02.009>
- [55] Gelman, A., Tuerlinckx, F.: Type S error rates for classical and Bayesian single and multiple comparison procedures. *Computational Statistics* **15**(3), 373–390 (2000) <https://doi.org/10.1007/s001800000040>

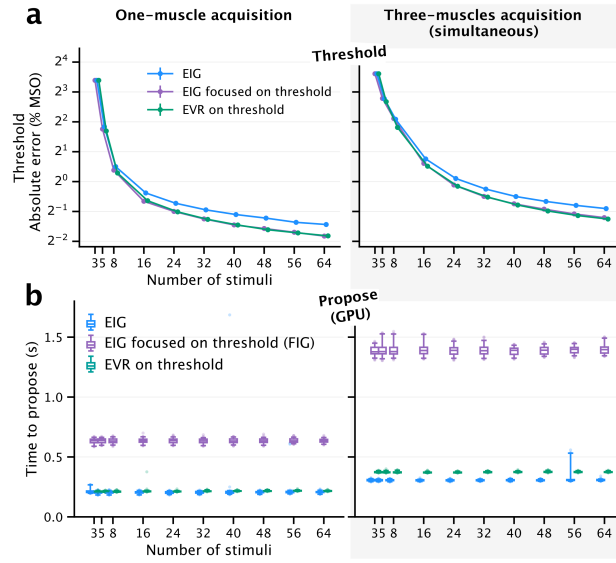
- [56] Kruschke, J.K.: Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General* **142**(2), 573–603 (2013) <https://doi.org/10.1037/a0029146>
- [57] Kruschke, J.: *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, 2nd edition edn. Academic Press, Boston (2014). <https://doi.org/10.1016/B978-0-12-405888-0.00001-5>
- [58] Vincent, B.T.: Hierarchical Bayesian estimation and hypothesis testing for delay discounting tasks. *Behavior Research Methods* **48**(4), 1608–1620 (2016) <https://doi.org/10.3758/s13428-015-0672-2>
- [59] Tran, D.M.D., McNair, N.A., Harris, J.A., Livesey, E.J.: Expected TMS excites the motor system less effectively than unexpected stimulation. *NeuroImage* **226**, 117541 (2021) <https://doi.org/10.1016/j.neuroimage.2020.117541>
- [60] Moral, P.D., Doucet, A., Jasra, A.: Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(3), 411–436 (2006) <https://doi.org/10.1111/j.1467-9868.2006.00553.x>
- [61] Chopin, N., Papaspiliopoulos, O.: *An Introduction to Sequential Monte Carlo*. Springer Series in Statistics. Springer, Cham (2020). <https://doi.org/10.1007/978-3-030-47845-2> . <https://link.springer.com/10.1007/978-3-030-47845-2>
- [62] Foster, A., Ivanova, D.R., Malik, I., Rainforth, T.: Deep Adaptive Design: Amortizing Sequential Bayesian Experimental Design. *arXiv* (2021). <https://doi.org/10.48550/arXiv.2103.02438>
- [63] Kullback, S., Leibler, R.A.: On information and sufficiency **22**(1), 79–86 (1951) <https://doi.org/10.1214/aoms/1177729694>
- [64] Huan, X., Marzouk, Y.M.: Simulation-based optimal Bayesian experimental design for nonlinear systems. *Journal of Computational Physics* **232**(1), 288–317 (2013) <https://doi.org/10.1016/j.jcp.2012.08.013>
- [65] Feng, C., Marzouk, Y.M.: A layered multiple importance sampling scheme for focused optimal Bayesian experimental design. *arXiv* (2019). <https://doi.org/10.48550/arXiv.1903.11187> . <http://arxiv.org/abs/1903.11187>
- [66] Nadaraya, E.A.: On estimating regression **9**(1), 141–142 (1964) <https://doi.org/10.1137/1109020>
- [67] Watson, G.S.: Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A* **26**, 359–372 (1964)
- [68] Vanlier, J., Tiemann, C.A., Hilbers, P.A.J., Riel, N.A.W.: A bayesian approach to targeted experiment design **28**(8), 1136–1142 (2012) <https://doi.org/10.1093/>

