

Interrogating theoretical models of neural computation with emergent property inference

Sean R. Bittner¹, Agostina Palmigiano¹, Alex T. Piet^{2,3,4}, Chunyu A. Duan⁵, Carlos D. Brody^{2,3,6}, Kenneth D. Miller¹, and John P. Cunningham⁷.

¹Department of Neuroscience, Columbia University,

²Princeton Neuroscience Institute,

³Princeton University,

⁴Allen Institute for Brain Science,

⁵Institute of Neuroscience, Chinese Academy of Sciences,

⁶Howard Hughes Medical Institute,

⁷Department of Statistics, Columbia University

¹ 1 Abstract

² A cornerstone of theoretical neuroscience is the circuit model: a system of equations that captures
³ a hypothesized neural mechanism. Such models are valuable when they give rise to an experimen-
⁴ tally observed phenomenon – whether behavioral or a pattern of neural activity – and thus can
⁵ offer insights into neural computation. The operation of these circuits, like all models, critically
⁶ depends on the choice of model parameters. A key step is then to identify the model parameters
⁷ consistent with observed phenomena: to solve the inverse problem. In this work, we present a
⁸ novel technique, emergent property inference (EPI), that brings the modern probabilistic modeling
⁹ toolkit to theoretical neuroscience. When theorizing circuit models, theoreticians predominantly
¹⁰ focus on reproducing computational properties rather than a particular dataset. Our method uses
¹¹ deep neural networks to learn parameter distributions with these computational properties. This
¹² methodology is introduced through a motivational example inferring conductance parameters in a
¹³ circuit model of the stomatogastric ganglion. Then, with recurrent neural networks of increasing
¹⁴ size, we show that EPI allows precise control over the behavior of inferred parameters, and that
¹⁵ EPI scales better in parameter dimension than alternative techniques. In the remainder of this
¹⁶ work, we present novel theoretical findings gained through the examination of complex parametric
¹⁷ structure captured by EPI. In a model of primary visual cortex, we discovered how connectivity
¹⁸ with multiple inhibitory subtypes shapes variability in the excitatory population. Finally, in a
¹⁹ model of superior colliculus, we identified and characterized two distinct regimes of connectivity

20 that facilitate switching between opposite tasks amidst interleaved trials, characterized each regime
21 via insights afforded by EPI, and found conditions where these circuit models reproduce results
22 from optogenetic silencing experiments. Beyond its scientific contribution, this work illustrates
23 the variety of analyses possible once deep learning is harnessed towards solving theoretical inverse
24 problems.

25 2 Introduction

26 The fundamental practice of theoretical neuroscience is to use a mathematical model to understand
27 neural computation, whether that computation enables perception, action, or some intermediate
28 processing. A neural circuit is systematized with a set of equations – the model – and these
29 equations are motivated by biophysics, neurophysiology, and other conceptual considerations [1–5].

30 The function of this system is governed by the choice of model *parameters*, which when configured
31 in a particular way, give rise to a measurable signature of a computation. The work of analyzing
32 a model then requires solving the inverse problem: given a computation of interest, how can we
33 reason about the distribution of parameters that give rise to it? The inverse problem is crucial for
34 reasoning about likely parameter values, uniquenesses and degeneracies, and predictions made by
35 the model [6–8].

36 Ideally, one carefully designs a model and analytically derives how computational properties deter-
37 mine model parameters. Seminal examples of this gold standard include our field’s understanding
38 of memory capacity in associative neural networks [9], chaos and autocorrelation timescales in ran-
39 dom neural networks [10], central pattern generation [11], the paradoxical effect [12], and decision
40 making [13]. Unfortunately, as circuit models include more biological realism, theory via analytical
41 derivation becomes intractable. Absent this analysis, statistical inference offers a toolkit by which
42 to solve the inverse problem by identifying, at least approximately, the distribution of parameters
43 that produce computations in a biologically realistic model [14–19].

44 Statistical inference, of course, requires quantification of the sometimes vague term *computation*.
45 In neuroscience, two perspectives are dominant. First, often we directly use an *exemplar dataset*:
46 a collection of samples that express the computation of interest, this data being gathered either
47 experimentally in the lab or from a computer simulation. Though a natural choice given its con-
48 nection to experiment [20], some drawbacks exist: these data are well known to have features
49 irrelevant to the computation of interest [21–23], confounding inferences made on such data. Re-

50 lated to this point, use of a conventional dataset encourages conventional data likelihoods or loss
51 functions, which focus on some global metric like squared error or marginal evidence, rather than
52 the computation itself.

53 Alternatively, researchers often quantify an *emergent property* (EP): a statistic of data that directly
54 quantifies the computation of interest, wherein the dataset is implicit. While such a choice may
55 seem esoteric, it is not: the above “gold standard” examples [9–13] all quantify and focus on
56 some derived feature of the data, rather than the data drawn from the model. An emergent
57 property is of course a dataset by another name, but it suggests different approach to solving
58 the same inverse problem: here we directly specify the desired emergent property – a statistic
59 of data drawn from the model – and the value we wish that property to have, and we set up
60 an optimization program to find the distribution of parameters that produce this computation.
61 This statistical framework is not new: it is intimately connected to the literature on approximate
62 bayesian computation [24–26], parameter sensitivity analyses [27–30], maximum entropy modeling
63 [31–33], and approximate bayesian inference [34, 35]; we detail these connections in Section 5.1.1.

64 The parameter distributions producing a computation may be curved or multimodal along vari-
65 ous parameter axes and combinations. It is by quantifying this complex structure that emergent
66 property inference offers scientific insight. Traditional approximation families (e.g. mean-field or
67 mixture of gaussians) are limited in the distributional structure they may learn. To address such re-
68 strictions on expressivity, advances in machine learning have used deep probability distributions as
69 flexible approximating families for such complicated distributions [36, 37] (see Section 5.1.2). How-
70 ever, the adaptation of deep probability distributions to the problem of theoretical circuit analysis
71 requires recent developments in deep learning for constrained optimization [38], and architectural
72 choices for efficient and expressive deep generative modeling [39, 40]. We detail our method, which
73 we call emergent property inference (EPI) in Section 3.2.

74 Equipped with this method, we demonstrate the capabilities of EPI and present novel theoretical
75 findings from its analysis. First, we show EPI’s ability to handle biologically realistic circuit models
76 using a five-neuron model of the stomatogastric ganglion [41]: a neural circuit whose parametric
77 degeneracy is closely studied [42]. Then, we show EPI’s scalability to high dimensional parameter
78 distributions by inferring connectivities of recurrent neural networks that exhibit stable, yet ampli-
79 fied responses – a hallmark of neural responses throughout the brain [43–45]. In a model of primary
80 visual cortex [46, 47], EPI reveals how the recurrent processing across different neuron-type popu-
81 lations shapes excitatory variability: a finding that we show is analytically intractable. Finally, we

82 investigated the possible connectivities of a superior colliculus model that allow execution of differ-
83 ent tasks on interleaved trials [48]. EPI discovered a rich distribution containing two connectivity
84 regimes with different solution classes. We queried the deep probability distribution learned by
85 EPI to produce a mechanistic understanding of neural responses in each regime. Intriguingly, the
86 inferred connectivities of each regime reproduced results from optogenetic inactivation experiments
87 in markedly different ways. These theoretical insights afforded by EPI illustrate the value of deep
88 inference for the interrogation of neural circuit models.

89 **3 Results**

90 **3.1 Motivating emergent property inference of theoretical models**

91 Consideration of the typical workflow of theoretical modeling clarifies the need for emergent prop-
92 erty inference. First, one designs or chooses an existing circuit model that, it is hypothesized,
93 captures the computation of interest. To ground this process in a well-known example, consider
94 the stomatogastric ganglion (STG) of crustaceans, a small neural circuit which generates multiple
95 rhythmic muscle activation patterns for digestion [49]. Despite full knowledge of STG connectivity
96 and a precise characterization of its rhythmic pattern generation, biophysical models of the STG
97 have complicated relationships between circuit parameters and computation [15, 42].

98 A subcircuit model of the STG [41] is shown schematically in Figure 1A. The fast population (f_1
99 and f_2) represents the subnetwork generating the pyloric rhythm and the slow population (s_1 and
100 s_2) represents the subnetwork of the gastric mill rhythm. The two fast neurons mutually inhibit
101 one another, and spike at a greater frequency than the mutually inhibiting slow neurons. The
102 hub neuron couples with either the fast or slow population, or both depending on modulatory
103 conditions. The jagged connections indicate electrical coupling having electrical conductance g_{el} ,
104 smooth connections in the diagram are inhibitory synaptic projections having strength g_{synA} onto
105 the hub neuron, and $g_{synB} = 5nS$ for mutual inhibitory connections. Note that the behavior of this
106 model will be critically dependent on its parameterization – the choices of conductance parameters
107 $\mathbf{z} = [g_{el}, g_{synA}]$.

108 Second, once the model is selected, one must specify what the model should produce. In this STG
109 model, we are concerned with neural spiking frequency, which emerges from the dynamics of the
110 circuit model (Figure 1B). An emergent property studied by Gutierrez et al. is the hub neuron firing
111 at an intermediate frequency between the intrinsic spiking rates of the fast and slow populations.

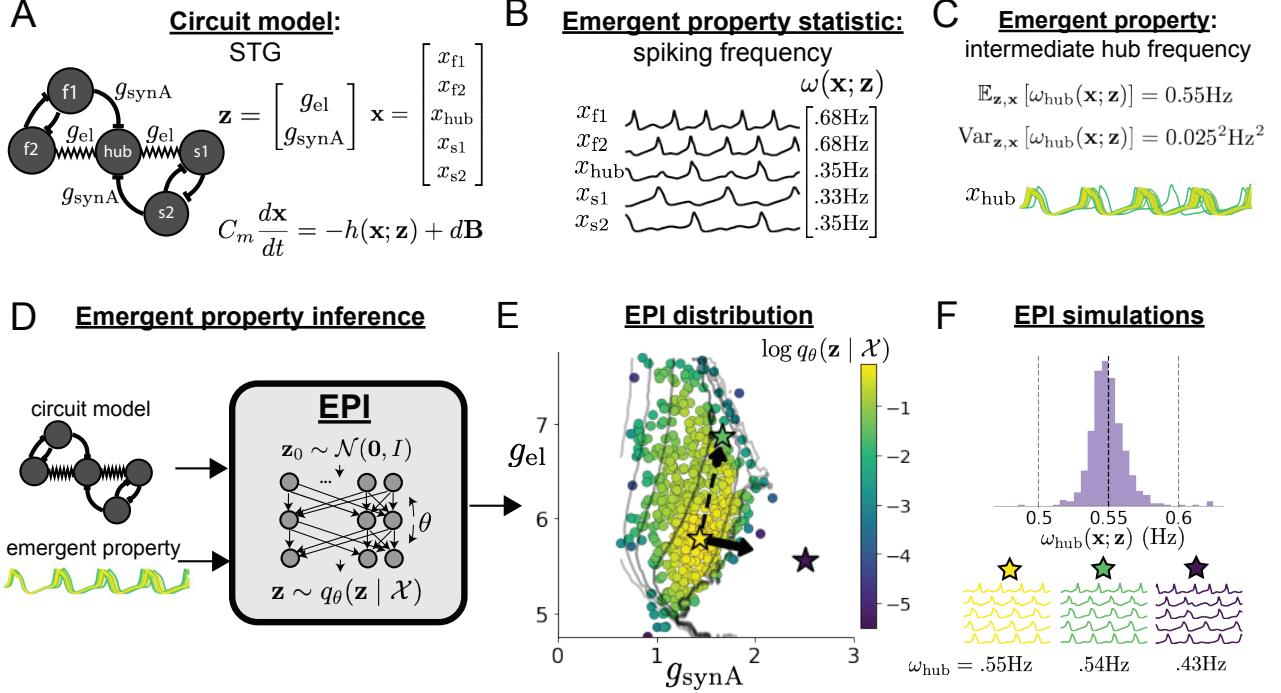


Figure 1: Emergent property inference in the stomatogastric ganglion. **A.** Conductance-based subcircuit model of the STG. **B.** Spiking frequency $\omega(\mathbf{x}; \mathbf{z})$ is an emergent property statistic. Simulated at $g_{\text{el}} = 4.5\text{nS}$ and $g_{\text{synA}} = 3\text{nS}$. **C.** The emergent property of intermediate hub frequency. Simulated activity traces are colored by log probability of generating parameters in the EPI distribution (Panel E). **D.** For a choice of circuit model and emergent property, EPI learns a deep probability distribution of parameters \mathbf{z} . **E.** The EPI distribution producing intermediate hub frequency. Samples are colored by log probability density. Contours of hub neuron frequency error are shown at levels of $.525, .53, \dots, .575$ Hz (dark to light gray away from mean). Dimension of sensitivity \mathbf{v}_1 (solid arrow) and robustness \mathbf{v}_2 (dashed arrow). **F** (Top) The predictions of the EPI distribution. The black and gray dashed lines show the mean and two standard deviations according the emergent property. (Bottom) Simulations at the starred parameter values.

112 This emergent property (EP) is shown in Figure 1C at an average frequency of 0.55Hz. To be
113 precise, we define intermediate hub frequency not strictly as 0.55Hz, but frequencies of moderate
114 deviation from 0.55Hz between the fast (.35Hz) and slow (.68Hz) frequencies.

115 Third, the model parameters producing the emergent property are inferred. By precisely quantify-
116 ing the emergent property of interest as a statistical feature of the model, we use emergent property
117 inference (EPI) to condition directly on this emergent property. Before presenting technical details
118 (in the following section), let us understand emergent property inference schematically. EPI (Fig-
119 ure 1D) takes, as input, the model and the specified emergent property, and as its output, returns
120 the parameter distribution (Figure 1E). This distribution – represented for clarity as samples from
121 the distribution – is a parameter distribution constrained such that the circuit model produces the
122 emergent property. Once EPI is run, the returned distribution can be used to efficiently gener-
123 ate additional parameter samples. Most importantly, the inferred distribution can be efficiently
124 queried to quantify the parametric structure that it captures. By quantifying the parametric struc-
125 ture governing the emergent property, EPI informs the central question of this inverse problem:
126 what aspects or combinations of model parameters have the desired emergent property?

127 3.2 Emergent property inference via deep generative models

128 EPI formalizes the three-step procedure of the previous section with deep probability distributions
129 [36, 37]. First, as is typical, we consider the model as a coupled set of noisy differential equations.
130 In this STG example, the model activity (or state) $\mathbf{x} = [x_{f1}, x_{f2}, x_{hub}, x_{s1}, x_{s2}]$ is the membrane
131 potential for each neuron, which evolves according to the biophysical conductance-based equation:

$$C_m \frac{d\mathbf{x}(t)}{dt} = -h(\mathbf{x}(t); \mathbf{z}) + d\mathbf{B} \quad (1)$$

132 where $C_m = 1\text{nF}$, and \mathbf{h} is a sum of the leak, calcium, potassium, hyperpolarization, electrical, and
133 synaptic currents, all of which have their own complicated dependence on activity \mathbf{x} and parameters
134 $\mathbf{z} = [g_{el}, g_{synA}]$, and $d\mathbf{B}$ is white gaussian noise [41] (see Section 5.2.1 for more detail).

135 Second, we determine that our model should produce the emergent property of “intermediate hub
136 frequency” (Figure 1C). We stipulate that the hub neuron’s spiking frequency – denoted by statistic
137 $\omega_{hub}(\mathbf{x})$ – is close to a frequency of 0.55Hz, between that of the slow and fast frequencies. Mathe-
138 matically, we define this emergent property with two constraints: that the mean hub frequency is
139 0.55Hz,

$$\mathbb{E}_{\mathbf{z}, \mathbf{x}} [\omega_{hub}(\mathbf{x}; \mathbf{z})] = 0.55 \quad (2)$$

140 and that the variance of the hub frequency is moderate

$$\text{Var}_{\mathbf{z}, \mathbf{x}} [\omega_{\text{hub}}(\mathbf{x}; \mathbf{z})] = 0.025^2. \quad (3)$$

141 In the emergent property of intermediate hub frequency, the statistic of hub neuron frequency is
142 an expectation over the distribution of parameters \mathbf{z} and the distribution of the data \mathbf{x} that those
143 parameters produce. We define the emergent property \mathcal{X} as the collection of these two constraints.
144 In general, an emergent property is a collection of constraints on statistical moments that together
145 define the computation of interest.

146 Third, we perform emergent property inference: we find a distribution over parameter configura-
147 tions \mathbf{z} of models that produce the emergent property; in other words, they satisfy the constraints
148 introduced in Equations 2 and 3. This distribution will be chosen from a family of probability
149 distributions $\mathcal{Q} = \{q_{\theta}(\mathbf{z}) : \theta \in \Theta\}$, defined by a deep neural network [36, 37] (Figure 1D, EPI box).
150 Deep probability distributions map a simple random variable \mathbf{z}_0 (e.g. an isotropic gaussian) through
151 a deep neural network with weights and biases θ to parameters $\mathbf{z} = g_{\theta}(\mathbf{z}_0)$ of a suitably compli-
152 cated distribution (see Section 5.1.2 for more details). Many distributions in \mathcal{Q} will respect the
153 emergent property constraints, so we select the most random (highest entropy) distribution, which
154 also means this approach is equivalent to bayesian variational inference (see Section 5.1.6). In EPI
155 optimization, stochastic gradient steps in θ are taken such that entropy is maximized, and the
156 emergent property \mathcal{X} is produced (see Section 5.1). We then denote the inferred EPI distribution
157 as $q_{\theta}(\mathbf{z} | \mathcal{X})$, since the structure of the learned parameter distribution is determined by weights
158 and biases θ , and this distribution is conditioned upon emergent property \mathcal{X} .

159 The structure of the inferred parameter distributions of EPI can be analyzed to reveal key infor-
160 mation about how the circuit model produces the emergent property. As probability in the EPI
161 distribution decreases away from the mode of $q_{\theta}(\mathbf{z} | \mathcal{X})$ (Figure 1E yellow star), the emergent prop-
162 erty deteriorates. Perturbing \mathbf{z} along a dimension in which $q_{\theta}(\mathbf{z} | \mathcal{X})$ changes little will not disturb
163 the emergent property, making this parameter combination *robust* with respect to the emergent
164 property. In contrast, if \mathbf{z} is perturbed along a dimension with strongly decreasing $q_{\theta}(\mathbf{z} | \mathcal{X})$,
165 that parameter combination is deemed *sensitive* [27, 30]. By querying the second order derivative
166 (Hessian) of $\log q_{\theta}(\mathbf{z} | \mathcal{X})$ at a mode, we can quantitatively identify how sensitive (or robust) each
167 eigenvector is by its eigenvalue; the more negative, the more sensitive and the closer to zero, the
168 more robust (see Section 5.2.4). Indeed, samples equidistant from the mode along these dimensions
169 of sensitivity (\mathbf{v}_1 , smaller eigenvalue) and robustness (\mathbf{v}_2 , greater eigenvalue) (Figure 1E, arrows)
170 agree with error contours (Figure 1E contours) and have diminished or preserved hub frequency,

171 respectively (Figure 1F activity traces). The directionality of \mathbf{v}_2 suggests that changes in conduction
 172 along this parameter combination will most preserve hub neuron firing between the intrinsic
 173 rates of the pyloric and gastric mill rhythms. Importantly and unlike alternative techniques, once
 174 an EPI distribution has been learned, the modes and Hessians of the distribution can be measured
 175 with trivial computation (see Section 5.1.2).

176 In the following sections, we demonstrate EPI on three neural circuit models across ranges of
 177 biological realism, neural system function, and network scale. First, we demonstrate the superior
 178 scalability of EPI compared to alternative techniques by inferring high-dimensional distributions
 179 of recurrent neural network connectivities that exhibit amplified, yet stable responses. Next, in a
 180 model of primary visual cortex [46,47], we show how EPI discovers parametric degeneracy, revealing
 181 how input variability across neuron types affects the excitatory population. Finally, in a model of
 182 superior colliculus [48], we used EPI to capture multiple parametric regimes of task switching, and
 183 queried the dimensions of parameter sensitivity to characterize each regime.

184 **3.3 Scaling inference of recurrent neural network connectivity with EPI**

185 To understand how EPI scales in comparison to existing techniques, we consider recurrent neu-
 186 ral networks (RNNs). Transient amplification is a hallmark of neural activity throughout cortex,
 187 and is often thought to be intrinsically generated by recurrent connectivity in the responding cor-
 188 tical area [43–45]. It has been shown that to generate such amplified, yet stabilized responses,
 189 the connectivity of RNNs must be non-normal [43, 50], and satisfy additional constraints [51]. In
 190 theoretical neuroscience, RNNs are optimized and then examined to show how dynamical systems
 191 could execute a given computation [52, 53], but such biologically realistic constraints on connec-
 192 tivity [43, 50, 51] are ignored for simplicity or because constrained optimization is difficult. In
 193 general, access to distributions of connectivity that produce theoretical criteria like stable amplifi-
 194 cation, chaotic fluctuations [10], or low tangling [54] would add scientific value to existing research
 195 with RNNs. Here, we use EPI to learn RNN connectivities producing stable amplification, and
 196 demonstrate the superior scalability and efficiency of EPI to alternative approaches.

197 We consider a rank-2 RNN with N neurons having connectivity $W = UV^\top$ and dynamics

$$\tau \dot{\mathbf{x}} = -\mathbf{x} + W\mathbf{x}, \quad (4)$$

198 where $U = [\mathbf{U}_1 \ \mathbf{U}_2] + g\chi^{(U)}$, $V = [\mathbf{V}_1 \ \mathbf{V}_2] + g\chi^{(V)}$, $\mathbf{U}_1, \mathbf{U}_2, \mathbf{V}_1, \mathbf{V}_2 \in [-1, 1]^N$, and $\chi_{i,j}^{(U)}, \chi_{i,j}^{(V)} \sim$
 199 $\mathcal{N}(0, 1)$. We infer connectivity parameters $\mathbf{z} = [\mathbf{U}_1, \mathbf{U}_2, \mathbf{V}_1, \mathbf{V}_2]$ that produce stable amplification.

