Rebuttal - Supporting Information

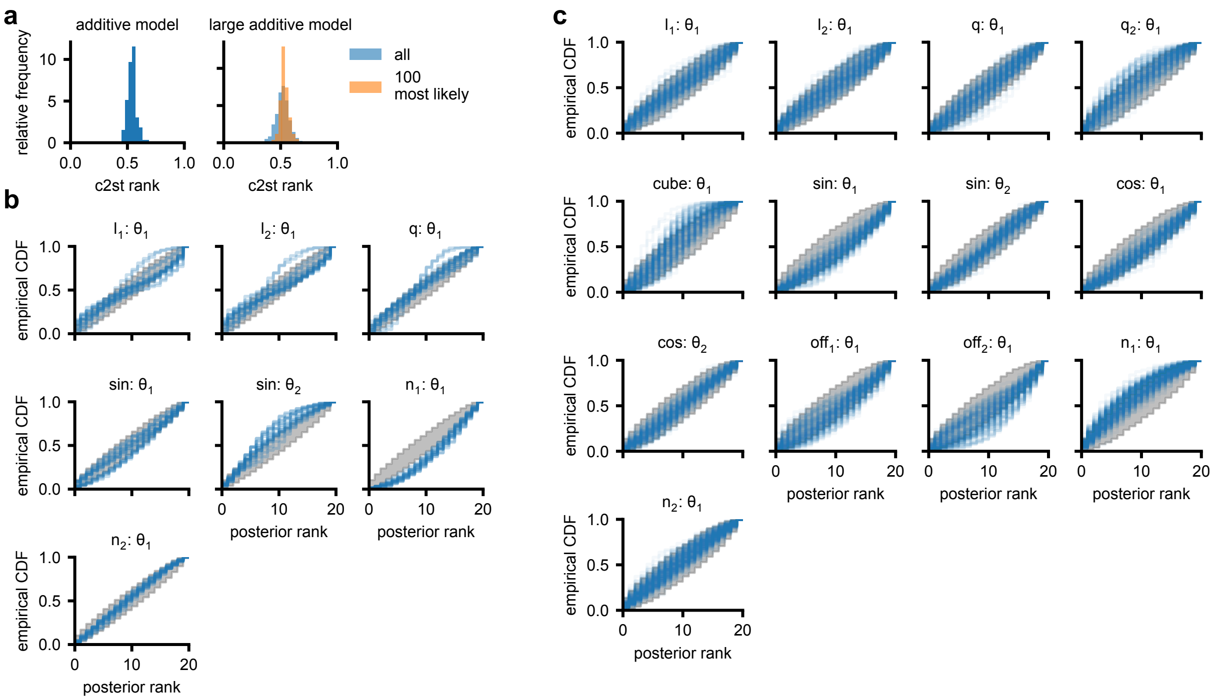

Figure R1: **Simulation-based calibration of the parameter posterior for the additive models. (a)** Histogram of the c2st ranks for the two additive models with 6 and 11 components (referred to as *small* and *large*). A value of 0.5 indicates a well calibrated posterior for which the rank statistics are indistinguishable from a uniform distribution. **(b)** Posterior calibration of the *small* additive model, by individual model parameters for all possible model component combinations. Grey regions indicate the 99% confidence intervals of a uniform distribution, given the provided number of samples. **(c)** Same as (b) for the *large* additive model and for all models with at least 50 samples in the test dataset of 100k samples. For all plots we ranked the true parameter θ_o against 1k posterior samples.