- [69] Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Routledge, New York (2018). <https://doi.org/10.1201/9781315140919>
- [70] Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261–272 (2020) <https://doi.org/10.1038/s41592-019-0686-2>
- [71] Halton, J.H.: On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals **2**(1), 84–90 (1960) <https://doi.org/10.1007/BF01386213>
- [72] Owen, A.B.: A randomized Halton algorithm in R. *arXiv* (2017). <https://doi.org/10.48550/arXiv.1706.02808> . <http://arxiv.org/abs/1706.02808>
- [73] Phan, D., Pradhan, N., Jankowiak, M.: Composable effects for flexible and accelerated probabilistic programming in numpyro. *arXiv* (2019) <https://doi.org/10.48550/arXiv.1912.11554>
- [74] Bingham, E., Chen, J.P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Singh, R., Szerlip, P.A., Horsfall, P., Goodman, N.D.: Pyro: Deep universal probabilistic programming. *Journal of Machine Learning Research* **20**, 28–1286 (2019) <https://doi.org/10.48550/arXiv.1810.09538>

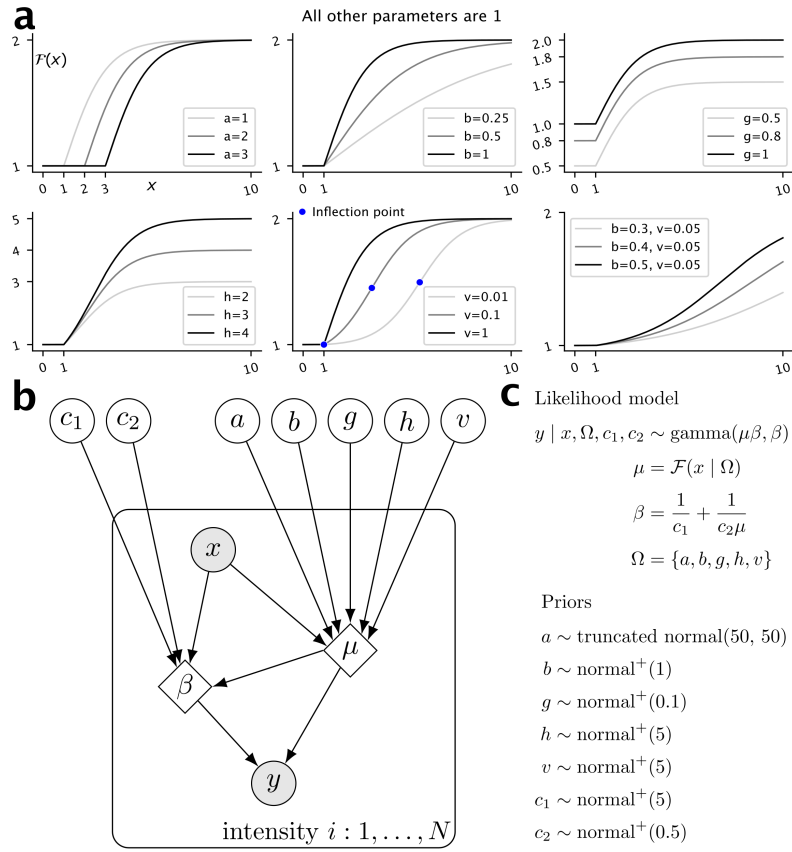
Supplementary information



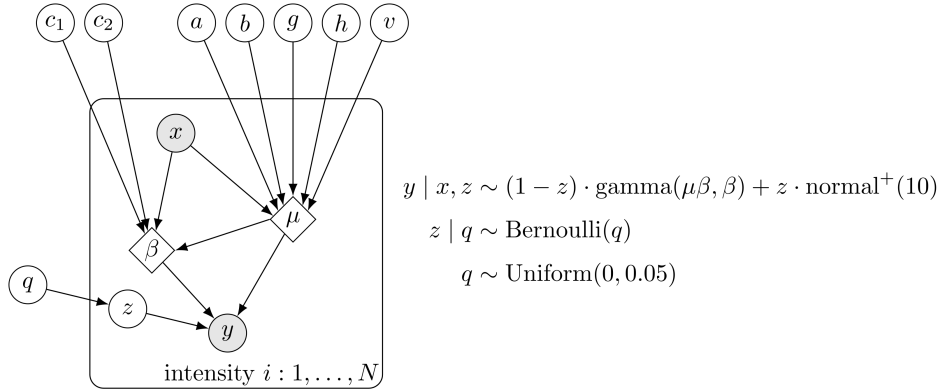
Supplementary Fig. S1. EIG proposals match EIG focused on threshold during early acquisition. (a) Comparison of stimuli proposed by EIG with those that would have been proposed by EIG focused on threshold at the same acquisition step across 200 simulations of EIG for $N \in [3, 8]$ stimuli. The dashed diagonal line indicates identical proposals. Bottom: corresponding acquisition values as a function of stimulus intensity centered at true threshold (dashed vertical line). Acquisition functions are normalized to $[0, 1]$ within each simulation run, showing that the relative information gain decreases as more samples are acquired. Lines indicate the median and shaded regions the interquartile range across simulations. (b) Same as (a), comparing EIG with EIG focused on saturation, whose proposals and acquisition functions do not agree.