200 Two conditions are necessary and sufficient for RNNs to exhibit stable amplification [51]: $\text{real}(\lambda_1) <$
 201 1 and $\lambda_1^s > 1$, where λ_1 is the eigenvalue of W with greatest real part and λ^s is the maximum
 202 eigenvalue of $W^s = \frac{W+W^\top}{2}$. RNNs with $\text{real}(\lambda_1) = 0.5 \pm 0.5$ and $\lambda_1^s = 1.5 \pm 0.5$ will be stable with
 203 modest decay rate ($\text{real}(\lambda_1)$ close to its upper bound of 1) and exhibit modest amplification (λ_1^s
 204 close to its lower bound of 1). EPI can naturally condition on this emergent property

$$\begin{aligned}\mathcal{X} : \mathbb{E}_{\mathbf{z}, \mathbf{x}} \begin{bmatrix} \text{real}(\lambda_1) \\ \lambda_1^s \end{bmatrix} &= \begin{bmatrix} 0.5 \\ 1.5 \end{bmatrix} \\ \text{Var}_{\mathbf{z}, \mathbf{x}} \begin{bmatrix} \text{real}(\lambda_1) \\ \lambda_1^s \end{bmatrix} &= \begin{bmatrix} 0.25^2 \\ 0.25^2 \end{bmatrix}. \end{aligned}\tag{5}$$

205 Variance constraints predicate that the majority of the distribution (within two standard deviations)
 206 are within the specified ranges.

207 For comparison, we infer the parameters \mathbf{z} likely to produce stable amplification using two alter-
 208 native simulation-based inference approaches. Sequential Monte Carlo approximate bayesian
 209 computation (SMC-ABC) [26] is a rejection sampling approach that uses SMC techniques to im-
 210 prove efficiency, and sequential neural posterior estimation (SNPE) [35] approximates posteriors
 211 with deep probability distributions (see Section 5.1.1). Unlike EPI, these statistical inference tech-
 212 niques do not constrain the predictions of the inferred distribution, so they were run by conditioning
 213 on an exemplar dataset $\mathbf{x}_0 = \boldsymbol{\mu}$, following standard practice with these methods [26, 35]. To com-
 214 pare the efficiency of these different techniques, we measured the time and number of simulations
 215 necessary for the distance of the predictive mean to be less than 0.5 from $\boldsymbol{\mu} = \mathbf{x}_0$ (see Section 5.3).

216 As the number of neurons N in the RNN, and thus the dimension of the parameter space $\mathbf{z} \in$
 217 $[-1, 1]^{4N}$, is scaled, we see that EPI converges at greater speed and at greater dimension than
 218 SMC-ABC and SNPE (Figure 2A). It also becomes most efficient to use EPI in terms of simulation
 219 count at $N = 50$ (Figure 2B). It is well known that ABC techniques struggle in parameter spaces
 220 of modest dimension [55], yet we were careful to assess the scalability of SNPE, which is a more
 221 closely related methodology to EPI. Between EPI and SNPE, we closely controlled the number
 222 of parameters in deep probability distributions by dimensionality (Figure 2-figure supplement 1),
 223 and tested more aggressive SNPE hyperparameter choices when SNPE failed to converge (Figure
 224 2-figure supplement 2). In this analysis, we see that deep inference techniques EPI and SNPE are
 225 far more amenable to inference of high dimensional RNN connectivities than rejection sampling
 226 techniques like SMC-ABC, and that EPI outperforms SNPE in both wall time (elapsed real time)
 227 and simulation count.

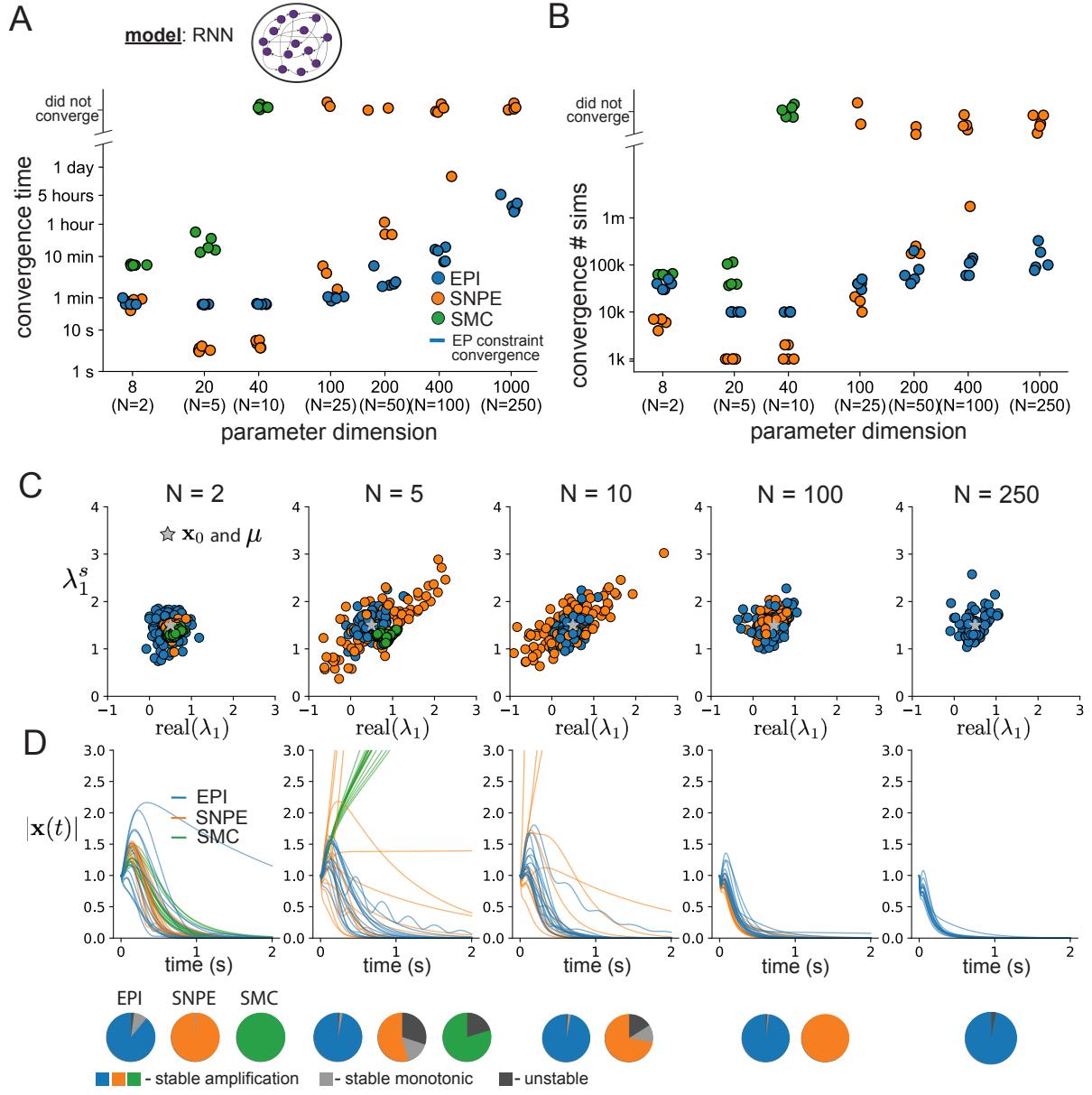


Figure 2: **A.** Wall time of EPI (blue), SNPE (orange), and SMC-ABC (green) to converge on RNN connectivities producing stable amplification. Each dot shows convergence time for an individual random seed. For reference, the mean wall time for EPI to achieve its full constraint convergence (means and variances) is shown (blue line). **B.** Simulation count of each algorithm to achieve convergence. Same conventions as A. **C.** The predictive distributions of connectivities inferred by EPI (blue), SNPE (orange), and SMC-ABC (green), with reference to $x_0 = \mu$ (gray star). **D.** Simulations of networks inferred by each method ($\tau = 100ms$). Each trace (15 per algorithm) corresponds to simulation of one z . (Below) Ratio of obtained samples producing stable amplification, stable monotonic decay, and instability.

228 No matter the number of neurons, EPI always produces connectivity distributions with mean and
229 variance of $\text{real}(\lambda_1)$ and λ_1^s according to \mathcal{X} (Figure 2C, blue). For the dimensionalities in which
230 SMC-ABC is tractable, the inferred parameters are concentrated and offset from the exemplar
231 dataset \mathbf{x}_0 (Figure 2C, green). When using SNPE, the predictions of the inferred parameters are
232 highly concentrated at some RNN sizes and widely varied in others (Figure 2C, orange). We see
233 these properties reflected in simulations from the inferred distributions: EPI produces a consistent
234 variety of stable, amplified activity norms $|\mathbf{x}(t)|$, SMC-ABC produces a limited variety of responses,
235 and the changing variety of responses from SNPE emphasizes the control of EPI on parameter pre-
236 diction (Figure 2D). Even for moderate neuron counts, the predictions of the inferred distribution
237 of SNPE are highly dependent on N and g , while EPI maintains the emergent property across
238 choices of RNN (see Section 5.3.5).

239 To understand these differences, note that EPI outperforms SNPE in high dimensions by using
240 gradient information (from $\nabla_{\mathbf{z}}[\text{real}(\lambda_1), \lambda_1^s]^{\top}$). This choice agrees with recent speculation that such
241 gradient information could improve the efficiency of simulation-based inference techniques [56],
242 as well as reflecting the classic tradeoff between gradient-based and sampling-based estimators
243 (scaling and speed versus generality). Since gradients of the emergent property are necessary
244 in EPI optimization, gradient tractability is a key criteria when determining the suitability of a
245 simulation-based inference technique. If the emergent property gradient is efficiently calculated,
246 EPI is a clear choice for inferring high dimensional parameter distributions. In the next two sections,
247 we use EPI for novel scientific insight by examining the structure of inferred distributions.

248 **3.4 EPI reveals how recurrence with multiple inhibitory subtypes governs ex-
249 citatory variability in a V1 model**

250 Dynamical models of excitatory (E) and inhibitory (I) populations with supralinear input-output
251 function have succeeded in explaining a host of experimentally documented phenomena in primary
252 visual cortex (V1). In a regime characterized by inhibitory stabilization of strong recurrent excita-
253 tion, these models give rise to paradoxical responses [12], selective amplification [43, 50], surround
254 suppression [57] and normalization [58]. Recent theoretical work [59] shows that stabilized E-I
255 models reproduce the effect of variability suppression [60]. Furthermore, experimental evidence
256 shows that inhibition is composed of distinct elements – parvalbumin (P), somatostatin (S), VIP
257 (V) – composing 80% of GABAergic interneurons in V1 [61–63], and that these inhibitory cell types
258 follow specific connectivity patterns (Figure 3A) [64]. Here, we use EPI on a model of V1 with

259 biologically realistic connectivity to show how the structure of input across neuron types affects
 260 the variability of the excitatory population – the population largely responsible for projecting to
 261 other brain areas [65].

262 We considered response variability of a nonlinear dynamical V1 circuit model (Figure 3A) with a
 263 state comprised of each neuron-type population’s rate $\mathbf{x} = [x_E, x_P, x_S, x_V]^\top$. Each population re-
 264 ceives recurrent input $W\mathbf{x}$, where W is the effective connectivity matrix (see Section 5.4) and an ex-
 265 ternal input with mean \mathbf{h} , which determines population rate via supralinear nonlinearity $\phi(\cdot) = [\cdot]_+^2$.
 266 The external input has an additive noisy component ϵ with variance $\sigma^2 = [\sigma_E^2, \sigma_P^2, \sigma_S^2, \sigma_V^2]$. This
 267 noise has a slower dynamical timescale $\tau_{\text{noise}} > \tau$ than the population rate, allowing fluctuations
 268 around a stimulus-dependent steady-state (Figure 3B). This model is the stochastic stabilized
 269 supralinear network (SSSN) [59]

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + \phi(W\mathbf{x} + \mathbf{h} + \epsilon), \quad (6)$$

270 generalized to have multiple inhibitory neuron types. It introduces stochasticity to four neuron-
 271 type models of V1 [46]. Stochasticity and inhibitory multiplicity introduce substantial complexity
 272 to the mathematical treatment of this problem (see Section 5.4.5) motivating the analysis of this
 273 model with EPI. Here, we consider fixed weights W and input \mathbf{h} [47], and study the effect of input
 274 variability $\mathbf{z} = [\sigma_E, \sigma_P, \sigma_S, \sigma_V]^\top$ on excitatory variability.

275 We quantify levels of E-population variability by studying two emergent properties

$$\begin{aligned} \mathcal{X}(5\text{Hz}) : \mathbb{E}_{\mathbf{z}, \mathbf{x}} [s_E(\mathbf{x}; \mathbf{z})] &= 5\text{Hz} & \mathcal{X}(10\text{Hz}) : \mathbb{E}_{\mathbf{z}, \mathbf{x}} [s_E(\mathbf{x}; \mathbf{z})] &= 10\text{Hz} \\ \text{Var}_{\mathbf{z}, \mathbf{x}} [s_E(\mathbf{x}; \mathbf{z})] &= 1\text{Hz}^2 & \text{Var}_{\mathbf{z}, \mathbf{x}} [s_E(\mathbf{x}; \mathbf{z})] &= 1\text{Hz}^2, \end{aligned} \quad (7)$$

276 where $s_E(\mathbf{x}; \mathbf{z})$ is the standard deviation of the stochastic E-population response about its steady
 277 state (Figure 3C). In the following analyses, we select 1Hz^2 variance such that the two emergent
 278 properties do not overlap in $s_E(\mathbf{z}; \mathbf{x})$.

279 First, we ran EPI to obtain parameter distribution $q_\theta(\mathbf{z} | \mathcal{X}(5\text{Hz}))$ producing E-population vari-
 280 ability around 5Hz (Figure 3D). From the marginal distribution of σ_E and σ_P (Figure 3D, top-left),
 281 we can see that $s_E(\mathbf{x}; \mathbf{z})$ is sensitive to various combinations of σ_E and σ_P . Alternatively, both
 282 σ_S and σ_V are degenerate with respect to $s_E(\mathbf{x}; \mathbf{z})$ evidenced by the unexpectedly high variability
 283 in those dimensions (Figure 3D, bottom-right). Together, these observations imply a curved path
 284 with respect to $s_E(\mathbf{x}; \mathbf{z})$ of 5Hz, which is indicated by the modes along σ_P (Figure 3E).

285 Figure 3E suggests a quadratic relationship in E-population fluctuations and the standard deviation
 286 of E- and P-population input; as the square of either σ_E or σ_P increases, the other compensates

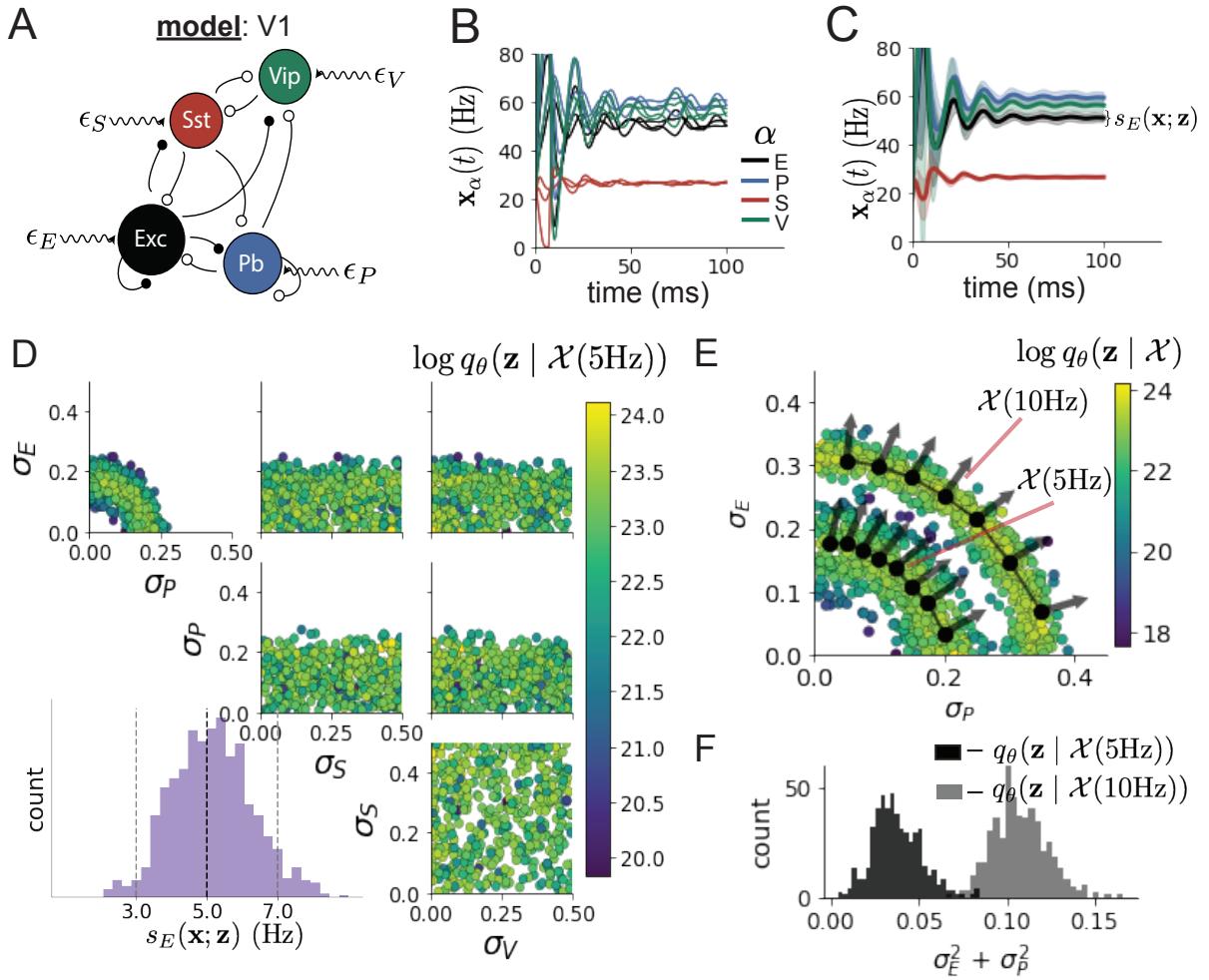


Figure 3: Emergent property inference in the stochastic stabilized supralinear network (SSSN)

A. Four-population model of primary visual cortex with excitatory (black), parvalbumin (blue), somatostatin (red), and VIP (green) neurons (excitatory and inhibitory projections filled and unfilled, respectively). Some neuron-types largely do not form synaptic projections to others ($|W_{\alpha_1, \alpha_2}| < 0.025$). Each neural population receives a baseline input \mathbf{h}_b , and the E- and P- populations also receive a contrast-dependent input \mathbf{h}_c . Additionally, each neural population receives a slow noisy input ϵ . **B.** Transient network responses of the SSSN model. Traces are independent trials with varying initialization $\mathbf{x}(0)$ and noise ϵ . **C.** Mean (solid line) and standard deviation $s_E(\mathbf{x}; \mathbf{z})$ (shading) across 100 trials. **D.** EPI distribution of noise parameters \mathbf{z} conditioned on E-population variability. The EPI predictive distribution of $s_E(\mathbf{x}; \mathbf{z})$ is show on the bottom-left. **E.** (Top) Enlarged visualization of the σ_E - σ_P marginal distribution of EPI $q_\theta(\mathbf{z} | \mathcal{X}(5\text{Hz}))$ and $q_\theta(\mathbf{z} | \mathcal{X}(10\text{Hz}))$. Each black dot shows the mode at each σ_P . The arrows show the most sensitive dimensions of the Hessian evaluated at these modes. **F.** The predictive distributions of $\sigma_E^2 + \sigma_P^2$ of each inferred distribution $q_\theta(\mathbf{z} | \mathcal{X}(5\text{Hz}))$ and $q_\theta(\mathbf{z} | \mathcal{X}(10\text{Hz}))$.

287 by decreasing to preserve the level of $s_E(\mathbf{x}; \mathbf{z})$. This quadratic relationship is preserved at greater
288 level of E-population variability $\mathcal{X}(10\text{Hz})$ (Figure 3E and Figure 3-figure supplement 1). Indeed,
289 the sum of squares of σ_E and σ_P is larger in $q_{\theta}(\mathbf{z} \mid \mathcal{X}(10\text{Hz}))$ than $q_{\theta}(\mathbf{z} \mid \mathcal{X}(5\text{Hz}))$ (Fig 3F,
290 $p < 1 \times 10^{-10}$), while the sum of squares of σ_S and σ_V are not significantly different in the two
291 EPI distributions (Figure 3-figure supplement 3, $p = .40$), in which parameters were bounded
292 from 0 to 0.5. The strong interaction between E- and P-population input variability on excitatory
293 variability is intriguing, since this circuit exhibits a paradoxical effect in the P-population (and
294 no other inhibitory types) (Figure 3-figure supplement 4), meaning that the E-population is P-
295 stabilized. Future research may uncover a link between the population of network stabilization and
296 compensatory interactions governing excitatory variability.

297 EPI revealed the quadratic dependence of excitatory variability on input variability to the E- and
298 P-populations, as well as its independence to input from the other two inhibitory populations.
299 In a simplified model ($\tau = \tau_{\text{noise}}$), it can be shown that surfaces of equal variance are ellipsoids
300 as a function of σ (see Section 5.4.5). Nevertheless, the sensitive and degenerate parameters are
301 intractable to predict mathematically, since the covariance matrix depends on the steady-state
302 solution of the network [59, 66], and terms in the covariance expression increase quadratically with
303 each additional neuron-type population (see also Section 5.4.5). By pointing out this mathematical
304 complexity, we emphasize the value of EPI for gaining understanding about theoretical models
305 when mathematical analysis becomes onerous or impractical.

306 3.5 EPI identifies two regimes of rapid task switching

307 It has been shown that rats can learn to switch from one behavioral task to the next on randomly
308 interleaved trials [67], and an important question is what neural mechanisms produce this compu-
309 tation. In this experimental setup, rats were given an explicit task cue on each trial, either Pro
310 or Anti. After a delay period, rats were shown a stimulus, and made a context (task) dependent
311 response (Figure 4A). In the Pro task, rats were required to orient towards the stimulus, while in
312 the Anti task, rats were required to orient away from the stimulus. Pharmacological inactivation
313 of the SC impaired rat performance, and time-specific optogenetic inactivation revealed a crucial
314 role for the SC on the cognitively demanding Anti trials [48]. These results motivated a nonlinear
315 dynamical model of the SC containing four functionally-defined neuron-type populations. In Duan
316 et al. 2021, a computationally intensive procedure was used to obtain a set of 373 connectivity
317 parameters that qualitatively reproduced these optogenetic inactivation results. To build upon

318 the insights of this previous work, we use the probabilistic tools afforded by EPI to identify and
 319 characterize two linked, yet distinct regimes of rapid task switching connectivity.