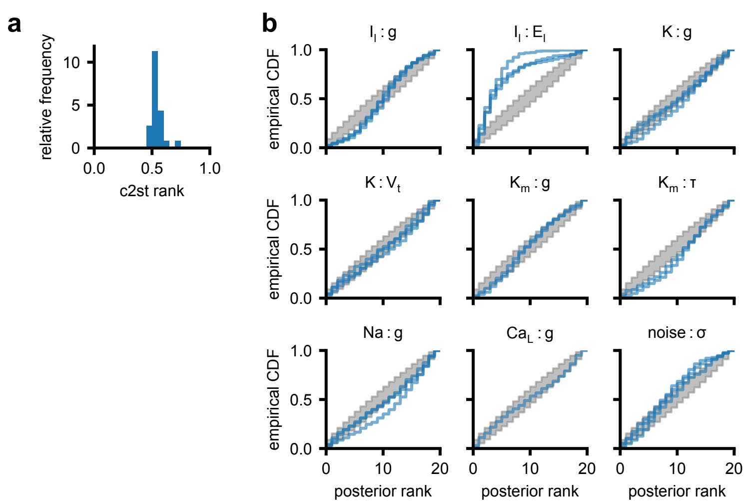

Figure R2: **Simulation-based calibration of the parameter posterior for the Hodgkin-Huxley model. (a)** Histogram of the c2st ranks. A value of 0.5 indicates a well calibrated posterior for which the rank statistics are indistinguishable from a uniform distribution. **(b)** Posterior calibration by individual model parameters for all possible model component combinations. Grey regions indicate the 99% confidence intervals of a uniform distribution, given the provided number of samples. For all plots we ranked the true parameter θ_o against 1k posterior samples.

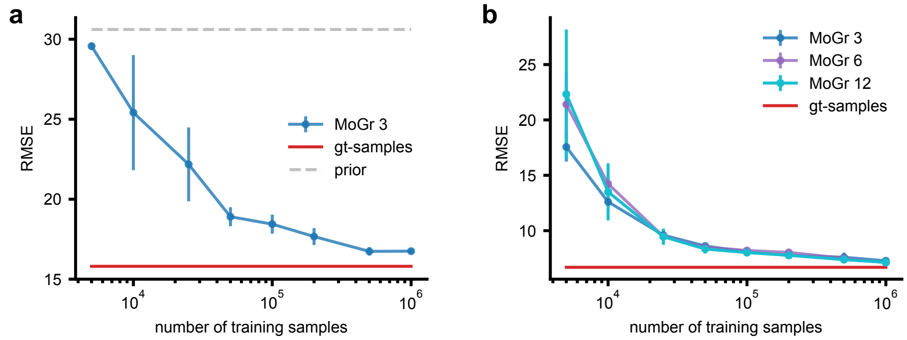

Figure R3: **SBMI performance scales to larger models and is robust to the number of mixture components. (a)** Posterior predictive performance for an additive model with 20 components (specified in Table R1) and $2^{20} \approx 1M$ possible combinations of model components. A categorical distribution can not be fitted anymore as we have very few to no samples for each model in a dataset of size $1M$. **(b)** Posterior predictive performance for the large model with 11 components, and $2^{11} = 2048$ possible combinations of model components (see main paper for details). We compare SBMI with 3, 6 and 12 components for the MoGr model posterior as well as for the mixture of Gaussian parameter posterior.

Table R1: **Details for the 1M additive model.**

Model Component	Token	Parameter Prior
$\theta_1 \cdot t$	l_1	$\theta_1 \sim \mathcal{U}(-2, 2)$
$\theta_1 \cdot t$	l_2	$\theta_1 \sim \mathcal{U}(-2, 2)$
$\theta_1 \cdot t^2$	q_1	$\theta_1 \sim \mathcal{U}(-0.5, 0.5)$
$(\theta_1 + t)^2$	q_2	$\theta_1 \sim \mathcal{U}(-5, 0)$
$\theta_1 \cdot t^3$	cub	$\theta_1 \sim \mathcal{U}(-0.1, 0.1)$
$\theta_1 \cdot \sin(\theta_2 t)$	sin	$\theta_1 \sim \mathcal{U}(0, 5)$ $\theta_2 \sim \mathcal{U}(0.5, 5)$
$\theta_1 \cdot \cos(\theta_2 t)$	cos	$\theta_1 \sim \mathcal{U}(0, 5)$ $\theta_2 \sim \mathcal{U}(0.5, 5)$
θ_1	$const_1$	$\theta_1 \sim \mathcal{U}(-5, 5)$
θ_1	$const_2$	$\theta_1 \sim \mathcal{U}(0, 10)$
$\theta_1 \cdot \tanh(t - \theta_2)$	$tanh_1$	$\theta_1 \sim \mathcal{U}(1, 10)$ $\theta_2 \sim \mathcal{U}(2, 8)$
$\theta_1 \cdot \tanh(\theta_2 - t)$	$tanh_2$	$\theta_1 \sim \mathcal{U}(1, 10)$ $\theta_2 \sim \mathcal{U}(2, 8)$
$\theta_1 \cdot \exp(-(t - \theta_2)^2)$	g_1	$\theta_1 \sim \mathcal{U}(1, 10)$ $\theta_2 \sim \mathcal{U}(2, 8)$
$\theta_1 \cdot \exp(-(t - \theta_2)^2/8$	g_2	$\theta_1 \sim \mathcal{U}(1, 10)$ $\theta_2 \sim \mathcal{U}(2, 8)$
$\theta_1 \cdot \mathrm{ReLU}(t - \theta_2)$	$relu_1$	$\theta_1 \sim \mathcal{U}(1, 5)$ $\theta_2 \sim \mathcal{U}(2, 8)$
$\theta_1 \cdot \mathrm{ReLU}(\theta_2 - t)$	$relu_2$	$\theta_1 \sim \mathcal{U}(1, 5)$ $\theta_2 \sim \mathcal{U}(2, 8)$
noise$_1$: $n_{t_i} \sim \mathcal{N}(0, \theta_1)$	n_1	$\theta_1 \sim \mathcal{U}(0.1, 2)$
noise$_2$: $n_{t_i} \sim (t_i + 1)\mathcal{N}(0, \theta_1)$	n_2	$\theta_1 \sim \mathcal{U}(0.5, 2)$
noise$_3$: $n_{t_i} \sim (11 - t_i)\mathcal{N}(0, \theta_1)$	n_3	$\theta_1 \sim \mathcal{U}(0.5, 2)$
noise$_4$: $n_{t_i} \sim (t_i^2 + 1)\mathcal{N}(0, \theta_1)$	n_4	$\theta_1 \sim \mathcal{U}(0.2, 1)$
noise$_5$: $n_{t_i} \sim (11 - t_i^2)\mathcal{N}(0, \theta_1)$	n_5	$\theta_1 \sim \mathcal{U}(0.2, 1)$