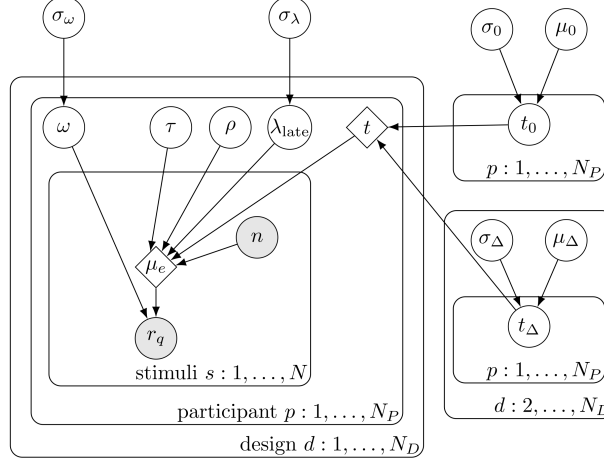
Supplementary Fig. S2. EVR matches EIG focused on threshold in accuracy while having faster proposal time. (a) Mean absolute error in threshold estimation when acquiring a single target muscle (left) and three muscles simultaneously (right). **(b)** Computation time for the proposal step.



Supplementary Fig. S3. Bayesian model specification for TMS data. (a) Effect of varying parameters of the rectified-logistic function. (b) Graphical model. Circular nodes represent random variables. Filled circular nodes represent observed data. Diamonds represent deterministic variables. Arrows indicate that a child node is informed by the distribution of its parent node. Plates denote re-instantiation of nodes. (c) Model specification with priors for TMS data. Here, $\mathcal{F}$ is the rectified-logistic function.