320 In this SC model, there are Pro- and Anti-populations in each hemisphere (left (L) and right (R))
 321 with activity variables $\mathbf{x} = [x_{LP}, x_{LA}, x_{RP}, x_{RA}]^\top$ [48]. The connectivity of these populations is
 322 parameterized by self sW , vertical vW , diagonal dW and horizontal hW connections (Figure 4B).
 323 The input \mathbf{h} is comprised of a positive cue-dependent signal to the Pro or Anti populations, a
 324 positive stimulus-dependent input to either the Left or Right populations, and a choice-period
 325 input to the entire network (see Section 5.5.1). Model responses are bounded from 0 to 1 as a
 326 function ϕ of an internal variable \mathbf{u}

$$\begin{aligned} \tau \frac{d\mathbf{u}}{dt} &= -\mathbf{u} + W\mathbf{x} + \mathbf{h} + d\mathbf{B} \\ \mathbf{x} &= \phi(\mathbf{u}). \end{aligned} \tag{8}$$

327 The model responds to the side with greater Pro neuron activation; e.g. the response is left if
 328 $x_{LP} > x_{RP}$ at the end of the trial. Here, we use EPI to determine the network connectivity
 329 $\mathbf{z} = [sW, vW, dW, hW]^\top$ that produces rapid task switching.

330 Rapid task switching is formalized mathematically as an emergent property with two statistics:
 331 accuracy in the Pro task $p_P(\mathbf{x}; \mathbf{z})$ and Anti task $p_A(\mathbf{x}; \mathbf{z})$. We stipulate that accuracy be on average
 332 .75 in each task with variance .075²

$$\begin{aligned} \mathcal{X} : \mathbb{E}_{\mathbf{z}} \begin{bmatrix} p_P(\mathbf{x}; \mathbf{z}) \\ p_A(\mathbf{x}; \mathbf{z}) \end{bmatrix} &= \begin{bmatrix} .75 \\ .75 \end{bmatrix} \\ \text{Var}_{\mathbf{z}} \begin{bmatrix} p_P(\mathbf{x}; \mathbf{z}) \\ p_A(\mathbf{x}; \mathbf{z}) \end{bmatrix} &= \begin{bmatrix} .075^2 \\ .075^2 \end{bmatrix}. \end{aligned} \tag{9}$$

333 75% accuracy is a realistic level of performance in each task, and with the chosen variance, inferred
 334 models will not exhibit fully random responses (50%), nor perfect performance (100%).

335 The EPI inferred distribution (Figure 4C) produces Pro and Anti task accuracies (Figure 4C,
 336 bottom-left) consistent with rapid task switching (Equation 9). This parameter distribution has
 337 rich structure that is not captured well by simple linear correlations (Figure 4-figure supplement
 338 1). Specifically, the shape of the EPI distribution is sharply bent, matching ground truth structure
 339 indicated by brute-force sampling (Figure 4-figure supplement 5). This is most saliently observed in
 340 the marginal distribution of $sW-hW$ (Figure 4C top-right), where anticorrelation between sW and
 341 hW switches to correlation with decreasing sW . By identifying the modes of the EPI distribution
 342 $\mathbf{z}^*(sW)$ at different values of sW (Figure 4C red/purple dots), we can quantify this change in

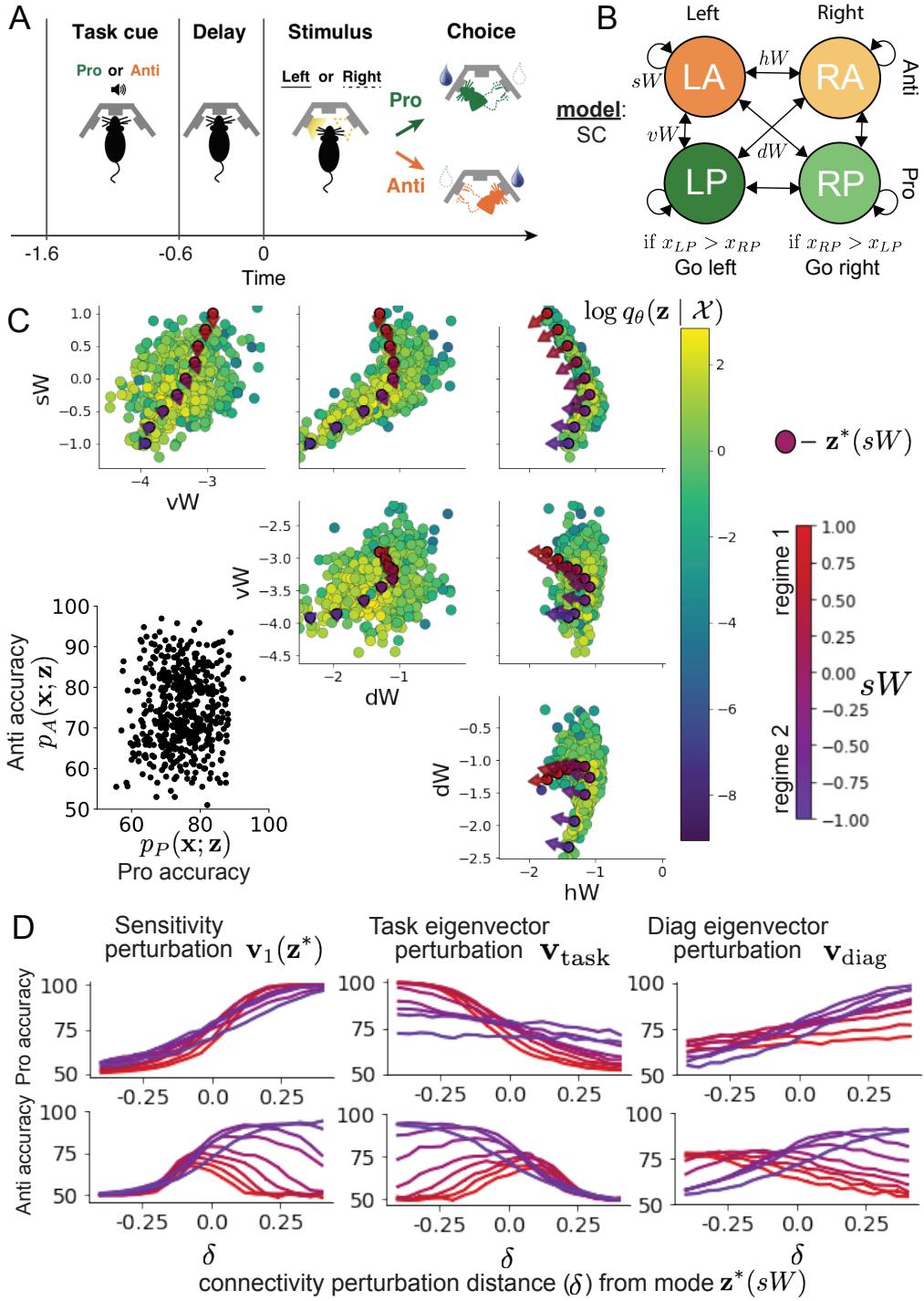


Figure 4: **A.** Rapid task switching behavioral paradigm (see text). **B.** Model of superior colliculus (SC). Neurons: LP - Left Pro, RP - Right Pro, LA - Left Anti, RA - Right Anti. Parameters: sW - self, hW - horizontal, vW - vertical, dW - diagonal weights. **C.** The EPI inferred distribution of rapid task switching networks. Red/purple parameters indicate modes $\mathbf{z}^*(sW)$ colored by sW . Sensitivity vectors $\mathbf{v}_1(\mathbf{z}^*)$ are shown by arrows. (Bottom-left) EPI predictive distribution of task accuracies. **D.** Mean and standard error ($N_{\text{test}} = 25$, bars not visible) of accuracy in Pro (top) and Anti (bottom) tasks after perturbing connectivity away from mode along $\mathbf{v}_1(\mathbf{z}^*)$ (left), \mathbf{v}_{task} (middle), and \mathbf{v}_{diag} (right).

343 distributional structure with the sensitivity dimension $\mathbf{v}_1(\mathbf{z})$ (Figure 4C red/purple arrows). Note
 344 that the directionality of these sensitivity dimensions at $\mathbf{z}^*(sW)$ changes distinctly with sW , and are
 345 perpendicular to the robust dimensions of the EPI distribution that preserve rapid task switching.
 346 These two directionalities of sensitivity motivate the distinction of connectivity into two regimes,
 347 which produce different types of responses in the Pro and Anti tasks (Figure 4-figure supplement
 348 2).
 349 When perturbing connectivity along the sensitivity dimension away from the modes

$$\mathbf{z} = \mathbf{z}^*(sW) + \delta\mathbf{v}_1(\mathbf{z}^*(sW)), \quad (10)$$

350 Pro accuracy monotonically increases in both regimes (Figure 4D, top-left). However, there is a
 351 stark difference between regimes in Anti accuracy. Anti accuracy falls in either direction of \mathbf{v}_1 in
 352 regime 1, yet monotonically increases along with Pro accuracy in regime 2 (Figure 4D, bottom-
 353 left). The sharp change in local structure of the EPI distribution is therefore explained by distinct
 354 sensitivities: Anti accuracy diminishes in only one or both directions of the sensitivity perturbation.
 355 To understand the mechanisms differentiating the two regimes, we can make connectivity pertur-
 356 bations along dimensions that only modify a single eigenvalue of the connectivity matrix. These
 357 eigenvalues λ_{all} , λ_{side} , λ_{task} , and λ_{diag} correspond to connectivity eigenmodes with intuitive roles in
 358 processing in this task (Figure 4-figure supplement 3A). For example, greater λ_{task} will strengthen
 359 internal representations of task, while greater λ_{diag} will amplify dominance of Pro and Anti pairs
 360 in opposite hemispheres (Section 5.5.7). Unlike the sensitivity dimension, the dimensions \mathbf{v}_a that
 361 perturb isolated connectivity eigenvalues λ_a for $a \in \{\text{all}, \text{side}, \text{task}, \text{diag}\}$ are independent of $\mathbf{z}^*(sW)$
 362 (see Section 5.5.7), e.g.

$$\mathbf{z} = \mathbf{z}^*(sW) + \delta\mathbf{v}_{\text{task}}. \quad (11)$$

363 Connectivity perturbation analyses reveal that decreasing λ_{task} has a very similar effect on Anti
 364 accuracy as perturbations along the sensitivity dimension (Figure 4D, middle). The similar effects
 365 of perturbations along the sensitivity dimension $\mathbf{v}_1(\mathbf{z}^*)$ and reduction of task eigenvalue (via per-
 366 turbations along $-\mathbf{v}_{\text{task}}$) suggest that there is a carefully tuned strength of task representation in
 367 connectivity regime 1, which if disturbed results in random Anti trial responses. Finally, we recog-
 368 nize that increasing λ_{diag} has opposite effects on Anti accuracy in each regime (Figure 4D, right).
 369 In the next section, we build on these mechanistic characterizations of each regime by examining
 370 their resilience to optogenetic inactivation.

371 **3.6 EPI inferred SC connectivities reproduce results from optogenetic inacti-**
372 **vation experiments**

373 During the delay period of this task, the circuit must prepare to execute the correct task according
374 to the presented cue. The circuit must then maintain a representation of task throughout the delay
375 period, which is important for correct execution of the Anti task. Duan et al. found that bilateral
376 optogenetic inactivation of SC during the delay period consistently decreased performance in the
377 Anti task, but had no effect on the Pro task (Figure 5A) [48]. The distribution of connectivities
378 inferred by EPI exhibited this same effect in simulation at high optogenetic strengths γ , which
379 reduce the network activities $\mathbf{x}(t)$ by a factor $1 - \gamma$ (Figure 5B) (see Section 5.5.8).

380 To examine how connectivity affects response to delay period inactivation, we grouped connectivi-
381 ties of the EPI distribution along the continuum linking regimes 1 and 2 of Section 3.5. $Z(sW)$ is
382 the set of EPI samples for which the closest mode was $\mathbf{z}^*(sW)$ (see Section 5.5.4). In the following
383 analyses, we examine how error, and the influence of connectivity eigenvalue on Anti error change
384 along this continuum of connectivities. Obtaining the parameter samples for these analysis with
385 the learned EPI distribution was more than 20,000 times faster than a brute force approach (see
386 Section 5.5.5).

387 The mean increase in Anti error of the EPI distribution is closest to the experimentally measured
388 value of 7% at $\gamma = 0.675$ (Figure 5B, black dot). At this level of optogenetic strength, regime 1
389 exhibits an increase in Anti error with delay period silencing (Figure 5C, left), while regime 2 does
390 not. In regime 1, greater λ_{task} and λ_{diag} decrease Anti error (Figure 5C, right). In other words,
391 stronger task representations and diagonal amplification make the SC model more resilient to delay
392 period silencing in the Anti task. This complements the finding from Duan et al. 2021 [48] that
393 λ_{task} and λ_{diag} improve Anti accuracy.

394 At roughly $\gamma = 0.85$ (Figure 5B, gray dot), the Anti error saturates, while Pro error remains
395 at zero. Following delay period inactivation at this optogenetic strength, there are strong sim-
396 ilarities in the responses of Pro and Anti trials during the choice period (Figure 5D, left). We
397 interpreted these similarities to suggest that delay period inactivation at this saturated level flips
398 the internal representation of task (from Anti to Pro) in the circuit model. A flipped task repre-
399 sentation would explain why the Anti error saturates at 50%: the average Anti accuracy in EPI
400 inferred connectivities is 75%, but is 25% when the internal representation is flipped during delay
401 period silencing. This hypothesis prescribes a model of Anti accuracy during delay period silencing

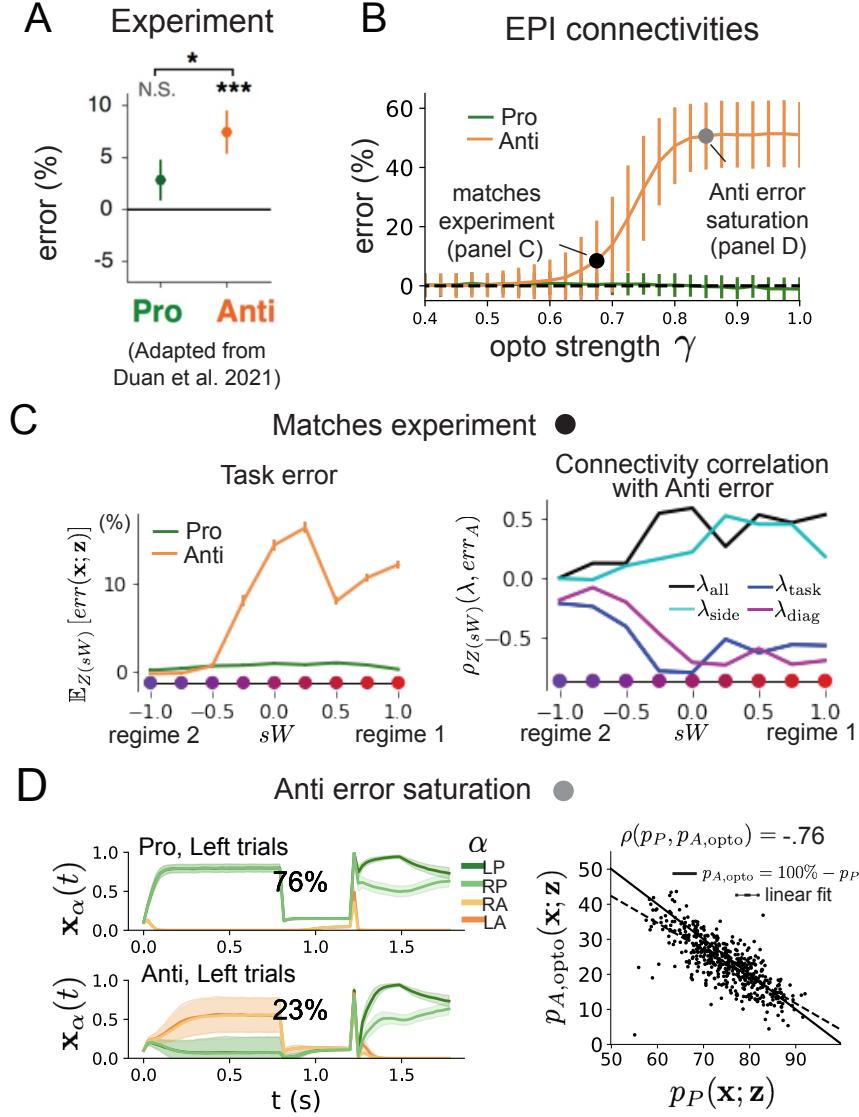


Figure 5: **A.** Mean and standard error (bars) across recording sessions of task error following delay period optogenetic inactivation in rats. **B.** Mean and standard deviation (bars) of task error induced by delay period inactivation of varying optogenetic strength γ across the EPI distribution. **C.** (Left) Mean and standard error of Pro and Anti error from regime 1 to regime 2 at $\gamma = 0.675$. (Right) Correlations of connectivity eigenvalues with Anti error from regime 1 to regime 2 at $\gamma = 0.675$. **D.** (Left) Mean and standard deviation (shading) of responses of the SC model at the mode of the EPI distribution to delay period inactivation at $\gamma = 0.85$. Accuracy in Pro (top) and Anti (bottom) task is shown as a percentage. (Right) Anti accuracy following delay period inactivation at $\gamma = 0.85$ versus accuracy in the Pro task across connectivities in the EPI distribution.

402 of $p_{A,\text{opto}} = 100\% - p_P$, which is fit closely across both regimes of the EPI inferred connectiv-
403 ities (Figure 5D, right). Similarities between Pro and Anti trial responses were not present at
404 the experiment-matching level of $\gamma = 0.675$ (Figure 5-figure supplement 2 left) and neither was
405 anticorrelation in p_P and $p_{A,\text{opto}}$ (Figure 5-figure supplement 2 right).

406 In summary, the connectivity inferred by EPI to perform rapid task switching replicated results
407 from optogenetic silencing experiments. We found that at levels of optogenetic strength matching
408 experimental levels of Anti error, only one regime actually exhibited the effect. This connectivity
409 regime is less resilient to optogenetic perturbation, and perhaps more biologically realistic. Finally,
410 we characterized the pathology in Anti error that occurs in both regimes when optogenetic strength
411 is increased to high levels, leading to a mechanistic hypothesis that is experimentally testable.
412 The probabilistic tools afforded by EPI yielded this insight: we identified two regimes and the
413 continuum of connectivities between them by taking gradients of parameter probabilities in the EPI
414 distribution, we identified sensitivity dimensions by measuring the Hessian of the EPI distribution,
415 and we obtained many parameter samples at each step along the continuum at an efficient rate.

416 4 Discussion

417 In neuroscience, machine learning has primarily been used to reveal structure in neural datasets [20].
418 Careful inference procedures are developed for these statistical models allowing precise, quantitative
419 reasoning, which clarifies the way data informs beliefs about the model parameters. However, these
420 statistical models often lack resemblance to the underlying biology, making it unclear how to go
421 from the structure revealed by these methods, to the neural mechanisms giving rise to it. In
422 contrast, theoretical neuroscience has primarily focused on careful models of neural circuits and
423 the production of emergent properties of computation, rather than measuring structure in neural
424 datasets. In this work, we improve upon parameter inference techniques in theoretical neuroscience
425 with emergent property inference, harnessing deep learning towards parameter inference in neural
426 circuit models (see Section 5.1.1).

427 Methodology for statistical inference in circuit models has evolved considerably in recent years.
428 Early work used rejection sampling techniques [24–26], but EPI and another recently developed
429 methodology [35] employ deep learning to improve efficiency and provide flexible approximations.
430 SNPE has been used for posterior inference of parameters in circuit models conditioned upon
431 exemplar data used to represent computation, but it does not infer parameter distributions that

432 only produce the computation of interest like EPI (see Section 3.3). When strict control over the
433 predictions of the inferred parameters is necessary, EPI uses a constrained optimization technique
434 [38] (see Section 5.1.4) to make inference conditioned on the emergent property possible.

435 A key difference between EPI and SNPE, is that EPI uses gradients of the emergent property
436 throughout optimization. In Section 3.3, we showed that such gradients confer beneficial scaling
437 properties, but a concern remains that emergent property gradients may be too computationally
438 intensive. Even in a case of close biophysical realism with an expensive emergent property gradient,
439 EPI was run successfully on intermediate hub frequency in a 5-neuron subcircuit model of the
440 STG (Section 3.1). However, conditioning on the pyloric rhythm [68] in a model of the pyloric
441 subnetwork model [15] proved to be prohibitive with EPI. The pyloric subnetwork requires many
442 time steps for simulation and many key emergent property statistics (e.g. burst duration and
443 phase gap) are not calculable or easily approximated with differentiable functions. In such cases,
444 SNPE, which does not require differentiability of the emergent property, has proven useful [35].
445 In summary, choice of deep inference technique should consider emergent property complexity and
446 differentiability, dimensionality of parameter space, and the importance of constraining the model
447 behavior predicted by the inferred parameter distribution.

448 In this paper, we demonstrate the value of deep inference for parameter sensitivity analyses at
449 both the local and global level. With these techniques, flexible deep probability distributions are
450 optimized to capture global structure by approximating the full distribution of suitable parame-
451 ters. Importantly, the local structure of this deep probability distribution can be quantified at
452 any parameter choice, offering instant sensitivity measurements after fitting. For example, the
453 global structure captured by EPI revealed two distinct parameter regimes, which had different
454 local structure quantified by the deep probability distribution (see Section 5.5). In comparison,
455 bayesian MCMC is considered a popular approach for capturing global parameter structure [69],
456 but there is no variational approximation (the deep probability distribution in EPI), so sensitiv-
457 ity information is not queryable and sampling remains slow after convergence. Local sensitivity
458 analyses (e.g. [27]) may be performed independently at individual parameter samples, but these
459 methods alone do not capture the full picture in nonlinear, complex distributions. In contrast,
460 deep inference yields a probability distribution that produces a wholistic assessment of parameter
461 sensitivity at the local and global level, which we used in this study to make novel insights into
462 a range of theoretical models. Together, the abilities to condition upon emergent properties, the
463 efficient inference algorithm, and the capacity for parameter sensitivity analyses make EPI a useful

464 method for addressing inverse problems in theoretical neuroscience.