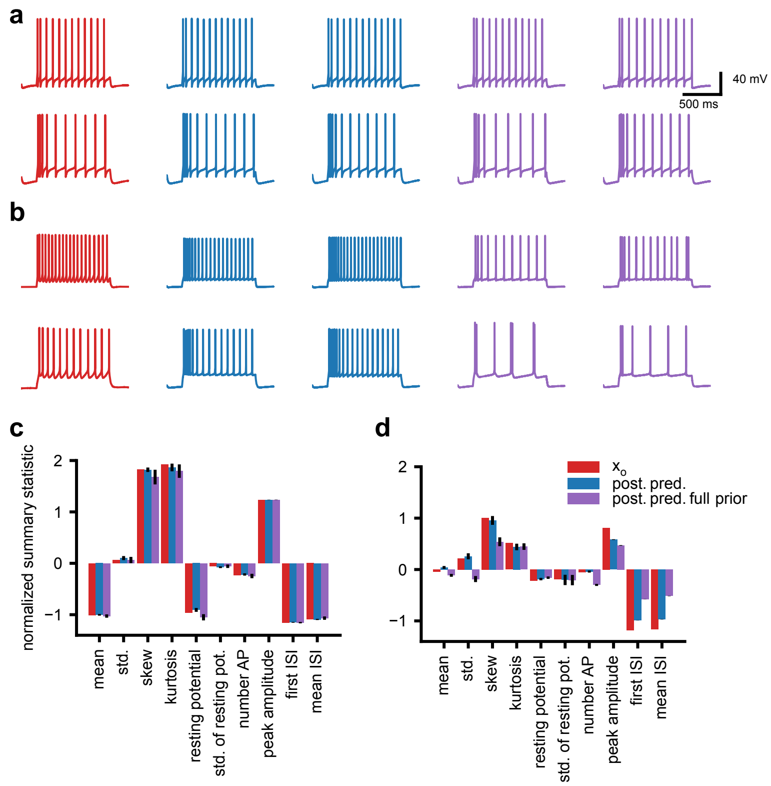

Figure R4: **Leveraging domain knowledge enhances posterior performance for the Hodgkin-Huxley model.** **(a)** Two synthetic samples (red) with two posterior predictive samples each (blue) (same as Fig. 7a of main paper) with additional two posterior predictives from a model trained on a fully connected model prior (violet). **(b)** Two voltage recordings from the Allen Cell database (red) with two posterior predictive samples each for the standard model (blue) and the model trained on a fully connected model prior (violet). **(c)** Ten example summary statistics for the upper trace in (a) and summary statistics for ten posterior predictive samples from the respective model (see also Fig. S9a for details). **(d)** Ten example summary statistics for the voltage recording from the Allen dataset shown in (b), upper trace, and summary statistics for ten posterior predictive samples from the respective model (see also Fig. S9b for details). We report mean±std. for ten posterior predictive samples.