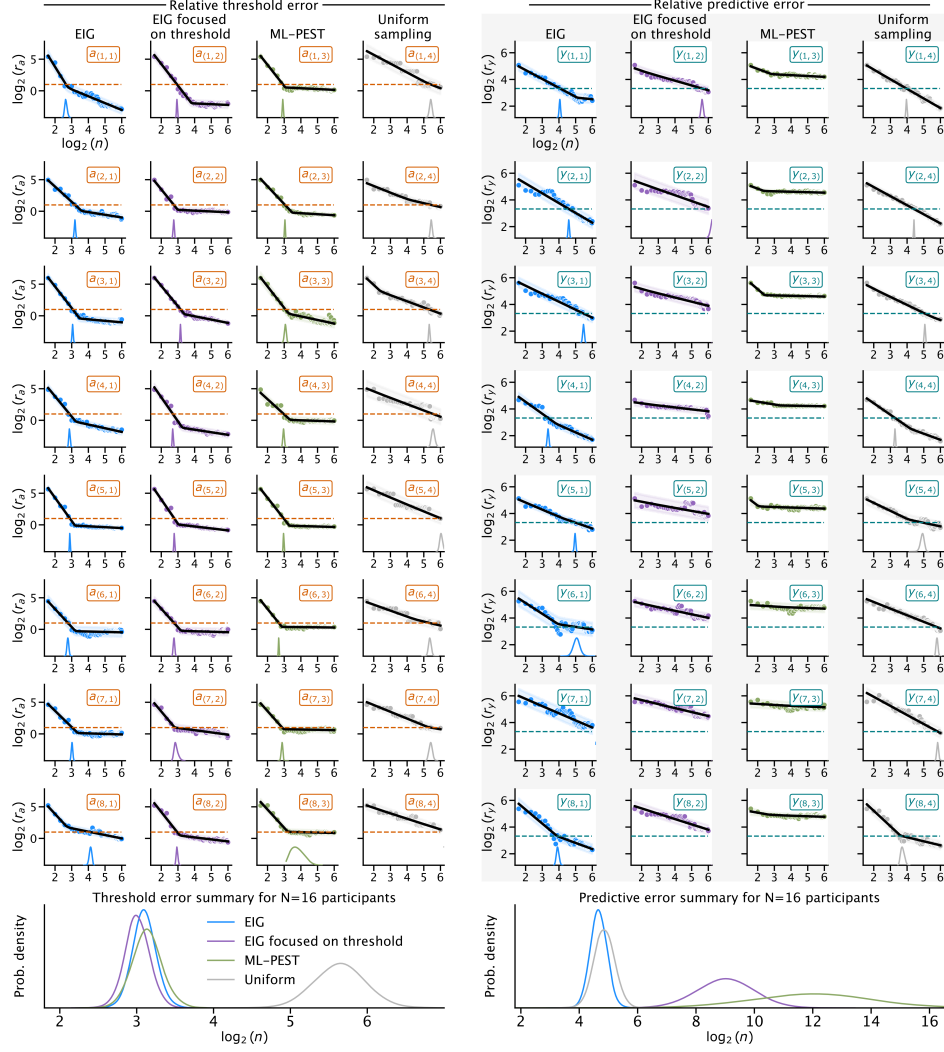


Supplementary Fig. S4. Mixture extension of the Bayesian model.

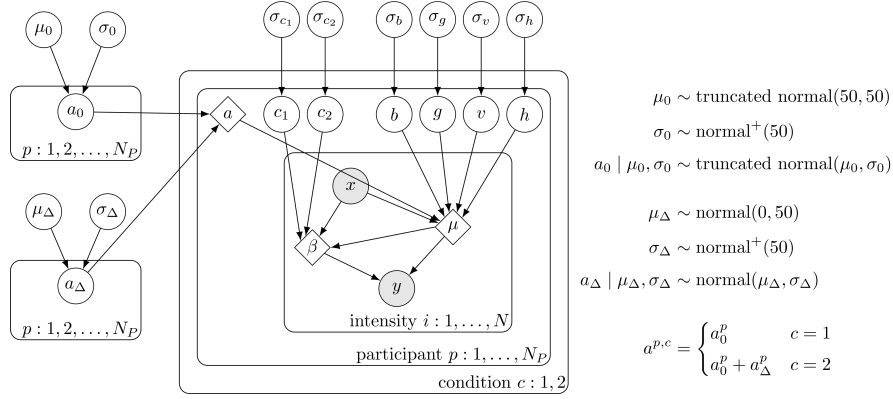


$$\begin{aligned}
\sigma_\lambda &\sim \text{lognormal}(0, 1) & \mu_0 &\sim \text{normal}(\log_2 24, 2) \\
\lambda_{\text{late}} | \sigma_\lambda &\sim \text{lognormal}(0, \sigma_\lambda) & \sigma_0 &\sim \text{normal}^+(2) \\
\rho &\sim \text{beta}(5, 2) & t_0 | \mu_0, \sigma_0 &\sim \text{normal}(\mu_0, \sigma_0) \\
\lambda_{\text{early}} &= \frac{\lambda_{\text{late}}}{\rho} & \mu_\Delta &\sim \text{normal}(0, 2) \\
\tau &\sim \text{normal}(\log_2 24, 2) & \sigma_\Delta &\sim \text{normal}^+(2) \\
\sigma_\omega &\sim \text{lognormal}(0, 1) & t_\Delta | \mu_\Delta, \sigma_\Delta &\sim \text{normal}(\mu_\Delta, \sigma_\Delta) \\
\omega | \sigma_\omega &\sim \text{normal}^+(\sigma_\omega) & t^{p,d} &= \begin{cases} t_0^p & d = 1 \\ t_0^p + t_\Delta^{p,d} & d > 1 \end{cases} \\
\log_2 r_q &\sim \text{normal}(\mu_q, \omega) \\
\mu_q &= \log_2 k_q - \lambda_{\text{early}} (\log_2 n - t) \\
&\quad - (\lambda_{\text{late}} - \lambda_{\text{early}}) [\max(0, \log_2 n - \tau) - \max(0, t - \tau)]
\end{aligned}$$