465 **Acknowledgements:**

466 This work was funded by NSF Graduate Research Fellowship, DGE-1644869, McKnight Endow-
467 ment Fund, NIH NINDS 5R01NS100066, Simons Foundation 542963, NSF NeuroNex Award, DBI-
468 1707398, The Gatsby Charitable Foundation, Simons Collaboration on the Global Brain Postdoc-
469 toral Fellowship, Chinese Postdoctoral Science Foundation, and International Exchange Program
470 Fellowship. We also acknowledge the Marine Biological Laboratory Methods in Computational
471 Neuroscience Course, where this work was discussed and explored in its early stages. Helpful con-
472 versations were had with Larry Abbott, Stephen Baccus, James Fitzgerald, Gabrielle Gutierrez,
473 Francesca Mastrogiosseppe, Srdjan Ostojic, Liam Paninski, and Dhruva Raman.

474 **Data availability statement:**

475 The datasets generated during this study have been made publicly available on Zenodo at this
476 address: <https://doi.org/10.5281/zenodo.4910010>.

477 **Code availability statement:**

478 All software written for this study is available at <https://github.com/cunningham-lab/epi>.

479 **References**

- 480 [1] Nancy Kopell and G Bard Ermentrout. Coupled oscillators and the design of central pattern
481 generators. *Mathematical biosciences*, 90(1-2):87–109, 1988.
- 482 [2] Eve Marder. From biophysics to models of network function. *Annual review of neuroscience*,
483 21(1):25–45, 1998.
- 484 [3] Larry F Abbott. Theoretical neuroscience rising. *Neuron*, 60(3):489–495, 2008.
- 485 [4] Xiao-Jing Wang. Neurophysiological and computational principles of cortical rhythms in cog-
486 nition. *Physiological reviews*, 90(3):1195–1268, 2010.
- 487 [5] Timothy O’Leary, Alexander C Sutton, and Eve Marder. Computational models in the age of
488 large datasets. *Current opinion in neurobiology*, 32:87–94, 2015.
- 489 [6] Ryan N Gutenkunst, Joshua J Waterfall, Fergal P Casey, Kevin S Brown, Christopher R
490 Myers, and James P Sethna. Universally sloppy parameter sensitivities in systems biology
491 models. *PLoS Comput Biol*, 3(10):e189, 2007.

- 492 [7] Kamil Erguler and Michael PH Stumpf. Practical limits for reverse engineering of dynamical
493 systems: a statistical analysis of sensitivity and parameter inferability in systems biology
494 models. *Molecular BioSystems*, 7(5):1593–1602, 2011.
- 495 [8] Brian K Mannakee, Aaron P Ragsdale, Mark K Transtrum, and Ryan N Gutenkunst. Sloppiness
496 and the geometry of parameter space. In *Uncertainty in Biology*, pages 271–299. Springer,
497 2016.
- 498 [9] John J Hopfield. Neural networks and physical systems with emergent collective computational
499 abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.
- 500 [10] Haim Sompolinsky, Andrea Crisanti, and Hans-Jurgen Sommers. Chaos in random neural
501 networks. *Physical review letters*, 61(3):259, 1988.
- 502 [11] Andrey V Olypher and Ronald L Calabrese. Using constraints on neuronal activity to reveal
503 compensatory changes in neuronal parameters. *Journal of Neurophysiology*, 98(6):3749–3758,
504 2007.
- 505 [12] Misha V Tsodyks, William E Skaggs, Terrence J Sejnowski, and Bruce L McNaughton. Para-
506 doxical effects of external modulation of inhibitory interneurons. *Journal of neuroscience*,
507 17(11):4382–4388, 1997.
- 508 [13] Kong-Fatt Wong and Xiao-Jing Wang. A recurrent network mechanism of time integration in
509 perceptual decisions. *Journal of Neuroscience*, 26(4):1314–1328, 2006.
- 510 [14] WR Foster, LH Ungar, and JS Schwaber. Significance of conductances in hodgkin-huxley
511 models. *Journal of neurophysiology*, 70(6):2502–2518, 1993.
- 512 [15] Astrid A Prinz, Dirk Bucher, and Eve Marder. Similar network activity from disparate circuit
513 parameters. *Nature neuroscience*, 7(12):1345–1352, 2004.
- 514 [16] Pablo Achard and Erik De Schutter. Complex parameter landscape for a complex neuron
515 model. *PLoS computational biology*, 2(7):e94, 2006.
- 516 [17] Dmitry Fisher, Itsaso Olasagasti, David W Tank, Emre RF Aksay, and Mark S Goldman.
517 A modeling framework for deriving the structural and functional architecture of a short-term
518 memory microcircuit. *Neuron*, 79(5):987–1000, 2013.

- 519 [18] Timothy O’Leary, Alex H Williams, Alessio Franci, and Eve Marder. Cell types, network
520 homeostasis, and pathological compensation from a biologically plausible ion channel expres-
521 sion model. *Neuron*, 82(4):809–821, 2014.
- 522 [19] Leandro M Alonso and Eve Marder. Visualization of currents in neural models with similar
523 behavior and different conductance densities. *Elife*, 8:e42722, 2019.
- 524 [20] Liam Paninski and John P Cunningham. Neural data science: accelerating the experiment-
525 analysis-theory cycle in large-scale neuroscience. *Current opinion in neurobiology*, 50:232–241,
526 2018.
- 527 [21] Christopher M Niell and Michael P Stryker. Modulation of visual responses by behavioral state
528 in mouse visual cortex. *Neuron*, 65(4):472–479, 2010.
- 529 [22] Aman B Saleem, Asli Ayaz, Kathryn J Jeffery, Kenneth D Harris, and Matteo Carandini.
530 Integration of visual motion and locomotion in mouse visual cortex. *Nature neuroscience*,
531 16(12):1864–1869, 2013.
- 532 [23] Simon Musall, Matthew T Kaufman, Ashley L Juavinett, Steven Gluf, and Anne K Church-
533 land. Single-trial neural dynamics are dominated by richly varied movements. *Nature neuro-
534 science*, 22(10):1677–1686, 2019.
- 535 [24] Mark A Beaumont, Wenyang Zhang, and David J Balding. Approximate bayesian computation
536 in population genetics. *Genetics*, 162(4):2025–2035, 2002.
- 537 [25] Paul Marjoram, John Molitor, Vincent Plagnol, and Simon Tavaré. Markov chain monte carlo
538 without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26):15324–15328,
539 2003.
- 540 [26] Scott A Sisson, Yanan Fan, and Mark M Tanaka. Sequential monte carlo without likelihoods.
541 *Proceedings of the National Academy of Sciences*, 104(6):1760–1765, 2007.
- 542 [27] Andreas Raue, Clemens Kreutz, Thomas Maiwald, Julie Bachmann, Marcel Schilling, Ursula
543 Klingmüller, and Jens Timmer. Structural and practical identifiability analysis of partially
544 observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–
545 1929, 2009.

- 546 [28] Johan Karlsson, Milena Anguelova, and Mats Jirstrand. An efficient method for structural
547 identifiability analysis of large dynamic systems. *IFAC Proceedings Volumes*, 45(16):941–946,
548 2012.
- 549 [29] Keegan E Hines, Thomas R Middendorf, and Richard W Aldrich. Determination of parameter
550 identifiability in nonlinear biophysical models: A bayesian approach. *Journal of General*
551 *Physiology*, 143(3):401–416, 2014.
- 552 [30] Dhruva V Raman, James Anderson, and Antonis Papachristodoulou. Delineating parameter
553 unidentifiabilities in complex models. *Physical Review E*, 95(3):032314, 2017.
- 554 [31] Gamaleldin F Elsayed and John P Cunningham. Structure in neural population recordings:
555 an expected byproduct of simpler phenomena? *Nature neuroscience*, 20(9):1310, 2017.
- 556 [32] Cristina Savin and Gašper Tkačik. Maximum entropy models as a tool for building precise
557 neural controls. *Current opinion in neurobiology*, 46:120–126, 2017.
- 558 [33] Wiktor Mlynarski, Michal Hledík, Thomas R Sokolowski, and Gašper Tkačik. Statistical
559 analysis and optimality of neural systems. *bioRxiv*, page 848374, 2020.
- 560 [34] Dustin Tran, Rajesh Ranganath, and David Blei. Hierarchical implicit models and likelihood-
561 free variational inference. In *Advances in Neural Information Processing Systems*, pages 5523–
562 5533, 2017.
- 563 [35] Pedro J Gonçalves, Jan-Matthis Lueckmann, Michael Deistler, Marcel Nonnenmacher, Kaan
564 Öcal, Giacomo Bassetto, Chaitanya Chintaluri, William F Podlaski, Sara A Haddad, Tim P
565 Vogels, et al. Training deep neural density estimators to identify mechanistic models of neural
566 dynamics. *bioRxiv*, page 838383, 2019.
- 567 [36] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows.
568 *International Conference on Machine Learning*, 2015.
- 569 [37] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji
570 Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *arXiv preprint*
571 *arXiv:1912.02762*, 2019.
- 572 [38] Gabriel Loaiza-Ganem, Yuanjun Gao, and John P Cunningham. Maximum entropy flow
573 networks. *International Conference on Learning Representations*, 2017.

- 574 [39] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp.
575 *Proceedings of the 5th International Conference on Learning Representations*, 2017.
- 576 [40] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolu-
577 tions. In *Advances in neural information processing systems*, pages 10215–10224, 2018.
- 578 [41] Gabrielle J Gutierrez, Timothy O’Leary, and Eve Marder. Multiple mechanisms switch an
579 electrically coupled, synaptically inhibited neuron between competing rhythmic oscillators.
580 *Neuron*, 77(5):845–858, 2013.
- 581 [42] Mark S Goldman, Jorge Golowasch, Eve Marder, and LF Abbott. Global structure, robustness,
582 and modulation of neuronal models. *Journal of Neuroscience*, 21(14):5229–5238, 2001.
- 583 [43] Brendan K Murphy and Kenneth D Miller. Balanced amplification: a new mechanism of
584 selective amplification of neural activity patterns. *Neuron*, 61(4):635–648, 2009.
- 585 [44] Guillaume Hennequin, Tim P Vogels, and Wulfram Gerstner. Optimal control of transient dy-
586 namics in balanced networks supports generation of complex movements. *Neuron*, 82(6):1394–
587 1406, 2014.
- 588 [45] Giulio Bondanelli, Thomas Deneux, Brice Bathellier, and Srdjan Ostojic. Population coding
589 and network dynamics during off responses in auditory cortex. *BioRxiv*, page 810655, 2019.
- 590 [46] Ashok Litwin-Kumar, Robert Rosenbaum, and Brent Doiron. Inhibitory stabilization and vi-
591 sual coding in cortical circuits with multiple interneuron subtypes. *Journal of neurophysiology*,
592 115(3):1399–1409, 2016.
- 593 [47] Agostina Palmigiano, Francesco Fumarola, Daniel P Mossing, Nataliya Kraynyukova, Hillel
594 Adesnik, and Kenneth Miller. Structure and variability of optogenetic responses identify the
595 operating regime of cortex. *bioRxiv*, 2020.
- 596 [48] Chunyu A Duan, Marino Pagan, Alex T Piet, Charles D Kopec, Athena Akrami, Alexander J
597 Riordan, Jeffrey C Erlich, and Carlos D Brody. Collicular circuits for flexible sensorimotor
598 routing. *Nature Neuroscience*, pages 1–11, 2021.
- 599 [49] Eve Marder and Vatsala Thirumalai. Cellular, synaptic and network effects of neuromodula-
600 tion. *Neural Networks*, 15(4-6):479–493, 2002.
- 601 [50] Mark S Goldman. Memory without feedback in a neural network. *Neuron*, 61(4):621–634,
602 2009.

- 603 [51] Giulio Bondanelli and Srdjan Ostojic. Coding with transient trajectories in recurrent neural
604 networks. *PLoS computational biology*, 16(2):e1007655, 2020.
- 605 [52] David Sussillo. Neural circuits as computational dynamical systems. *Current opinion in*
606 *neurobiology*, 25:156–163, 2014.
- 607 [53] Omri Barak. Recurrent neural networks as versatile tools of neuroscience research. *Current*
608 *opinion in neurobiology*, 46:1–6, 2017.
- 609 [54] Abigail A Russo, Sean R Bittner, Sean M Perkins, Jeffrey S Seely, Brian M London, Antonio H
610 Lara, Andrew Miri, Najja J Marshall, Adam Kohn, Thomas M Jessell, et al. Motor cortex
611 embeds muscle-like commands in an untangled population response. *Neuron*, 97(4):953–966,
612 2018.
- 613 [55] Scott A Sisson, Yanan Fan, and Mark Beaumont. *Handbook of approximate Bayesian compu-*
614 *tation*. CRC Press, 2018.
- 615 [56] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference.
616 *Proceedings of the National Academy of Sciences*, 2020.
- 617 [57] Hirofumi Ozeki, Ian M Finn, Evan S Schaffer, Kenneth D Miller, and David Ferster. Inhibitory
618 stabilization of the cortical network underlies visual surround suppression. *Neuron*, 62(4):578–
619 592, 2009.
- 620 [58] Daniel B Rubin, Stephen D Van Hooser, and Kenneth D Miller. The stabilized supralinear
621 network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron*,
622 85(2):402–417, 2015.
- 623 [59] Guillaume Hennequin, Yashar Ahmadian, Daniel B Rubin, Máté Lengyel, and Kenneth D
624 Miller. The dynamical regime of sensory cortex: stable dynamics around a single stimulus-
625 tuned attractor account for patterns of noise variability. *Neuron*, 98(4):846–860, 2018.
- 626 [60] Mark M. Churchland, Byron M. Yu, John P. Cunningham, Leo P. Sugrue, Marlene R. Cohen,
627 Greg S. Corrado, William T. Newsome, Andrew M. Clark, Paymon Hosseini, Benjamin B.
628 Scott, David C. Bradley, Matthew A. Smith, Adam Kohn, J. Anthony Movshon, Katherine
629 M. Armstrong, Tirin Moore, Steve W. Chang, Lawrence H. Snyder, Stephen G. Lisberger,
630 Nicholas J. Priebe, Ian M. Finn, David Ferster, Stephen I. Ryu, Gopal Santhanam, Maneesh
631 Sahani, and Krishna V. Shenoy. Stimulus onset quenches neural variability: a widespread
632 cortical phenomenon. *Nat. Neurosci.*, 13(3):369–378, 2010.

- 633 [61] Henry Markram, Maria Toledo-Rodriguez, Yun Wang, Anirudh Gupta, Gilad Silberberg, and
634 Caizhi Wu. Interneurons of the neocortical inhibitory system. *Nature reviews neuroscience*,
635 5(10):793, 2004.
- 636 [62] Bernardo Rudy, Gordon Fishell, SooHyun Lee, and Jens Hjerling-Leffler. Three groups of
637 interneurons account for nearly 100% of neocortical gabaergic neurons. *Developmental neuro-*
638 *biology*, 71(1):45–61, 2011.
- 639 [63] Robin Tremblay, Soohyun Lee, and Bernardo Rudy. GABAergic Interneurons in the Neocortex:
640 From Cellular Properties to Circuits. *Neuron*, 91(2):260–292, 2016.
- 641 [64] Carsten K Pfeffer, Mingshan Xue, Miao He, Z Josh Huang, and Massimo Scanziani. Inhi-
642 bition of inhibition in visual cortex: the logic of connections between molecularly distinct
643 interneurons. *Nature Neuroscience*, 16(8):1068, 2013.
- 644 [65] Daniel J Felleman and David C Van Essen. Distributed hierarchical processing in the primate
645 cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, 1(1):1–47, 1991.
- 646 [66] C Gardiner. Stochastic methods: A Handbook for the Natural and Social Sciences, 2009.
- 647 [67] Chunyu A Duan, Jeffrey C Erlich, and Carlos D Brody. Requirement of prefrontal and midbrain
648 regions for rapid executive control of behavior in the rat. *Neuron*, 86(6):1491–1503, 2015.
- 649 [68] Eve Marder and Allen I Selverston. *Dynamic biological networks: the stomatogastric nervous*
650 *system*. MIT press, 1992.
- 651 [69] Mark Girolami and Ben Calderhead. Riemann manifold langevin and hamiltonian monte
652 carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*,
653 73(2):123–214, 2011.
- 654 [70] Dimitri P Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic
655 press, 2014.
- 656 [71] Lawrence Saul and Michael Jordan. A mean field learning algorithm for unsupervised neural
657 networks. In *Learning in graphical models*, pages 541–554. Springer, 1998.
- 658 [72] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and
659 Edward Teller. Equation of state calculations by fast computing machines. *The journal of*
660 *chemical physics*, 21(6):1087–1092, 1953.

- 661 [73] W Keith Hastings. Monte carlo sampling methods using markov chains and their applications.
662 1970.
- 663 [74] Ben Calderhead and Mark Girolami. Statistical analysis of nonlinear dynamical systems using
664 differential geometric sampling methods. *Interface focus*, 1(6):821–835, 2011.
- 665 [75] Andrew Golightly and Darren J Wilkinson. Bayesian parameter inference for stochastic bio-
666 chemical network models using particle markov chain monte carlo. *Interface focus*, 1(6):807–
667 820, 2011.
- 668 [76] Oksana A Chkrebtii, David A Campbell, Ben Calderhead, Mark A Girolami, et al. Bayesian
669 solution uncertainty quantification for differential equations. *Bayesian Analysis*, 11(4):1239–
670 1267, 2016.
- 671 [77] Juliane Liepe, Paul Kirk, Sarah Filippi, Tina Toni, Chris P Barnes, and Michael PH Stumpf.
672 A framework for parameter estimation and model selection from experimental data in systems
673 biology using approximate bayesian computation. *Nature protocols*, 9(2):439–456, 2014.
- 674 [78] Sean R Bittner, Agostina Palmigiano, Kenneth D Miller, and John P Cunningham. Degener-
675 ate solution networks for theoretical neuroscience. *Computational and Systems Neuroscience
676 Meeting (COSYNE), Lisbon, Portugal*, 2019.
- 677 [79] Sean R Bittner, Alex T Piet, Chunyu A Duan, Agostina Palmigiano, Kenneth D Miller,
678 Carlos D Brody, and John P Cunningham. Examining models in theoretical neuroscience with
679 degenerate solution networks. *Bernstein Conference 2019, Berlin, Germany*, 2019.
- 680 [80] Marcel Nonnenmacher, Pedro J Goncalves, Giacomo Bassetto, Jan-Matthis Lueckmann, and
681 Jakob H Macke. Robust statistical inference for simulation-based models in neuroscience. In
682 *Bernstein Conference 2018, Berlin, Germany*, 2018.
- 683 [81] Deistler Michael, , Pedro J Goncalves, Kaan Oecal, and Jakob H Macke. Statistical inference for
684 analyzing sloppiness in neuroscience models. In *Bernstein Conference 2019, Berlin, Germany*,
685 2019.
- 686 [82] Jan-Matthis Lueckmann, Pedro J Goncalves, Giacomo Bassetto, Kaan Öcal, Marcel Nonnen-
687 macher, and Jakob H Macke. Flexible statistical inference for mechanistic models of neural
688 dynamics. In *Advances in Neural Information Processing Systems*, pages 1289–1299, 2017.

- 689 [83] George Papamakarios, David Sterratt, and Iain Murray. Sequential neural likelihood: Fast
690 likelihood-free inference with autoregressive flows. In *The 22nd International Conference on*
691 *Artificial Intelligence and Statistics*, pages 837–848. PMLR, 2019.
- 692 [84] Joeri Hermans, Volodimir Begy, and Gilles Louppe. Likelihood-free mcmc with amortized
693 approximate ratio estimators. In *International Conference on Machine Learning*, pages 4239–
694 4248. PMLR, 2020.
- 695 [85] Martin J Wainwright, Michael I Jordan, et al. Graphical models, exponential families, and
696 variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.
- 697 [86] Sean R Bittner and John P Cunningham. Approximating exponential family models (not
698 single distributions) with a two-network architecture. *arXiv preprint arXiv:1903.07515*, 2019.
- 699 [87] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary
700 differential equations. In *Advances in neural information processing systems*, pages 6571–6583,
701 2018.
- 702 [88] Xuechen Li, Ting-Kam Leonard Wong, Ricky TQ Chen, and David Duvenaud. Scalable
703 gradients for stochastic differential equations. *arXiv preprint arXiv:2001.01328*, 2020.
- 704 [89] Maria Pia Saccomani, Stefania Audoly, and Leontina D’Angiò. Parameter identifiability of
705 nonlinear systems: the role of initial conditions. *Automatica*, 39(4):619–632, 2003.
- 706 [90] Stefan Hengl, Clemens Kreutz, Jens Timmer, and Thomas Maiwald. Data-based identifiability
707 analysis of non-linear dynamical models. *Bioinformatics*, 23(19):2612–2618, 2007.
- 708 [91] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density
709 estimation. In *Advances in Neural Information Processing Systems*, pages 2338–2347, 2017.
- 710 [92] Durk P Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling.
711 Improved variational inference with inverse autoregressive flow. *Advances in neural information*
712 *processing systems*, 29:4743–4751, 2016.
- 713 [93] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International*
714 *Conference on Learning Representations*, 2015.
- 715 [94] Emmanuel Klinger, Dennis Rickert, and Jan Hasenauer. pyabc: distributed, likelihood-free
716 inference. *Bioinformatics*, 34(20):3591–3593, 2018.

717 [95] David S Greenberg, Marcel Nonnenmacher, and Jakob H Macke. Automatic posterior trans-
718 formation for likelihood-free inference. *International Conference on Machine Learning*, 2019.

719 [96] Daniel P Mossing, Julia Veit, Agostina Palmigiano, Kenneth D. Miller, and Hillel Adesnik.
720 Antagonistic inhibitory subnetworks control cooperation and competition across cortical space.
721 *bioRxiv*, 2021.

722 **5 Methods**

723 **5.1 Emergent property inference (EPI)**

724 Solving inverse problems is an important part of theoretical neuroscience, since we must understand
725 how neural circuit models and their parameter choices produce computations. Recently, research on
726 machine learning methodology for neuroscience has focused on finding latent structure in large-scale
727 neural datasets, while research in theoretical neuroscience generally focuses on developing precise
728 neural circuit models that can produce computations of interest. By quantifying computation
729 into an *emergent property* through statistics of the emergent activity of neural circuit models, we
730 can adapt the modern technique of deep probabilistic inference towards solving inverse problems
731 in theoretical neuroscience. Here, we introduce a novel method for statistical inference, which
732 uses deep networks to learn parameter distributions constrained to produce emergent properties of
733 computation.

734 Consider model parameterization \mathbf{z} , which is a collection of scientifically meaningful variables that
735 govern the complex simulation of data \mathbf{x} . For example (see Section 3.1), \mathbf{z} may be the electrical
736 conductance parameters of an STG subcircuit, and \mathbf{x} the evolving membrane potentials of the five
737 neurons. In terms of statistical modeling, this circuit model has an intractable likelihood $p(\mathbf{x} | \mathbf{z})$,
738 which is predicated by the stochastic differential equations that define the model. From a theoretical
739 perspective, we are less concerned about the likelihood of an exemplar dataset \mathbf{x} , but rather the
740 emergent property of intermediate hub frequency (which implies a consistent dataset \mathbf{x}).

741 In this work, emergent properties \mathcal{X} are defined through the choice of emergent property statistic
742 $f(\mathbf{x}; \mathbf{z})$ (which is a vector of one or more statistics), and its means $\boldsymbol{\mu}$, and variances $\boldsymbol{\sigma}^2$:

$$\mathcal{X} : \mathbb{E}_{\mathbf{z}, \mathbf{x}} [f(\mathbf{x}; \mathbf{z})] = \boldsymbol{\mu}, \quad \text{Var}_{\mathbf{z}, \mathbf{x}} [f(\mathbf{x}; \mathbf{z})] = \boldsymbol{\sigma}^2. \quad (12)$$

743 In general, an emergent property may be a collection of first-, second-, or higher-order moments
744 of a group of statistics, but this study focuses on the case written in Equation 12. In the STG
745 example, intermediate hub frequency is defined by mean and variance constraints on the statistic
746 of hub neuron frequency $\omega_{\text{hub}}(\mathbf{x}; \mathbf{z})$ (Equations 2 and 3). Precisely, the emergent property statistics
747 $f(\mathbf{x}; \mathbf{z})$ must have means $\boldsymbol{\mu}$ and variances $\boldsymbol{\sigma}^2$ over the EPI distribution of parameters ($\mathbf{z} \sim q_{\boldsymbol{\theta}}(\mathbf{z})$) and
748 the data produced by those parameters ($\mathbf{x} \sim p(\mathbf{x} | \mathbf{z})$), where the inferred parameter distribution
749 $q_{\boldsymbol{\theta}}(\mathbf{z})$ itself is parameterized by deep network weights and biases $\boldsymbol{\theta}$.

750 In EPI, a deep probability distribution $q_{\boldsymbol{\theta}}(\mathbf{z})$ is optimized to approximate the parameter distribution

751 producing the emergent property \mathcal{X} . In contrast to simpler classes of distributions like the gaussian
 752 or mixture of gaussians, deep probability distributions are far more flexible and capable of fitting
 753 rich structure [36, 37]. In deep probability distributions, a simple random variable $\mathbf{z}_0 \sim q_0(\mathbf{z}_0)$ (we
 754 choose an isotropic gaussian) is mapped deterministically via a sequence of deep neural network
 755 layers ($g_1, \dots g_l$) parameterized by weights and biases $\boldsymbol{\theta}$ to the support of the distribution of interest:

$$\mathbf{z} = g_{\boldsymbol{\theta}}(\mathbf{z}_0) = g_l(\dots g_1(\mathbf{z}_0)) \sim q_{\boldsymbol{\theta}}(\mathbf{z}). \quad (13)$$

756 Such deep probability distributions embed the inferred distribution in a deep network. Once op-
 757 timized, this deep network representation of a distribution has remarkably useful properties: fast
 758 sampling and probability evaluations. Importantly, fast probability evaluations confer fast gradient
 759 and Hessian calculations as well.

760 Given this choice of circuit model and emergent property \mathcal{X} , $q_{\boldsymbol{\theta}}(\mathbf{z})$ is optimized via the neural
 761 network parameters $\boldsymbol{\theta}$ to find a maximally entropic distribution $q_{\boldsymbol{\theta}}^*$ within the deep variational
 762 family $\mathcal{Q} = \{q_{\boldsymbol{\theta}}(\mathbf{z}) : \boldsymbol{\theta} \in \Theta\}$ that produces the emergent property \mathcal{X} :

$$\begin{aligned} q_{\boldsymbol{\theta}}(\mathbf{z} | \mathcal{X}) &= q_{\boldsymbol{\theta}}^*(\mathbf{z}) = \operatorname{argmax}_{q_{\boldsymbol{\theta}} \in \mathcal{Q}} H(q_{\boldsymbol{\theta}}(\mathbf{z})) \\ \text{s.t. } \mathcal{X} &: \mathbb{E}_{\mathbf{z}, \mathbf{x}} [f(\mathbf{x}; \mathbf{z})] = \boldsymbol{\mu}, \operatorname{Var}_{\mathbf{z}, \mathbf{x}} [f(\mathbf{x}; \mathbf{z})] = \boldsymbol{\sigma}^2, \end{aligned} \quad (14)$$

763 where $H(q_{\boldsymbol{\theta}}(\mathbf{z})) = \mathbb{E}_{\mathbf{z}} [-\log q_{\boldsymbol{\theta}}(\mathbf{z})]$ is entropy. By maximizing the entropy of the inferred distribution
 764 $q_{\boldsymbol{\theta}}$, we select the most random distribution in family \mathcal{Q} that satisfies the constraints of the emergent
 765 property. Since entropy is maximized in Equation 14, EPI is equivalent to bayesian variational
 766 inference (see Section 5.1.6), which is why we specify the inferred distribution of EPI as conditioned
 767 upon emergent property \mathcal{X} with the notation $q_{\boldsymbol{\theta}}(\mathbf{z} | \mathcal{X})$. To run this constrained optimization, we
 768 use an augmented lagrangian objective, which is the standard approach for constrained optimization
 769 [70], and the approach taken to fit Maximum Entropy Flow Networks (MEFNs) [38]. This procedure
 770 is detailed in Section 5.1.4 and the pseudocode in Algorithm 1.

771 In the remainder of Section 5.1, we will explain the finer details and motivation of the EPI method.
 772 First, we explain related approaches and what EPI introduces to this domain (Section 5.1.1). Sec-
 773 ond, we describe the special class of deep probability distributions used in EPI called normalizing
 774 flows (Section 5.1.2). Then, we establish the known relationship between maximum entropy dis-
 775 tributions and exponential families (Section 5.1.3). Next, we explain the constrained optimization
 776 technique used to solve Equation 14 (Section 5.1.4). Then, we demonstrate the details of this opti-
 777 mization in a toy example (Section 5.1.5). Finally, we explain how EPI is equivalent to variational
 778 inference (Section 5.1.6).

779 **5.1.1 Related approaches**

780 When bayesian inference problems lack conjugacy, scientists use approximate inference methods like
781 variational inference (VI) [71] and Markov chain Monte Carlo (MCMC) [72,73]. After optimization,
782 variational methods return a parameterized posterior distribution, which we can analyze. Also, the
783 variational approximation is often chosen such that it permits fast sampling. In contrast MCMC
784 methods only produce samples from the approximated posterior distribution. No parameterized
785 distribution is estimated, and additional samples are always generated with the same sampling
786 complexity. Inference in models defined by systems of differential has been demonstrated with
787 MCMC [69], although this approach requires tractable likelihoods. Advancements have introduced
788 sampling [74], likelihood approximation [75], and uncertainty quantification techniques [76] to make
789 MCMC approaches more efficient and expand the class of applicable models.

790 Simulation-based inference [56] is model parameter inference in the absence of a tractable likeli-
791 hood function. The most prevalent approach to simulation-based inference is approximate bayesian
792 computation (ABC) [24], in which satisfactory parameter samples are kept from random prior sam-
793 pling according to a rejection heuristic. The obtained set of parameters do not have a probabilities,
794 and further insight about the model must be gained from examination of the parameter set and
795 their generated activity. Methodological advances to ABC methods have come through the use of
796 Markov chain Monte Carlo (MCMC-ABC) [25] and sequential Monte Carlo (SMC-ABC) [26] sam-
797 pling techniques. SMC-ABC is considered state-of-the-art ABC, yet this approach still struggles to
798 scale in dimensionality [55] (cf. Figure 2). Still, this method has enjoyed much success in systems
799 biology [77]. Furthermore, once a parameter set has been obtained by SMC-ABC from a finite set
800 of particles, the SMC-ABC algorithm must be run again from scratch with a new population of
801 initialized particles to obtain additional samples.

802 For scientific model analysis, we seek a parameter distribution represented by an approximating
803 distribution as in variational inference [71]: a variational approximation that once optimized yields
804 fast analytic calculations and samples. For the reasons described above, ABC and MCMC tech-
805 niques are not suitable, since they only produce a set of parameter samples lacking probabilities
806 and have unchanging sampling rate. EPI infers parameters in circuit models using the MEFN [38]
807 algorithm with a deep variational approximation. The deep neural network of EPI (Figure 1E) de-
808 fines the parametric form (with weights and biases as variational parameters θ) of the variational
809 approximation of the inferred parameter distribution $q_\theta(\mathbf{z} | \mathbf{x})$. The EPI optimization is enabled
810 using stochastic gradient techniques in the spirit of likelihood-free variational inference [34]. The

811 analytic relationship between EPI and variational inference is explained in Section 5.1.6.

812 We note that, during our preparation and early presentation of this work [78, 79], another work
813 has arisen with broadly similar goals: bringing statistical inference to mechanistic models of neural
814 circuits [35, 80, 81]. We are encouraged by this general problem being recognized by others in the
815 community, and we emphasize that these works offer complementary neuroscientific contributions
816 (different theoretical models of focus) and use different technical methodologies (ours is built on
817 our prior work [38], theirs similarly [82]).

818 The method EPI differs from SNPE in some key ways. SNPE belongs to a “sequential” class
819 of recently developed simulation-based inference methods in which two neural networks are used
820 for posterior inference. This first neural network is a deep probability distribution (normalizing
821 flow) used to estimate the posterior $p(\mathbf{z} | \mathbf{x})$ (SNPE) or the likelihood $p(\mathbf{x} | \mathbf{z})$ (sequential neural
822 likelihood (SNL) [83]). A recent approach uses an unconstrained neural network to estimate the
823 likelihood ratio (sequential neural ratio estimation (SNRE) [84]). In SNL and SNRE, MCMC
824 sampling techniques are used to obtain samples from the approximated posterior. This contrasts
825 with EPI and SNPE, which use deep probability distributions to model parameters, which facilitates
826 immediate measurements of sample probability, gradient, or Hessian for system analysis. The
827 second neural network in this sequential class of methods is the amortizer. This unconstrained
828 deep network maps data \mathbf{x} (or statistics $f(\mathbf{x}; \mathbf{z})$ or model parameters \mathbf{z}) to the weights and biases of
829 the first neural network. These methods are optimized on a conditional density (or ratio) estimation
830 objective. The data used to optimize this objective are generated via an adaptive procedure, in
831 which training data pairs $(\mathbf{x}_i, \mathbf{z}_i)$ become sequentially closer to the true data and posterior.

832 The approximating fidelity of the deep probability distribution in sequential approaches is opti-
833 mized to generalize across the training distribution of the conditioning variable. This generalization
834 property of the sequential methods can reduce the accuracy at the singular posterior of interest.
835 Whereas in EPI, the entire expressivity of the deep probability distribution is dedicated to learning
836 a single distribution as well as possible. The well-known inverse mapping problem of exponential
837 families [85] prohibits an amortization-based approach in EPI, since EPI learns an exponential fam-
838 ily distribution parameterized by its mean (in contrast to its natural parameter, see Section 5.1.3).
839 However, we have shown that the same two-network architecture of the sequential simulation-based
840 inference methods can be used for amortized inference in intractable exponential family posteriors
841 when using their natural parameterization [86].

842 Finally, one important differentiating factor between EPI and sequential simulation-based infer-

ence methods is that EPI leverages gradients $\nabla_{\mathbf{z}} f(\mathbf{x}; \mathbf{z})$ during optimization. These gradients can
 843 improve convergence time and scalability, as we have shown on an example conditioning low-rank
 844 RNN connectivity on the property of stable amplification (see Section 3.3). With EPI, we prove out
 845 the suggestion that a deep inference technique can improve efficiency by leveraging these emergent
 846 property gradients when they are tractable. Sequential simulation-based inference techniques may
 847 be better suited for scientific problems where $\nabla_{\mathbf{z}} f(\mathbf{x}; \mathbf{z})$ is intractable or unavailable, like when
 848 there is a nondifferentiable emergent property. However, the sequential simulation-based inference
 849 techniques cannot constrain the predictions of the inferred distribution in the manner of EPI.
 850

851 Structural identifiability analysis involves the measurement of sensitivity and unidentifiabilities in
 852 scientific models. Around a single parameter choice, one can measure the Jacobian. One approach
 853 for this calculation that scales well is EAR [28]. A popular efficient approach for systems of ODEs
 854 has been neural ODE adjoint [87] and its stochastic adaptation [88]. Casting identifiability as a
 855 statistical estimation problem, the profile likelihood works via iterated optimization while holding
 856 parameters fixed [27]. An exciting recent method is capable of recovering the functional form of such
 857 unidentifiabilities away from a point by following degenerate dimensions of the fisher information
 858 matrix [30]. Global structural non-identifiabilities can be found for models with polynomial or
 859 rational dynamics equations using DAISY [89], or through mean optimal transformations [90].
 860 With EPI, we have all the benefits given by a statistical inference method plus the ability to query
 861 the first- or second-order gradient of the probability of the inferred distribution at any chosen
 862 parameter value. The second-order gradient of the log probability (the Hessian), which is directly
 863 afforded by EPI distributions, produces quantified information about parametric sensitivity of the
 864 emergent property in parameter space (see Section 3.2).

865 5.1.2 Deep probability distributions and normalizing flows

866 Deep probability distributions are comprised of multiple layers of fully connected neural networks
 867 (Equation 13). When each neural network layer is restricted to be a bijective function, the sample
 868 density can be calculated using the change of variables formula at each layer of the network. For
 869 $\mathbf{z}_i = g_i(\mathbf{z}_{i-1})$,

$$p(\mathbf{z}_i) = p(g_i^{-1}(\mathbf{z}_i)) \left| \det \frac{\partial g_i^{-1}(\mathbf{z}_i)}{\partial \mathbf{z}_i} \right| = p(\mathbf{z}_{i-1}) \left| \det \frac{\partial g_i(\mathbf{z}_{i-1})}{\partial \mathbf{z}_{i-1}} \right|^{-1}. \quad (15)$$

870 However, this computation has cubic complexity in dimensionality for fully connected layers. By
 871 restricting our layers to normalizing flows [36, 37] – bijective functions with fast log determinant

872 Jacobian computations, which confer a fast calculation of the sample log probability. Fast log
873 probability calculation confers efficient optimization of the maximum entropy objective (see Section
874 5.1.4).

875 We use the real NVP [39] normalizing flow class, because its coupling architecture confers both
876 fast sampling (forward) and fast log probability evaluation (backward). Fast probability evaluation
877 facilitates fast gradient and Hessian evaluation of log probability throughout parameter space.
878 Glow permutations were used in between coupling stages [40]. This is in contrast to autoregressive
879 architectures [91, 92], in which only one of the forward or backward passes can be efficient. In this
880 work, normalizing flows are used as flexible parameter distribution approximations $q_{\theta}(\mathbf{z})$ having
881 weights and biases θ . We specify the architecture used in each application by the number of real
882 NVP affine coupling stages, and the number of neural network layers and units per layer of the
883 conditioning functions.

884 When calculating Hessians of log probabilities in deep probability distributions, it is important to
885 consider the normalizing flow architecture. With autoregressive architectures [91, 92], fast sam-
886 pling and fast log probability evaluations are mutually exclusive. That makes these architectures
887 undesirable for EPI, where efficient sampling is important for optimization, and log probability
888 evaluation speed predicates the efficiency of gradient and Hessian calculations. With real NVP
889 coupling architectures, we get both fast sampling and fast Hessians making both optimization and
890 scientific analysis efficient.

891 5.1.3 Maximum entropy distributions and exponential families

892 The inferred distribution of EPI is a maximum entropy distribution, which have fundamental links
893 to exponential family distributions. A maximum entropy distribution of form:

$$p^*(\mathbf{z}) = \underset{p \in \mathcal{P}}{\operatorname{argmax}} H(p(\mathbf{z})) \quad (16)$$

s.t. $\mathbb{E}_{\mathbf{z} \sim p}[T(\mathbf{z})] = \boldsymbol{\mu}_{\text{opt}},$

894 where $T(\mathbf{z})$ is the sufficient statistics vector and $\boldsymbol{\mu}_{\text{opt}}$ a vector of their mean values, will have
895 probability density in the exponential family:

$$p^*(\mathbf{z}) \propto \exp(\boldsymbol{\eta}^\top T(\mathbf{z})). \quad (17)$$

896 The mappings between the mean parameterization $\boldsymbol{\mu}_{\text{opt}}$ and the natural parameterization $\boldsymbol{\eta}$ are
897 formally hard to identify except in special cases [85].

898 In this manuscript, emergent properties are defined by statistics $f(\mathbf{x}; \mathbf{z})$ having a fixed mean $\boldsymbol{\mu}$ and
 899 variance σ^2 as in Equation 12. The variance constraint is a second moment constraint on $f(\mathbf{x}; \mathbf{z})$:

$$\text{Var}_{\mathbf{z}, \mathbf{x}} [f(\mathbf{x}; \mathbf{z})] = \mathbb{E}_{\mathbf{z}, \mathbf{x}} [(f(\mathbf{x}; \mathbf{z}) - \boldsymbol{\mu})^2]. \quad (18)$$

900 As a general maximum entropy distribution (Equation 16), the sufficient statistics vector contains
 901 both first and second order moments of $f(\mathbf{x}; \mathbf{z})$

$$T(\mathbf{z}) = \begin{bmatrix} \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x}|\mathbf{z})} [f(\mathbf{x}; \mathbf{z})] \\ \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x}|\mathbf{z})} [(f(\mathbf{x}; \mathbf{z}) - \boldsymbol{\mu})^2] \end{bmatrix}, \quad (19)$$

902 which are constrained to the chosen means and variances

$$\boldsymbol{\mu}_{\text{opt}} = \begin{bmatrix} \boldsymbol{\mu} \\ \sigma^2 \end{bmatrix}. \quad (20)$$

903 Thus, $\boldsymbol{\mu}_{\text{opt}}$ is used to denote the mean parameter of the maximum entropy distribution defined by
 904 the emergent property (all constraints), while $\boldsymbol{\mu}$ is only the mean of $f(\mathbf{x}; \mathbf{z})$. The subscript “opt” of
 905 $\boldsymbol{\mu}_{\text{opt}}$ is chosen since it contains all of the constraint values to which the EPI optimization algorithm
 906 must adhere.

907 5.1.4 Augmented lagrangian optimization

908 To optimize $q_{\boldsymbol{\theta}}(\mathbf{z})$ in Equation 14, the constrained maximum entropy optimization is executed using
 909 the augmented lagrangian method. The following objective is minimized:

$$L(\boldsymbol{\theta}; \boldsymbol{\eta}_{\text{opt}}, c) = -H(q_{\boldsymbol{\theta}}) + \boldsymbol{\eta}_{\text{opt}}^\top R(\boldsymbol{\theta}) + \frac{c}{2} \|R(\boldsymbol{\theta})\|^2 \quad (21)$$

910 where there are average constraint violations

$$R(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{z} \sim q_{\boldsymbol{\theta}}(\mathbf{z})} [T(\mathbf{z}) - \boldsymbol{\mu}_{\text{opt}}], \quad (22)$$

911 $\boldsymbol{\eta}_{\text{opt}} \in \mathbb{R}^m$ are the lagrange multipliers where m is the number of total constraints

$$m = |\boldsymbol{\mu}_{\text{opt}}| = |T(\mathbf{z})| = 2|f(\mathbf{x}; \mathbf{z})|, \quad (23)$$

912 and c is the penalty coefficient. The mean parameter $\boldsymbol{\mu}_{\text{opt}}$ and sufficient statistics $T(\mathbf{z})$ are de-
 913 termined by the means $\boldsymbol{\mu}$ and variances σ^2 of the emergent property statistics $f(\mathbf{x}; \mathbf{z})$ defined in
 914 Equation 14. Specifically, $T(\mathbf{z})$ is a concatenation of the first and second moments (Equation 19)
 915 and $\boldsymbol{\mu}_{\text{opt}}$ is a concatenation of their constraints $\boldsymbol{\mu}$ and σ^2 (Equation 20). (Although, note that

916 this algorithm is written for general $T(\mathbf{z})$ and $\boldsymbol{\mu}_{\text{opt}}$ to satisfy the more general class of emergent
 917 properties.) The lagrange multipliers $\boldsymbol{\eta}_{\text{opt}}$ are closely related to the natural parameters $\boldsymbol{\eta}$ of expo-
 918 nential families (see Section 5.1.6). Weights and biases $\boldsymbol{\theta}$ of the deep probability distribution are
 919 optimized according to Equation 21 using the Adam optimizer with learning rate 10^{-3} [93].

920 The gradient with respect to entropy $H(q_{\boldsymbol{\theta}}(\mathbf{z}))$ can be expressed using the reparameterization trick
 921 as an expectation of the negative log density of parameter samples \mathbf{z} over the randomness in the
 922 parameterless initial distribution $q_0(\mathbf{z}_0)$:

$$H(q_{\boldsymbol{\theta}}(\mathbf{z})) = \int -q_{\boldsymbol{\theta}}(\mathbf{z}) \log(q_{\boldsymbol{\theta}}(\mathbf{z})) d\mathbf{z} = \mathbb{E}_{\mathbf{z} \sim q_{\boldsymbol{\theta}}} [-\log(q_{\boldsymbol{\theta}}(\mathbf{z}))] = \mathbb{E}_{\mathbf{z}_0 \sim q_0} [-\log(q_{\boldsymbol{\theta}}(g_{\boldsymbol{\theta}}(\mathbf{z}_0)))]. \quad (24)$$

923 Thus, the gradient of the entropy of the deep probability distribution can be estimated as an
 924 average of gradients with respect to the base distribution \mathbf{z}_0 :

$$\nabla_{\boldsymbol{\theta}} H(q_{\boldsymbol{\theta}}(\mathbf{z})) = \mathbb{E}_{\mathbf{z}_0 \sim q_0} [-\nabla_{\boldsymbol{\theta}} \log(q_{\boldsymbol{\theta}}(g_{\boldsymbol{\theta}}(\mathbf{z}_0)))] . \quad (25)$$

925 The gradients of the log density of the deep probability distribution are tractable through the use
 926 of normalizing flows (see Section 5.1.2).