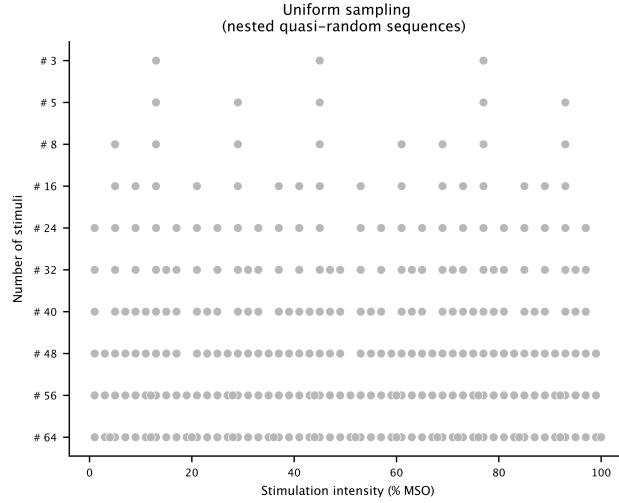
Supplementary Fig. S5. Paired comparison model for the number of stimuli needed to reach acceptable error. Here $d = 1$ represents EIG, and $d > 1$ may represent EIG focused on threshold, ML-PEST, or uniform sampling. A priori, the model assumes no difference in the number of stimuli needed to reach an acceptable median error k_q , as indicated by the weak prior on μ_Δ centered at zero.



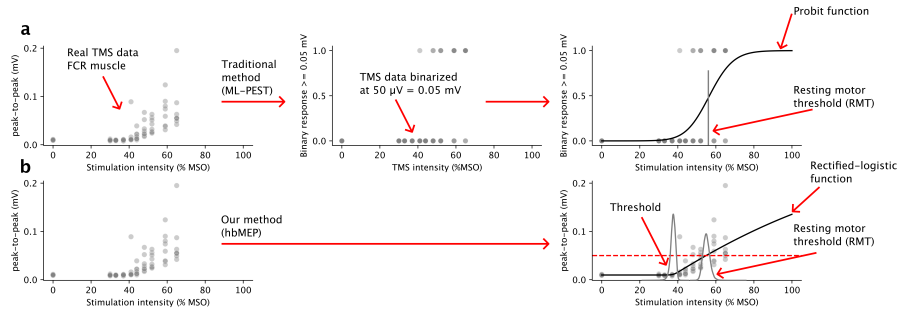
Supplementary Fig. S6. Example piecewise linear functions estimated using the paired comparison model comparing the number of stimuli needed to reach acceptable error across designs. Each row, except the last, shows one participant, with columns corresponding to EIG, EIG focused on threshold, ML-PEST, and uniform sampling. Only the first eight of the 16 participants are shown. Left: relative threshold error r_a , with the orange horizontal dashed line indicating $\log_2 k_a$, corresponding to acceptable error of 2% ($k_a = 2$). Right: relative predictive error r_v , with the teal horizontal dashed line indicating $\log_2 k_y$, corresponding to acceptable error of 10% ($k_y = 10$). The number of stimuli n and relative errors are shown in log-log space, with dots indicating the relative errors and black lines the estimated piecewise linear functions. Inset: posterior distribution of the number of stimuli needed to reach the corresponding acceptable error. A missing or truncated posterior indicates that the estimated number of stimuli needed to reach acceptable error extends beyond the range shown. Bottom row: posterior distributions of population-level location μ_0 for EIG and $(\mu_0 + \mu_\Delta)$ for each comparison design, summarizing the number of stimuli needed to reach acceptable error across 16 participants.



Supplementary Fig. S7. Paired comparison model for detecting a shift in threshold between pre- and post-intervention phases. Here $c = 1$ and $c = 2$ represent pre- and post-intervention conditions, respectively. A priori, the model assumes no shift, as indicated by the weak prior on μ_Δ centered at zero.



Supplementary Fig. S8. Uniform sampling using nested quasi-random sequence. Stimuli remain approximately evenly distributed over $[0, 100]\%$ MSO at any number of acquired samples from $n = 3$ to $n = 64$, with each larger set containing all samples from the smaller sets.



Supplementary Fig. S9. ML-PEST is designed for the resting motor threshold. (a) ML-PEST binarizes the MEP sizes using the conventional 50 μV cutoff and fits a probit function to the resulting binary responses. The resting motor threshold (RMT) is the stimulation intensity at which the fitted probability of observing an MEP larger than 50 μV is 0.5. ML-PEST provides a point estimate of the RMT. **(b)** In contrast, hbMEP directly models the observed MEP sizes using the rectified-logistic function and defines the threshold as the stimulation intensity at which the expected MEP size deflects from the baseline response. The RMT can additionally be obtained from hbMEP as a postprocessing step, which agrees with the ML-PEST estimate. However, the threshold cannot be estimated using ML-PEST. Additionally, ML-PEST fails when the baseline response exceeds the fixed 50 μV cutoff, whereas hbMEP can accommodate different baseline response levels since it estimates the baseline within the same model.