927 The full EPI optimization algorithm is detailed in Algorithm 1. The lagrangian parameters $\boldsymbol{\eta}_{\text{opt}}$
 928 are initialized to zero and adapted following each augmented lagrangian epoch, which is a period of
 929 optimization with fixed $(\boldsymbol{\eta}_{\text{opt}}, c)$ for a given number of stochastic gradient descent (SGD) iterations.
 930 A low value of c is used initially, and conditionally increased after each epoch based on constraint
 931 error reduction. The penalty coefficient is updated based on the result of a hypothesis test regarding
 932 the reduction in constraint violation. The p-value of $\mathbb{E}[|R(\boldsymbol{\theta}_{k+1})|] > \gamma \mathbb{E}[|R(\boldsymbol{\theta}_k)|]$ is computed,
 933 and c_{k+1} is updated to βc_k with probability $1 - p$. The other update rule is $\boldsymbol{\eta}_{\text{opt},k+1} = \boldsymbol{\eta}_{\text{opt},k} +$
 934 $c_k \frac{1}{n} \sum_{i=1}^n (T(\mathbf{z}^{(i)}) - \boldsymbol{\mu}_{\text{opt}})$ given a batch size n and $\mathbf{z}^{(i)} \sim q_{\boldsymbol{\theta}}(\mathbf{z})$. Throughout the study, $\gamma = 0.25$,
 935 while β was chosen to be either 2 or 4. The batch size of EPI also varied according to application.

936 In general, c and $\boldsymbol{\eta}_{\text{opt}}$ should start at values encouraging entropic growth early in optimization.
 937 With each training epoch in which the update rule for c is invoked, the constraint satisfaction
 938 terms are increasingly weighted, which generally results in decreased entropy (e.g. see Figure 1-
 939 figure supplement 1C). This encourages the discovery of suitable regions of parameter space, and
 940 the subsequent refinement of the distribution to produce the emergent property. The momentum
 941 parameters of the Adam optimizer are reset at the end of each augmented lagrangian epoch, which
 942 proceeds for i_{max} iterations. In this work, we used a maximum number of augmented lagrangian
 943 epochs $k_{\text{max}} \geq 5$.

Algorithm 1: Emergent property inference

```

1 initialize  $\boldsymbol{\theta}$  by fitting  $q_{\boldsymbol{\theta}}$  to an isotropic gaussian of mean  $\boldsymbol{\mu}_{\text{init}}$  and variance  $\sigma_{\text{init}}^2$ 
2 initialize  $c_0 > 0$  and  $\boldsymbol{\eta}_{\text{opt},0} = \mathbf{0}$ .
3 for Augmented lagrangian epoch  $k = 1, \dots, k_{\max}$  do
4   for SGD iteration  $i = 1, \dots, i_{\max}$  do
5     Sample  $\mathbf{z}_0^{(1)}, \dots, \mathbf{z}_0^{(n)} \sim q_0$ , get transformed variable  $\mathbf{z}^{(j)} = g_{\boldsymbol{\theta}}(\mathbf{z}_0^{(j)})$ ,  $j = 1, \dots, n$ 
6     Update  $\boldsymbol{\theta}$  by descending its stochastic gradient (using ADAM optimizer [93]).  


$$\begin{aligned}\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \boldsymbol{\eta}_{\text{opt},k}, c) = & \frac{1}{n} \sum_{j=1}^n \nabla_{\boldsymbol{\theta}} \log q_{\boldsymbol{\theta}}(\mathbf{z}^{(j)}) + \frac{1}{n} \sum_{j=1}^n \nabla_{\boldsymbol{\theta}} \left( T(\mathbf{z}^{(j)}) - \boldsymbol{\mu}_{\text{opt}} \right) \boldsymbol{\eta}_{\text{opt},k} \\ & + c_k \frac{2}{n} \sum_{j=1}^{\frac{n}{2}} \nabla_{\boldsymbol{\theta}} \left( T(\mathbf{z}^{(j)}) - \boldsymbol{\mu}_{\text{opt}} \right) \cdot \frac{2}{n} \sum_{j=\frac{n}{2}+1}^n \left( T(\mathbf{z}^{(j)}) - \boldsymbol{\mu}_{\text{opt}} \right)\end{aligned}$$

7   end
8   Sample  $\mathbf{z}_0^{(1)}, \dots, \mathbf{z}_0^{(n)} \sim q_0$ , get transformed variable  $\mathbf{z}^{(j)} = g_{\boldsymbol{\theta}}(\mathbf{z}_0^{(j)})$ ,  $j = 1, \dots, n$ 
9   Update  $\boldsymbol{\eta}_{\text{opt},k+1} = \boldsymbol{\eta}_{\text{opt},k} + c_k \frac{1}{n} \sum_{j=1}^n (T(\mathbf{z}^{(j)}) - \boldsymbol{\mu}_{\text{opt}})$ .
10  Update  $c_{k+1} > c_k$  (see text for detail).
11 end

```

944 Rather than starting optimization from some $\boldsymbol{\theta}$ drawn from a randomized distribution, we found
 945 that initializing $q_{\boldsymbol{\theta}}(\mathbf{z})$ to approximate an isotropic gaussian distribution conferred more stable, con-
 946 sistent optimization. The parameters of the gaussian initialization were chosen on an application-
 947 specific basis. Throughout the study, we chose isotropic Gaussian initializations with mean $\boldsymbol{\mu}_{\text{init}}$ at
 948 the center of the support of the distribution and some variance σ_{init}^2 , except for one case, where an
 949 initialization informed by random search was used (see Section 5.2). Deep probability distributions
 950 were fit to these gaussian initializations using 10,000 iterations of stochastic gradient descent on
 951 the evidence lower bound (as in [86]) with Adam optimizer and a learning rate of 10^{-3} .

952 To assess whether the EPI distribution $q_{\boldsymbol{\theta}}(\mathbf{z})$ produces the emergent property, we assess whether
 953 each individual constraint on the means and variances of $f(\mathbf{x}; \mathbf{z})$ is satisfied. We consider the EPI
 954 to have converged when a null hypothesis test of constraint violations $R(\boldsymbol{\theta})_i$ being zero is accepted
 955 for all constraints $i \in \{1, \dots, m\}$ at a significance threshold $\alpha = 0.05$. This significance threshold is
 956 adjusted through Bonferroni correction according to the number of constraints m . The p-values for
 957 each constraint are calculated according to a two-tailed nonparametric test, where 200 estimations
 958 of the sample mean $R(\boldsymbol{\theta})^i$ are made using N_{test} samples of $\mathbf{z} \sim q_{\boldsymbol{\theta}}(\mathbf{z})$ at the end of the augmented
 959 lagrangian epoch. Of all k_{\max} augmented lagrangian epochs, we select the EPI inferred distribution
 960 as that which satisfies the convergence criteria and has greatest entropy.

961 When assessing the suitability of EPI for a particular modeling question, there are some important
 962 technical considerations. First and foremost, as in any optimization problem, the defined emergent
 963 property should always be appropriately conditioned (constraints should not have wildly different
 964 units). Furthermore, if the program is underconstrained (not enough constraints), the distribution
 965 grows (in entropy) unstably unless mapped to a finite support. If overconstrained, there is no
 966 parameter set producing the emergent property, and EPI optimization will fail (appropriately).

967 5.1.5 Example: 2D LDS

968 To gain intuition for EPI, consider a two-dimensional linear dynamical system (2D LDS) model
 969 (Figure 1-figure supplement 1A):

$$\tau \frac{d\mathbf{x}}{dt} = A\mathbf{x} \quad (26)$$

970 with

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}. \quad (27)$$

971 To run EPI with the dynamics matrix elements as the free parameters $\mathbf{z} = [a_{1,1}, a_{1,2}, a_{2,1}, a_{2,2}]$
972 (fixing $\tau = 1s$), the emergent property statistics $f(\mathbf{x}; \mathbf{z})$ were chosen to contain parts of the primary
973 eigenvalue of A , which predicate frequency, $\text{imag}(\lambda_1)$, and the growth/decay, $\text{real}(\lambda_1)$, of the system

$$f(\mathbf{x}; \mathbf{z}) \triangleq \begin{bmatrix} \text{real}(\lambda_1)(\mathbf{x}; \mathbf{z}) \\ \text{imag}(\lambda_1)(\mathbf{x}; \mathbf{z}) \end{bmatrix} \quad (28)$$

974 λ_1 is the eigenvalue of greatest real part when the imaginary component is zero, and alternatively
975 that of positive imaginary component when the eigenvalues are complex conjugate pairs. To learn
976 the distribution of real entries of A that produce a band of oscillating systems around 1Hz, we for-
977 malized this emergent property as $\text{real}(\lambda_1)$ having mean zero with variance 0.25^2 , and the oscillation
978 frequency $\frac{\text{imag}(\lambda_1)}{2\pi}$ having mean 1Hz with variance 0.1Hz^2 :

$$\begin{aligned} \mathcal{X} : \mathbb{E}_{\mathbf{z}, \mathbf{x}} [f(\mathbf{x}; \mathbf{z})] &\triangleq \mathbb{E}_{\mathbf{z}, \mathbf{x}} \begin{bmatrix} \text{real}(\lambda_1)(\mathbf{x}; \mathbf{z}) \\ \text{imag}(\lambda_1)(\mathbf{x}; \mathbf{z}) \end{bmatrix} = \begin{bmatrix} 0 \\ 2\pi \end{bmatrix} \triangleq \boldsymbol{\mu} \\ \text{Var}_{\mathbf{z}, \mathbf{x}} [f(\mathbf{x}; \mathbf{z})] &\triangleq \text{Var}_{\mathbf{z}, \mathbf{x}} \begin{bmatrix} \text{real}(\lambda_1)(\mathbf{x}; \mathbf{z}) \\ \text{imag}(\lambda_1)(\mathbf{x}; \mathbf{z}) \end{bmatrix} = \begin{bmatrix} 0.25^2 \\ (\frac{\pi}{5})^2 \end{bmatrix} \triangleq \boldsymbol{\sigma}^2. \end{aligned} \quad (29)$$

979 To write the emergent property \mathcal{X} in the form required for the augmented lagrangian optimization
980 (Section 5.1.4), we concatenate these first and second moment constraints into a vector of sufficient
981 statistics $T(\mathbf{z})$ and constraint values $\boldsymbol{\mu}_{\text{opt}}$.

$$\mathbb{E}_{\mathbf{z}} [T(\mathbf{z})] \triangleq \mathbb{E}_{\mathbf{z}} \begin{bmatrix} \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x}|\mathbf{z})} [\text{real}(\lambda_1)(\mathbf{x}; \mathbf{z})] \\ \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x}|\mathbf{z})} [\text{imag}(\lambda_1)(\mathbf{x}; \mathbf{z})] \\ \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x}|\mathbf{z})} [(\text{real}(\lambda_1)(\mathbf{x}; \mathbf{z}) - 0)^2] \\ \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x}|\mathbf{z})} [(\text{imag}(\lambda_1)(\mathbf{x}; \mathbf{z}) - 2\pi)^2] \end{bmatrix} = \begin{bmatrix} 0 \\ 2\pi \\ 0.25^2 \\ (\frac{\pi}{5})^2 \end{bmatrix} \triangleq \boldsymbol{\mu}_{\text{opt}}. \quad (30)$$

982 From now on in all scientific applications (Sections 5.2-5.5, we specify how the EPI optimization
983 was setup by specifying $f(\mathbf{x}; \mathbf{z})$, $\boldsymbol{\mu}$, and $\boldsymbol{\sigma}^2$.

984 Unlike the models we presented in the main text, this model admits an analytical form for the
985 mean emergent property statistics given parameter \mathbf{z} , since the eigenvalues can be calculated using
986 the quadratic formula:

$$\lambda = \frac{(\frac{a_{1,1}+a_{2,2}}{\tau}) \pm \sqrt{(\frac{a_{1,1}+a_{2,2}}{\tau})^2 + 4(\frac{a_{1,2}a_{2,1}-a_{1,1}a_{2,2}}{\tau})}}{2}. \quad (31)$$

987 We study this example, because the inferred distribution is curved and multimodal, and we can
988 compare the result of EPI to analytically derived contours of the emergent property statistics.

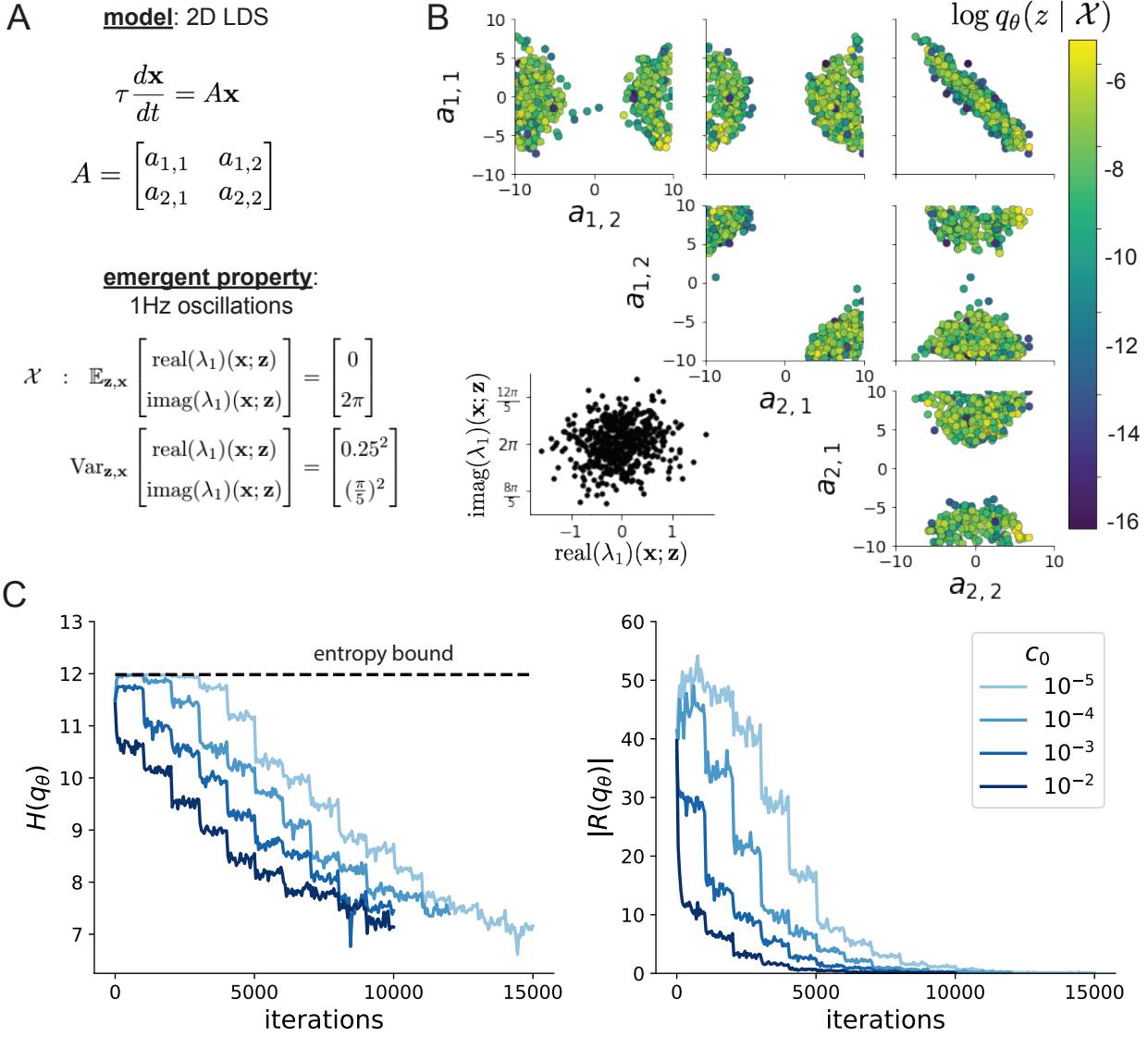


Figure 1-figure supplement 1: **A.** Two-dimensional linear dynamical system model, where real entries of the dynamics matrix A are the parameters. **B.** The EPI distribution for a two-dimensional linear dynamical system with $\tau = 1$ that produces an average of 1Hz oscillations with some small amount of variance. Dashed lines indicate the parameter axes. **C.** Entropy throughout the optimization. At the beginning of each augmented lagrangian epoch ($i_{\max} = 2,000$ iterations), the entropy dipped due to the shifted optimization manifold where emergent property constraint satisfaction is increasingly weighted. **D.** Emergent property moments throughout optimization. At the beginning of each augmented lagrangian epoch, the emergent property moments adjust closer to their constraints.

989 Despite the simple analytic form of the emergent property statistics, the EPI distribution in this
 990 example is not simply determined. Although $\mathbb{E}_{\mathbf{z}} [T(\mathbf{z})]$ is calculable directly via a closed form
 991 function, the distribution $q_{\boldsymbol{\theta}}^*(\mathbf{z} | \mathcal{X})$ cannot be derived directly. This fact is due to the formally hard
 992 problem of the backward mapping: finding the natural parameters $\boldsymbol{\eta}$ from the mean parameters $\boldsymbol{\mu}$
 993 of an exponential family distribution [85]. Instead, we used EPI to approximate this distribution
 994 (Figure 1-figure supplement 1B). We used a real NVP normalizing flow architecture three coupling
 995 layers and two-layer neural networks of 50 units per layer, mapped onto a support of $z_i \in [-10, 10]$.
 996 (see Section 5.1.2).

997 Even this relatively simple system has nontrivial (though intuitively sensible) structure in the
 998 parameter distribution. To validate our method, we analytically derived the contours of the proba-
 999 bility density from the emergent property statistics and values. In the $a_{1,1}$ - $a_{2,2}$ plane, the black line
 1000 at $\text{real}(\lambda_1) = \frac{a_{1,1}+a_{2,2}}{2} = 0$, dashed black line at the standard deviation $\text{real}(\lambda_1) = \frac{a_{1,1}+a_{2,2}}{2} \pm 0.25$,
 1001 and the dashed gray line at twice the standard deviation $\text{real}(\lambda_1) = \frac{a_{1,1}+a_{2,2}}{2} \pm 0.5$ follow the contour
 1002 of probability density of the samples (Figure 1-figure supplement 2A). The distribution precisely
 1003 reflects the desired statistical constraints and model degeneracy in the sum of $a_{1,1}$ and $a_{2,2}$. Intu-
 1004 itively, the parameters equivalent with respect to emergent property statistic $\text{real}(\lambda_1)$ have similar
 1005 log densities.

1006 To explain the bimodality of the EPI distribution, we examined the imaginary component of λ_1 .
 1007 When $\text{real}(\lambda_1) = a_{1,1} + a_{2,2} = 0$ (which is the case on average in \mathcal{X}), we have

$$\text{imag}(\lambda_1) = \begin{cases} \sqrt{\frac{a_{1,1}a_{2,2}-a_{1,2}a_{2,1}}{\tau}}, & \text{if } a_{1,1}a_{2,2} < a_{1,2}a_{2,1} \\ 0 & \text{otherwise} \end{cases}. \quad (32)$$

1008 In Figure 1-figure supplement 2B, we plot the contours of $\text{imag}(\lambda_1)$ where $a_{1,1}a_{2,2}$ is fixed to 0 at one
 1009 standard deviation ($\frac{\pi}{5}$, black dashed) and two standard deviations ($\frac{2\pi}{5}$, gray dashed) from the mean
 1010 of 2π . This validates the curved multimodal structure of the inferred distribution learned through
 1011 EPI. Subtler combinations of model and emergent property will have more complexity, further
 1012 motivating the use of EPI for understanding these systems. As we expect, the distribution results
 1013 in samples of two-dimensional linear systems oscillating near 1Hz (Figure 1-figure supplement 3).

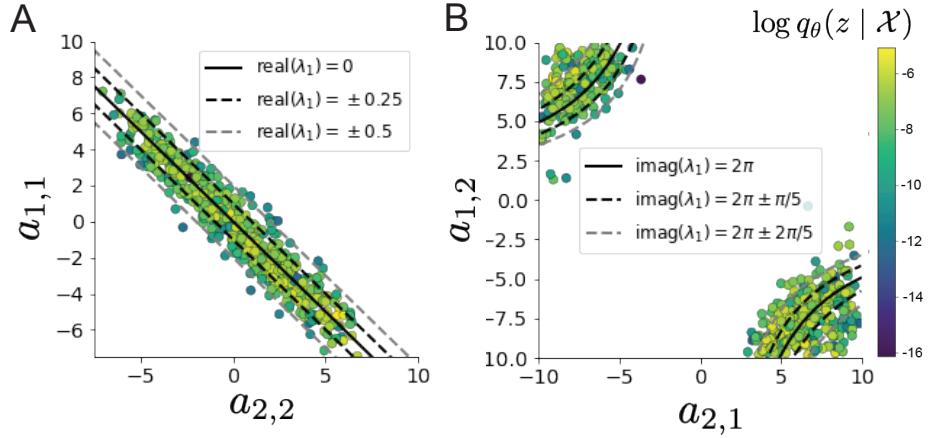


Figure 1-figure supplement 2: **A.** Probability contours in the $a_{1,1}$ - $a_{2,2}$ plane were derived from the relationship to emergent property statistic of growth/decay factor $\text{real}(\lambda_1)$. **B.** Probability contours in the $a_{1,2}$ - $a_{2,1}$ plane were derived from the emergent property statistic of oscillation frequency $2\pi\text{imag}(\lambda_1)$.

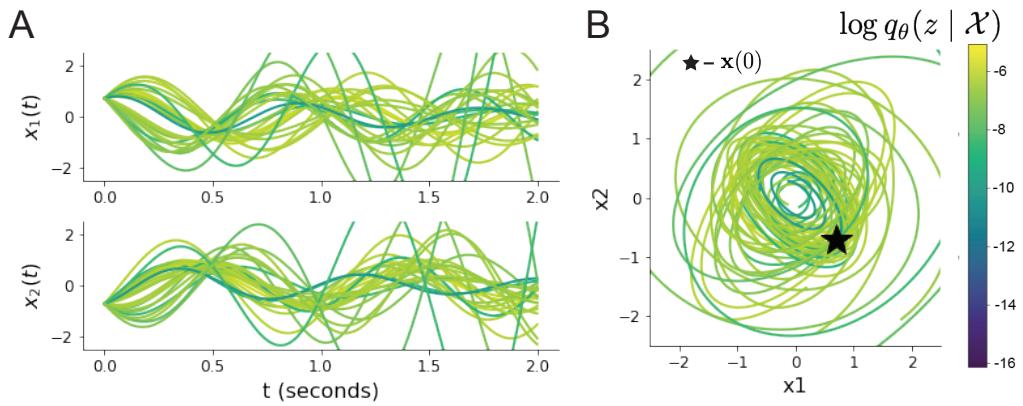


Figure 1-figure supplement 3: Sampled dynamical systems $\mathbf{z} \sim q_\theta(\mathbf{z} \mid \mathcal{X})$ and their simulated activity from $\mathbf{x}(t=0) = [\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}]$ colored by log probability. **A.** Each dimension of the simulated trajectories throughout time. **B.** The simulated trajectories in phase space.

1014 **5.1.6 EPI as variational inference**

1015 In variational inference, a posterior approximation q_{θ}^* is chosen from within some variational family
 1016 \mathcal{Q} to be as close as possible to the posterior under the KL divergence criteria

$$q_{\theta}^*(\mathbf{z}) = \operatorname{argmin}_{q_{\theta} \in \mathcal{Q}} KL(q_{\theta}(\mathbf{z}) \parallel p(\mathbf{z} \mid \mathbf{x})). \quad (33)$$

1017 This KL divergence can be written in terms of entropy of the variational approximation:

$$KL(q_{\theta}(\mathbf{z}) \parallel p(\mathbf{z} \mid \mathbf{x})) = \mathbb{E}_{\mathbf{z} \sim q_{\theta}} [\log(q_{\theta}(\mathbf{z}))] - \mathbb{E}_{\mathbf{z} \sim q_{\theta}} [\log(p(\mathbf{z} \mid \mathbf{x}))] \quad (34)$$

1018

$$= -H(q_{\theta}) - \mathbb{E}_{\mathbf{z} \sim q_{\theta}} [\log(p(\mathbf{x} \mid \mathbf{z})) + \log(p(\mathbf{z})) - \log(p(\mathbf{x}))] \quad (35)$$

1019 Since the marginal distribution of the data $p(\mathbf{x})$ (or ‘‘evidence’’) is independent of θ , variational
 1020 inference is executed by optimizing the remaining expression. This is usually framed as maximizing
 1021 the evidence lower bound (ELBO)

$$\operatorname{argmin}_{q_{\theta} \in \mathcal{Q}} KL(q_{\theta} \parallel p(\mathbf{z} \mid \mathbf{x})) = \operatorname{argmax}_{q_{\theta} \in \mathcal{Q}} H(q_{\theta}) + \mathbb{E}_{\mathbf{z} \sim q_{\theta}} [\log(p(\mathbf{x} \mid \mathbf{z})) + \log(p(\mathbf{z}))]. \quad (36)$$

1022 Now, we will show how the maximum entropy problem of EPI is equivalent to variational inference.

1023 In general, a maximum entropy problem (as in Equation 16) has an equivalent lagrange dual form:

$$\begin{aligned} \operatorname{argmax}_{q \in \mathcal{Q}} H(q(\mathbf{z})) &\iff \operatorname{argmax}_{q \in \mathcal{Q}} H(q(\mathbf{z})) + \boldsymbol{\eta}^{*\top} \mathbb{E}_{\mathbf{z} \sim q} [T(\mathbf{z})], \\ \text{s.t. } \mathbb{E}_{\mathbf{z} \sim q} [T(\mathbf{z})] &= \mathbf{0} \end{aligned} \quad (37)$$

1024 with lagrange multipliers $\boldsymbol{\eta}^*$. By moving the lagrange multipliers within the expectation

$$q^* = \operatorname{argmax}_{q \in \mathcal{Q}} H(q(\mathbf{z})) + \mathbb{E}_{\mathbf{z} \sim q} \left[\boldsymbol{\eta}^{*\top} T(\mathbf{z}) \right], \quad (38)$$

1025 inserting a $\log \exp(\cdot)$ within the expectation,

$$q^* = \operatorname{argmax}_{q \in \mathcal{Q}} H(q(\mathbf{z})) + \mathbb{E}_{\mathbf{z} \sim q} \left[\log \exp \left(\boldsymbol{\eta}^{*\top} T(\mathbf{z}) \right) \right], \quad (39)$$

1026 and finally choosing $T(\cdot)$ to be likelihood averaged statistics as in EPI

$$q^* = \operatorname{argmax}_{q \in \mathcal{Q}} H(q(\mathbf{z})) + \mathbb{E}_{\mathbf{z} \sim q} \left[\log \exp \left(\boldsymbol{\eta}^{*\top} \begin{bmatrix} \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x} \mid \mathbf{z})} [\phi_1(\mathbf{x}; \mathbf{z})] \\ \dots \\ \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x} \mid \mathbf{z})} [\phi_m(\mathbf{x}; \mathbf{z})] \end{bmatrix} \right) \right], \quad (40)$$

1027 we can compare directly to the objective used in variational inference (Equation 36). We see
 1028 that EPI is exactly variational inference with an exponential family likelihood defined by sufficient

1029 statistics $T(\mathbf{z}) = \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x}|\mathbf{z})} [\phi(\mathbf{x}; \mathbf{z})]$, and where the natural parameter $\boldsymbol{\eta}^*$ is predicated by the mean
 1030 parameter $\boldsymbol{\mu}_{\text{opt}}$. Equation 40 implies that EPI uses an improper (or uniform) prior, which is easily
 1031 changed.

1032 This derivation of the equivalence between EPI and variational inference emphasizes why defining
 1033 a statistical inference program by its mean parameterization $\boldsymbol{\mu}_{\text{opt}}$ is so useful. With EPI, one can
 1034 clearly define the emergent property \mathcal{X} that the model of interest should produce through intuitive
 1035 selection of $\boldsymbol{\mu}_{\text{opt}}$ for a given $T(\mathbf{z})$. Alternatively, figuring out the correct natural parameters $\boldsymbol{\eta}^*$ for
 1036 the same $T(\mathbf{z})$ that produces \mathcal{X} is a formally hard problem.

1037 5.2 Stomatogastric ganglion

1038 In Section 3.1 and 3.2, we used EPI to infer conductance parameters in a model of the stomatogastric
 1039 ganglion (STG) [41]. This 5-neuron circuit model represents two subcircuits: that generating the
 1040 pyloric rhythm (fast population) and that generating the gastric mill rhythm (slow population).
 1041 The additional neuron (the IC neuron of the STG) receives inhibitory synaptic input from both
 1042 subcircuits, and can couple to either rhythm dependent on modulatory conditions. There is also
 1043 a parametric regime in which this neuron fires at an intermediate frequency between that of the
 1044 fast and slow populations [41], which we infer with EPI as a motivational example. This model
 1045 is not to be confused with an STG subcircuit model of the pyloric rhythm [68], which has been
 1046 statistically inferred in other studies [15, 35].

1047 5.2.1 STG model

1048 We analyze how the parameters $\mathbf{z} = [g_{el}, g_{synA}]$ govern the emergent phenomena of intermediate
 1049 hub frequency in a model of the stomatogastric ganglion (STG) [41] shown in Figure 1A with
 1050 activity $\mathbf{x} = [x_{f1}, x_{f2}, x_{hub}, x_{s1}, x_{s2}]$, using the same hyperparameter choices as Gutierrez et al.
 1051 Each neuron’s membrane potential $x_\alpha(t)$ for $\alpha \in \{f1, f2, hub, s1, s2\}$ is the solution of the following
 1052 stochastic differential equation:

$$C_m \frac{dx_\alpha}{dt} = -[h_{leak}(\mathbf{x}; \mathbf{z}) + h_{Ca}(\mathbf{x}; \mathbf{z}) + h_K(\mathbf{x}; \mathbf{z}) + h_{hyp}(\mathbf{x}; \mathbf{z}) + h_{elec}(\mathbf{x}; \mathbf{z}) + h_{syn}(\mathbf{x}; \mathbf{z})] + dB. \quad (41)$$

1053 The input current of each neuron is the sum of the leak, calcium, potassium, hyperpolarization,
 1054 electrical and synaptic currents. Each current component is a function of all membrane potentials
 1055 and the conductance parameters \mathbf{z} . Finally, we include gaussian noise dB to the model of Gutierrez
 1056 et al. so that the model stochastic, although this is not required by EPI.

1057 The capacitance of the cell membrane was set to $C_m = 1nF$. Specifically, the currents are the
 1058 difference in the neuron's membrane potential and that current type's reversal potential multiplied
 1059 by a conductance:

$$1060 \quad h_{leak}(\mathbf{x}; \mathbf{z}) = g_{leak}(x_\alpha - V_{leak}) \quad (42)$$

$$1061 \quad h_{elec}(\mathbf{x}; \mathbf{z}) = g_{el}(x_\alpha^{post} - x_\alpha^{pre}) \quad (43)$$

$$1062 \quad h_{syn}(\mathbf{x}; \mathbf{z}) = g_{syn}S_\infty^{pre}(x_\alpha^{post} - V_{syn}) \quad (44)$$

$$1063 \quad h_{Ca}(\mathbf{x}; \mathbf{z}) = g_{Ca}M_\infty(x_\alpha - V_{Ca}) \quad (45)$$

$$1064 \quad h_K(\mathbf{x}; \mathbf{z}) = g_KN(x_\alpha - V_K) \quad (46)$$

$$1064 \quad h_{hyp}(\mathbf{x}; \mathbf{z}) = g_hH(x_\alpha - V_{hyp}). \quad (47)$$

1065 The reversal potentials were set to $V_{leak} = -40mV$, $V_{Ca} = 100mV$, $V_K = -80mV$, $V_{hyp} = -20mV$,
 1066 and $V_{syn} = -75mV$. The other conductance parameters were fixed to $g_{leak} = 1 \times 10^{-4}\mu S$. g_{Ca} ,
 1067 g_K , and g_{hyp} had different values based on fast, intermediate (hub) or slow neuron. The fast
 1068 conductances had values $g_{Ca} = 1.9 \times 10^{-2}$, $g_K = 3.9 \times 10^{-2}$, and $g_{hyp} = 2.5 \times 10^{-2}$. The intermediate
 1069 conductances had values $g_{Ca} = 1.7 \times 10^{-2}$, $g_K = 1.9 \times 10^{-2}$, and $g_{hyp} = 8.0 \times 10^{-3}$. Finally, the
 1070 slow conductances had values $g_{Ca} = 8.5 \times 10^{-3}$, $g_K = 1.5 \times 10^{-2}$, and $g_{hyp} = 1.0 \times 10^{-2}$.

1071 Furthermore, the Calcium, Potassium, and hyperpolarization channels have time-dependent gating
 1072 dynamics dependent on steady-state gating variables M_∞ , N_∞ and H_∞ , respectively:

$$1073 \quad M_\infty = 0.5 \left(1 + \tanh \left(\frac{x_\alpha - v_1}{v_2} \right) \right) \quad (48)$$

$$1074 \quad \frac{dN}{dt} = \lambda_N(N_\infty - N) \quad (49)$$

$$1075 \quad N_\infty = 0.5 \left(1 + \tanh \left(\frac{x_\alpha - v_3}{v_4} \right) \right) \quad (50)$$

$$1076 \quad \lambda_N = \phi_N \cosh \left(\frac{x_\alpha - v_3}{2v_4} \right) \quad (51)$$

$$1077 \quad \frac{dH}{dt} = \frac{(H_\infty - H)}{\tau_h} \quad (52)$$

$$1078 \quad H_\infty = \frac{1}{1 + \exp \left(\frac{x_\alpha + v_5}{v_6} \right)} \quad (53)$$

$$1078 \quad \tau_h = 272 - \left(\frac{-1499}{1 + \exp \left(\frac{-x_\alpha + v_7}{v_8} \right)} \right). \quad (54)$$

1079 where we set $v_1 = 0mV$, $v_2 = 20mV$, $v_3 = 0mV$, $v_4 = 15mV$, $v_5 = 78.3mV$, $v_6 = 10.5mV$,
 1080 $v_7 = -42.2mV$, $v_8 = 87.3mV$, $v_9 = 5mV$, and $v_{th} = -25mV$.

1081 Finally, there is a synaptic gating variable as well:

$$S_\infty = \frac{1}{1 + \exp\left(\frac{v_{th} - x_\alpha}{v_9}\right)}. \quad (55)$$

1082 When the dynamic gating variables are considered, this is actually a 15-dimensional nonlinear
 1083 dynamical system. The gaussian noise $d\mathbf{B}$ has variance $(1 \times 10^{-12})^2$ A², and introduces variability
 1084 in frequency at each parameterization \mathbf{z} .

1085 5.2.2 Hub frequency calculation

1086 In order to measure the frequency of the hub neuron during EPI, the STG model was simulated for
 1087 $T = 300$ time steps of $dt = 25\text{ms}$. The chosen dt and T were the most computationally convenient
 1088 choices yielding accurate frequency measurement. We used a basis of complex exponentials with
 1089 frequencies from 0.0-1.0 Hz at 0.01Hz resolution to measure frequency from simulated time series

$$\Phi = [0.0, 0.01, \dots, 1.0]^\top \dots \quad (56)$$

1090 To measure spiking frequency, we processed simulated membrane potentials with a relu (spike
 1091 extraction) and low-pass filter with averaging window of size 20, then took the frequency with the
 1092 maximum absolute value of the complex exponential basis coefficients of the processed time-series.
 1093 The first 20 temporal samples of the simulation are ignored to account for initial transients.

1094 To differentiate through the maximum frequency identification, we used a soft-argmax Let $X_\alpha \in$
 1095 $\mathcal{C}^{|\Phi|}$ be the complex exponential filter bank dot products with the signal $x_\alpha \in \mathbb{R}^N$, where $\alpha \in$
 1096 {f1, f2, hub, s1, s2}. The soft-argmax is then calculated using temperature parameter $\beta_\psi = 100$

$$\psi_\alpha = \text{softmax}(\beta_\psi |X_\alpha| \odot i), \quad (57)$$

1097 where $i = [0, 1, \dots, 100]$. The frequency is then calculated as

$$\omega_\alpha = 0.01\psi_\alpha \text{Hz}. \quad (58)$$

1098 Intermediate hub frequency, like all other emergent properties in this work, is defined by the mean
 1099 and variance of the emergent property statistics. In this case, we have one statistic, hub neuron
 1100 frequency, where the mean was chosen to be 0.55Hz,(Equation 2) and variance was chosen to be
 1101 0.025^2 Hz² (Equation 3).

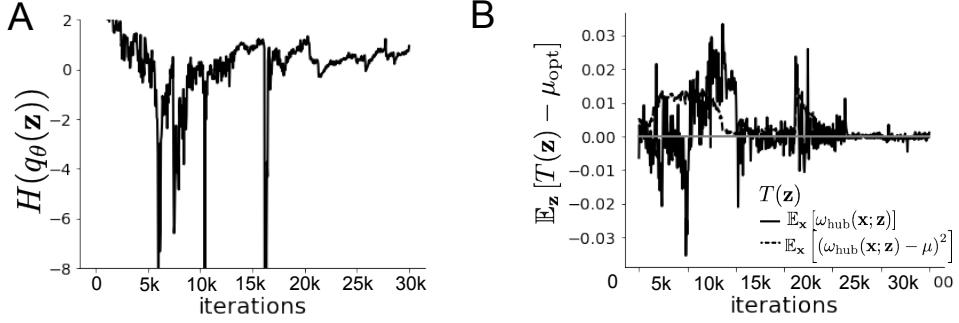


Figure 1-figure supplement 4: EPI optimization of the STG model producing network syncing.

A. Entropy throughout optimization. **B.** The emergent property statistic means and variances converge to their constraints at 25,000 iterations following the fifth augmented lagrangian epoch.

1102 **5.2.3 EPI details for the STG model**

1103 EPI was run for the STG model using

$$f(\mathbf{x}; \mathbf{z}) = \omega_{\text{hub}}(\mathbf{x}; \mathbf{z}), \quad (59)$$

1104

$$\boldsymbol{\mu} = [0.55], \quad (60)$$

1105 and

$$\boldsymbol{\sigma}^2 = [0.025^2] \quad (61)$$

1106 (see Sections 5.1.3-5.1.4, and example in Section 5.1.5). Throughout optimization, the augmented
1107 lagrangian parameters η and c , were updated after each epoch of $i_{\max} = 5,000$ iterations (see
1108 Section 5.1.4). The optimization converged after five epochs (Figure 1-figure supplement 4).

1109 For EPI in Fig 1E, we used a real NVP architecture with three coupling layers and two-layer
1110 neural networks of 25 units per layer. The normalizing flow architecture mapped $\mathbf{z}_0 \sim \mathcal{N}(\mathbf{0}, I)$ to
1111 a support of $\mathbf{z} = [g_{\text{el}}, g_{\text{synA}}] \in [4, 8] \times [0.01, 4]$, initialized to a gaussian approximation of samples
1112 returned by a preliminary ABC search. We did not include $g_{\text{synA}} < 0.01$, for numerical stability.
1113 EPI optimization was run with an augmented lagrangian coefficient of $c_0 = 10^5$, hyperparameter
1114 $\beta = 2$, a batch size $n = 400$, and we simulated one $\mathbf{x}^{(i)}$ per $\mathbf{z}^{(i)}$. The architecture converged with
1115 criteria $N_{\text{test}} = 100$.

1116 **5.2.4 Hessian sensitivity vectors**

1117 To quantify the second-order structure of the EPI distribution, we evaluated the Hessian of the log
1118 probability $\frac{\partial^2 \log q(\mathbf{z}|\mathcal{X})}{\partial \mathbf{z} \mathbf{z}^\top}$. The eigenvector of this Hessian with most negative eigenvalue is defined as
1119 the sensitivity dimension \mathbf{v}_1 , and all subsequent eigenvectors are ordered by increasing eigenvalue.
1120 These eigenvalues are quantifications of how fast the emergent property deteriorates via the param-
1121 eter combination of their associated eigenvector. In Figure 1D, the sensitivity dimension v_1 (solid)
1122 and the second eigenvector of the Hessian v_2 (dashed) are shown evaluated at the mode of the
1123 distribution. Since the Hessian eigenvectors have sign degeneracy, the visualized directions in 2-D
1124 parameter space were chosen to have positive g_{synA} . The length of the arrows is inversely propor-
1125 tional to the square root of the absolute value of their eigenvalues $\lambda_1 = -10.7$ and $\lambda_2 = -3.22$. For
1126 the same magnitude perturbation away from the mode, intermediate hub frequency only diminishes
1127 along the sensitivity dimension \mathbf{v}_1 (Figure 1E-F).

1128 **5.3 Scaling EPI for stable amplification in RNNs**

1129 **5.3.1 Rank-2 RNN model**

1130 We examined the scaling properties of EPI by learning connectivities of RNNs of increasing size
1131 that exhibit stable amplification. Rank-2 RNN connectivity was modeled as $W = UV^\top$, where
1132 $U = [\mathbf{U}_1 \ \mathbf{U}_2] + g\chi^{(W)}$, $V = [\mathbf{V}_1 \ \mathbf{V}_2] + g\chi^{(V)}$, and $\chi_{i,j}^{(W)}, \chi_{i,j}^{(V)} \sim \mathcal{N}(0, 1)$. This RNN model has
1133 dynamics

$$\tau \dot{\mathbf{x}} = -\mathbf{x} + W\mathbf{x}. \quad (62)$$

1134 In this analysis, we inferred connectivity parameterizations $\mathbf{z} = [\mathbf{U}_1^\top, \mathbf{U}_2^\top, \mathbf{V}_1^\top, \mathbf{V}_2^\top]^\top \in [-1, 1]^{(4N)}$
1135 that produced stable amplification using EPI, SMC-ABC [26], and SNPE [35] (see Section Related
1136 Methods).

1137 **5.3.2 Stable amplification**

1138 For this RNN model to be stable, all real eigenvalues of W must be less than 1: $\text{real}(\lambda_1) < 1$,
1139 where λ_1 denotes the greatest real eigenvalue of W . For a stable RNN to amplify at least one input
1140 pattern, the symmetric connectivity $W^s = \frac{W+W^\top}{2}$ must have an eigenvalue greater than 1: $\lambda_1^s > 1$,
1141 where λ^s is the maximum eigenvalue of W^s . These two conditions are necessary and sufficient for
1142 stable amplification in RNNs [51].

1143 **5.3.3 EPI details for RNNs**

1144 We defined the emergent property of stable amplification with means of these eigenvalues (0.5
 1145 and 1.5, respectively) that satisfy these conditions. To complete the emergent property definition,
 1146 we chose variances (0.25^2) about those means such that samples rarely violate the eigenvalue
 1147 constraints. To write the emergent property of Equation 5 in terms of the EPI optimization, we
 1148 have

$$f(\mathbf{x}; \mathbf{z}) = \begin{bmatrix} \text{real}(\lambda_1)(\mathbf{x}; \mathbf{z}) \\ \lambda_1^s(\mathbf{x}; \mathbf{z}) \end{bmatrix}, \quad (63)$$

1149

$$\boldsymbol{\mu} = \begin{bmatrix} 0.5 \\ 1.5 \end{bmatrix}, \quad (64)$$

1150 and

$$\boldsymbol{\sigma}^2 = \begin{bmatrix} 0.25^2 \\ 0.25^2 \end{bmatrix} \quad (65)$$

1151 (see Sections 5.1.3-5.1.4, and example in Section 5.1.5). Gradients of maximum eigenvalues of Her-
 1152 mitian matrices like W^s are available with modern automatic differentiation tools. To differentiate
 1153 through the $\text{real}(\lambda_1)$, we solved the following equation for eigenvalues of rank-2 matrices using the
 1154 rank reduced matrix $W^r = V^\top U$

$$\lambda_{\pm} = \frac{\text{Tr}(W^r) \pm \sqrt{\text{Tr}(W^r)^2 - 4\text{Det}(W^r)}}{2}. \quad (66)$$

1155 For EPI in Figure 2, we used a real NVP architecture with three coupling layers of affine transfor-
 1156 mations parameterized by two-layer neural networks of 100 units per layer. The initial distribution
 1157 was a standard isotropic gaussian $\mathbf{z}_0 \sim \mathcal{N}(\mathbf{0}, I)$ mapped to the support of $\mathbf{z}_i \in [-1, 1]$. We used
 1158 an augmented lagrangian coefficient of $c_0 = 10^3$, a batch size $n = 200$, $\beta = 4$, and we simulated
 1159 one $\mathbf{W}^{(i)}$ per $\mathbf{z}^{(i)}$. We chose to use $i_{\max} = 500$ iterations per augmented lagrangian epoch and
 1160 emergent property constraint convergence was evaluated at $N_{\text{test}} = 200$ (Figure 2B blue line, and
 1161 Figure 2C-D blue). It was fastest to initialize the EPI distribution on a Tesla V100 GPU, and then
 1162 subsequently optimize it on a CPU with 32 cores. EPI timing measurements accounted for this
 1163 initialization period.

1164 **5.3.4 Methodological comparison**

1165 We compared EPI to two alternative simulation-based inference techniques, since the likelihood
 1166 of these eigenvalues given \mathbf{z} is not available. Approximate bayesian computation (ABC) [24] is a

1167 rejection sampling technique for obtaining sets of parameters \mathbf{z} that produce activity \mathbf{x} close to some
 1168 observed data \mathbf{x}_0 . Sequential Monte Carlo approximate bayesian computation (SMC-ABC) is the
 1169 state-of-the-art ABC method, which leverages SMC techniques to improve sampling speed. We ran
 1170 SMC-ABC with the pyABC package [94] to infer RNNs with stable amplification: connectivities
 1171 having eigenvalues within an ϵ -defined l_2 distance of

$$\mathbf{x}_0 = \begin{bmatrix} \text{real}(\lambda_1) \\ \lambda_1^s \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1.5 \end{bmatrix}. \quad (67)$$

1172 SMC-ABC was run with a uniform prior over $\mathbf{z} \in [-1, 1]^{(4N)}$, a population size of 1,000 particles
 1173 with simulations parallelized over 32 cores, and a multivariate normal transition model.

1174 SNPE, the next approach in our comparison, is far more similar to EPI. Like EPI, SNPE treats pa-
 1175 rameters in mechanistic models with deep probability distributions, yet the two learning algorithms
 1176 are categorically different. SNPE uses a two-network architecture to approximate the posterior dis-
 1177 tribution of the model conditioned on observed data \mathbf{x}_0 . The amortizing network maps observations
 1178 \mathbf{x}_i to the parameters of the deep probability distribution. The weights and biases of the parameter
 1179 network are optimized by sequentially augmenting the training data with additional pairs $(\mathbf{z}_i, \mathbf{x}_i)$
 1180 based on the most recent posterior approximation. This sequential procedure is important to get
 1181 training data \mathbf{z}_i to be closer to the true posterior, and \mathbf{x}_i to be closer to the observed data. For
 1182 the deep probability distribution architecture, we chose a masked autoregressive flow with affine
 1183 couplings (the default choice), three transforms, 50 hidden units, and a normalizing flow mapping
 1184 to the support as in EPI. This architectural choice closely tracked the size of the architecture used
 1185 by EPI (Figure 2-figure supplement 1). As in SMC-ABC, we ran SNPE with $\mathbf{x}_0 = \mu$. All SNPE
 1186 optimizations were run for a limit of 1.5 days, or until two consecutive rounds resulted in a valida-
 1187 tion log probability lower than the maximum observed for that random seed. It was always faster
 1188 to run SNPE on a CPU with 32 cores rather than on a Tesla V100 GPU.

1189 To compare the efficiency of these algorithms for inferring RNN connectivity distributions producing
 1190 stable amplification, we develop a convergence criteria that can be used across methods. While EPI
 1191 has its own hypothesis testing convergence criteria for the emergent property, it would not make
 1192 sense to use this criteria on SNPE and SMC-ABC which do not constrain the means and variances
 1193 of their predictions. Instead, we consider EPI and SNPE to have converged after completing its
 1194 most recent optimization epoch (EPI) or round (SNPE) in which the distance $\|\mathbb{E}_{\mathbf{z}, \mathbf{x}} [f(\mathbf{x}; \mathbf{z})] - \mu\|_2$
 1195 is less than 0.5. We consider SMC-ABC to have converged once the population produces samples
 1196 within the $\epsilon = 0.5$ ball ensuring stable amplification.

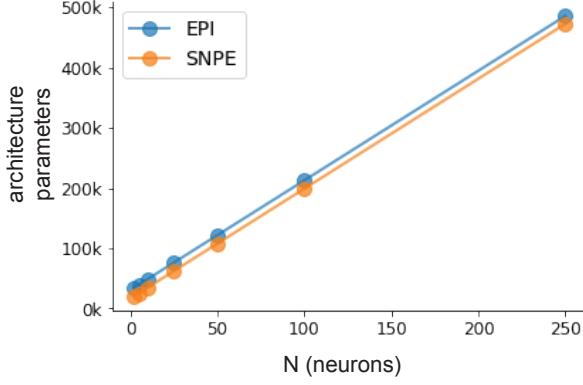


Figure 2-figure supplement 1: Number of parameters in deep probability distribution architectures of EPI (blue) and SNPE (orange) by RNN size (N).

When assessing the scalability of SNPE, it is important to check that alternative hyperparameterizations could not yield better performance. Key hyperparameters of the SNPE optimization are the number of simulations per round n_{round} , the number of atoms used in the atomic proposals of the SNPE-C algorithm [95], and the batch size n . To match EPI, we used a batch size of $n = 200$ for $N \leq 25$, however we found $n = 1,000$ to be helpful for SNPE in higher dimensions. While $n_{\text{round}} = 1,000$ yielded SNPE convergence for $N \leq 25$, we found that a substantial increase to $n_{\text{round}} = 25,000$ yielded more consistent convergence at $N = 50$ (Figure 2-figure supplement 2A). By increasing n_{round} , we also necessarily increase the duration of each round. At $N = 100$, we tried two hyperparameter modifications. As suggested in [95], we increased n_{atom} by an order of magnitude to improve gradient quality, but this had little effect on the optimization (much overlap between same random seeds) (Figure 2-figure supplement 2B). Finally, we increased n_{round} by an order of magnitude, which yielded convergence in one case, but no others. We found no way to improve the convergence rate of SNPE without making more aggressive hyperparameter choices requiring high numbers of simulations. In Figure 2C-D, we show samples from the random seed resulting in emergent property convergence at greatest entropy (EPI), the random seed resulting in greatest validation log probability (SNPE), and the result of all converged random seeds (SMC).

5.3.5 Effect of RNN parameters on EPI and SNPE inferred distributions

To clarify the difference in objectives of EPI and SNPE, we show their results on RNN models with different numbers of neurons N and random strength g . The parameters inferred by EPI

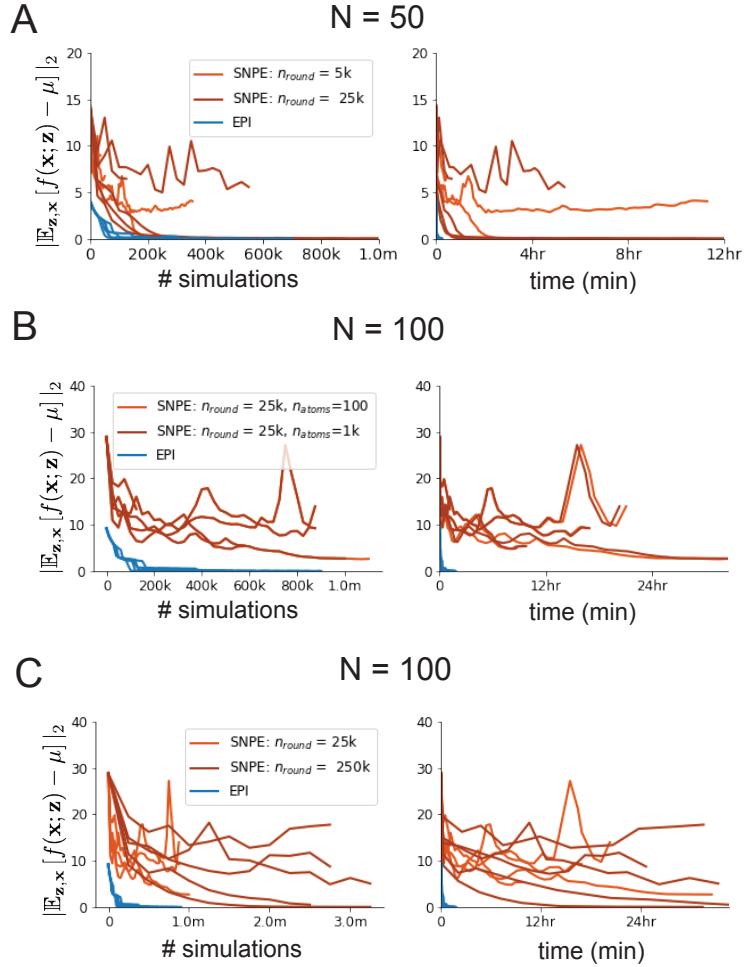


Figure 2-figure supplement 2: SNPE convergence was enabled by increasing n_{round} , not n_{atom} . **A.** Difference of mean predictions \mathbf{x}_0 throughout optimization at $N = 50$ with by simulation count (left) and wall time (right) of SNPE with $n_{\text{round}} = 5,000$ (light orange), SNPE with $n_{\text{round}} = 25,000$ (dark orange), and EPI (blue). Each line shows an individual random seed. **B.** Same conventions as A at $N = 100$ of SNPE with $n_{\text{atom}} = 100$ (light orange) and $n_{\text{atom}} = 1,000$ (dark orange). **C.** Same conventions as A at $N = 100$ of SNPE with $n_{\text{round}} = 25,000$ (light orange) and $n_{\text{round}} = 250,000$ (dark orange).

1216 consistently produces the same mean and variance of $\text{real}(\lambda_1)$ and λ_1^s , while those inferred by
1217 SNPE change according to the model definition (Figure 2-figure supplement 3A). For $N = 2$ and
1218 $g = 0.01$, the SNPE posterior has greater concentration in eigenvalues around \mathbf{x}_0 than at $g = 0.1$,
1219 where the model has greater randomness (Figure 2-figure supplement 3B top, orange). At both
1220 levels of g when $N = 2$, the posterior of SNPE has lower entropy than EPI at convergence (Figure
1221 2-figure supplement 3B top). However at $N = 10$, SNPE results in a predictive distribution
1222 of more widely dispersed eigenvalues (Figure 2-figure supplement 3A bottom), and an inferred
1223 posterior with greater entropy than EPI (Figure 2-figure supplement 3B bottom). We highlight
1224 these differences not to focus on an insightful trend, but to emphasize that these methods optimize
1225 different objectives with different implications.

1226 Note that SNPE converges when it's validation log probability has saturated after several rounds
1227 of optimization (Figure 2-figure supplement 3C), and that EPI converges after several epochs of
1228 its own optimization to enforce the emergent property constraints (Figure 2-figure supplement 3D
1229 blue). Importantly, as SNPE optimizes its posterior approximation, the predictive means change,
1230 and at convergence may be different than \mathbf{x}_0 (Figure 2-figure supplement 3D orange, left). It is
1231 sensible to assume that predictions of a well-approximated SNPE posterior should closely reflect
1232 the data on average (especially given a uniform prior and a low degree of stochasticity), however
1233 this is not a given. Furthermore, no aspect of the SNPE optimization controls the variance of the
1234 predictions (Figure 2-figure supplement 3D orange, right).

1235 5.4 Primary visual cortex

1236 5.4.1 V1 model

1237 E-I circuit models, rely on the assumption that inhibition can be studied as an indivisible unit,
1238 despite ample experimental evidence showing that inhibition is instead composed of distinct ele-
1239 ments [63]. In particular three types of genetically identified inhibitory cell-types – parvalbumin
1240 (P), somatostatin (S), VIP (V) – compose 80% of GABAergic interneurons in V1 [61–63], and
1241 follow specific connectivity patterns (Figure 3A) [64], which lead to cell-type specific computa-
1242 tions [47, 96]. Currently, how the subdivision of inhibitory cell-types, shapes correlated variability
1243 by reconfiguring recurrent network dynamics is not understood.

1244 In the stochastic stabilized supralinear network [59], population rate responses \mathbf{x} to mean input \mathbf{h} ,

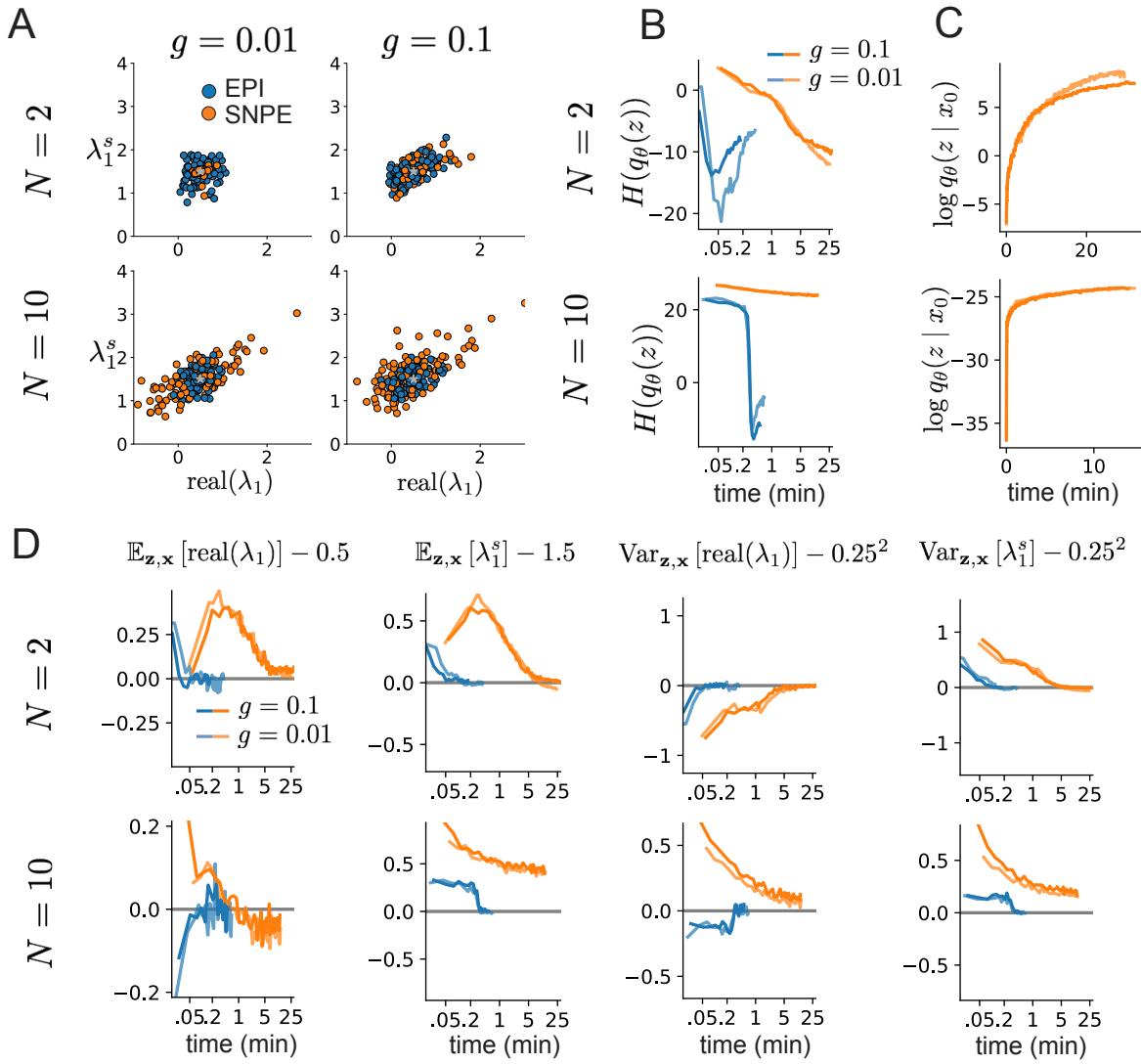


Figure 2-figure supplement 3: Model characteristics affect predictions of posteriors inferred by SNPE, while predictions of parameters inferred by EPI remain fixed. **A.** Predictive distribution of EPI (blue) and SNPE (orange) inferred connectivity of RNNs exhibiting stable amplification with $N = 2$ (top), $N = 10$ (bottom), $g = 0.01$ (left), and $g = 0.1$ (right). **B.** Entropy of parameter distribution approximations throughout optimization with $N = 2$ (top), $N = 10$ (bottom), $g = 0.1$ (dark shade), and $g = 0.01$ (light shade). **C.** Validation log probabilities throughout SNPE optimization. Same conventions as B. **D.** Adherence to EPI constraints. Same conventions as B.

₁₂₄₅ recurrent input $W\mathbf{x}$ and slow noise ϵ are governed by

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + \phi(W\mathbf{x} + \mathbf{h} + \epsilon), \quad (68)$$

₁₂₄₆ where the noise is an Ornstein-Uhlenbeck process $\epsilon \sim OU(\tau_{\text{noise}}, \sigma)$

$$\tau_{\text{noise}} d\epsilon_\alpha = -\epsilon_\alpha dt + \sqrt{2\tau_{\text{noise}}} \tilde{\sigma}_\alpha dB \quad (69)$$

₁₂₄₇ with $\tau_{\text{noise}} = 5\text{ms} > \tau = 1\text{ms}$. The noisy process is parameterized as

$$\tilde{\sigma}_\alpha = \sigma_\alpha \sqrt{1 + \frac{\tau}{\tau_{\text{noise}}}}, \quad (70)$$

₁₂₄₈ so that σ parameterizes the variance of the noisy input in the absence of recurrent connectivity
₁₂₄₉ ($W = \mathbf{0}$). As contrast $c \in [0, 1]$ increases, input to the E- and P-populations increases relative to
₁₂₅₀ a baseline input $\mathbf{h} = \mathbf{h}_b + c\mathbf{h}_c$. Connectivity (W_{fit}) and input ($\mathbf{h}_{b,\text{fit}}$ and $\mathbf{h}_{c,\text{fit}}$) parameters were fit
₁₂₅₁ using the deterministic V1 circuit model [47]

$$W_{\text{fit}} = \begin{bmatrix} W_{EE} & W_{EP} & W_{ES} & W_{EV} \\ W_{PE} & W_{PP} & W_{PS} & W_{PV} \\ W_{SE} & W_{SP} & W_{SS} & W_{SV} \\ W_{VE} & W_{VP} & W_{VS} & W_{VV} \end{bmatrix} = \begin{bmatrix} 2.18 & -1.19 & -.594 & -.229 \\ 1.66 & -.651 & -.680 & -.242 \\ .895 & -5.22 \times 10^{-3} & -1.51 \times 10^{-4} & -.761 \\ 3.34 & -2.31 & -.254 & -2.52 \times 10^{-4} \end{bmatrix}, \quad (71)$$

$$\mathbf{h}_{b,\text{fit}} = \begin{bmatrix} .416 \\ .429 \\ .491 \\ .486 \end{bmatrix}, \quad (72)$$

₁₂₅₂ and

$$\mathbf{h}_{c,\text{fit}} = \begin{bmatrix} .359 \\ .403 \\ 0 \\ 0 \end{bmatrix}. \quad (73)$$

₁₂₅₃ To obtain rates on a realistic scale (100-fold greater), we map these fitted parameters to an equivalence class
₁₂₅₄

$$W = \begin{bmatrix} W_{EE} & W_{EP} & W_{ES} & W_{EV} \\ W_{PE} & W_{PP} & W_{PS} & W_{PV} \\ W_{SE} & W_{SP} & W_{SS} & W_{SV} \\ W_{VE} & W_{VP} & W_{VS} & W_{VV} \end{bmatrix} = \begin{bmatrix} .218 & -.119 & -.0594 & -.0229 \\ .166 & -.0651 & -.068 & -.0242 \\ .0895 & -5.22 \times 10^{-4} & -1.51 \times 10^{-5} & -.0761 \\ .334 & -.231 & -.0254 & -2.52 \times 10^{-5} \end{bmatrix}, \quad (74)$$

$$\mathbf{h}_b = \begin{bmatrix} h_{b,E} \\ h_{b,P} \\ h_{b,S} \\ h_{b,V} \end{bmatrix} = \begin{bmatrix} 4.16 \\ 4.29 \\ 4.91 \\ 4.86 \end{bmatrix}, \quad (75)$$

1255 and

$$\mathbf{h}_c = \begin{bmatrix} h_{c,E} \\ h_{c,P} \\ h_{c,S} \\ h_{c,V} \end{bmatrix} = \begin{bmatrix} 3.59 \\ 4.03 \\ 0 \\ 0 \end{bmatrix}. \quad (76)$$

1256 Circuit responses are simulated using $T = 200$ time steps at $dt = 0.5\text{ms}$ from an initial condition
 1257 drawn from $\mathbf{x}(0) \sim U[10\text{Hz}, 25\text{Hz}]$. Standard deviation of the E-population $s_E(\mathbf{x}; \mathbf{z})$ is calculated
 1258 as the square root of the temporal variance from $t_{ss} = 75\text{ms}$ to $Tdt = 100\text{ms}$

$$s_E(\mathbf{x}; \mathbf{z}) = \sqrt{\mathbb{E}_{t>t_{ss}} [(x_E(t) - \mathbb{E}_{t>t_{ss}} [x_E(t)])^2]}. \quad (77)$$

1259 5.4.2 EPI details for the V1 model

1260 To write the emergent properties of Equation 7 in terms of the EPI optimization, we have

$$f(\mathbf{x}; \mathbf{z}) = s_E(\mathbf{x}; \mathbf{z}), \quad (78)$$

$$\boldsymbol{\mu} = \begin{bmatrix} 5 \end{bmatrix} \quad (79)$$

1262 (or $\boldsymbol{\mu} = \begin{bmatrix} 10 \end{bmatrix}$), and

$$\boldsymbol{\sigma}^2 = \begin{bmatrix} 1^2 \end{bmatrix} \quad (80)$$

1263 (see Sections 5.1.3-5.1.4, and example in Section 5.1.5).

1264 For EPI in Figure 3D-E and Figure 3-figure supplement 1, we used a real NVP architecture with
 1265 three coupling layers and two-layer neural networks of 50 units per layer. The normalizing flow

1266 architecture mapped $z_0 \sim \mathcal{N}(\mathbf{0}, I)$ to a support of $\mathbf{z} = [\sigma_E, \sigma_P, \sigma_S, \sigma_V] \in [0.0, 0.5]^4$. EPI optimization
 1267 was run using three different random seeds for architecture initialization $\boldsymbol{\theta}$ with an augmented
 1268 lagrangian coefficient of $c_0 = 10^{-1}$, $\beta = 2$, a batch size $n = 100$, and simulated 100 trials to
 1269 calculate average $s_E(\mathbf{x}; \mathbf{z})$ for each $\mathbf{z}^{(i)}$. We used $i_{\max} = 2,000$ iterations per epoch. The distribu-
 1270 tions shown are those of the architectures converging with criteria $N_{\text{test}} = 100$ at greatest entropy
 1271 across three random seeds. Optimization details are shown in Figure 3-figure supplement 2. The
 1272 sums of squares of each pair of parameters are shown for each EPI distribution in Figure 3-figure
 1273 supplement 3. The plots are histograms of 500 samples from each EPI distribution from which the
 1274 significance p -values of Section 3.4 are determined.

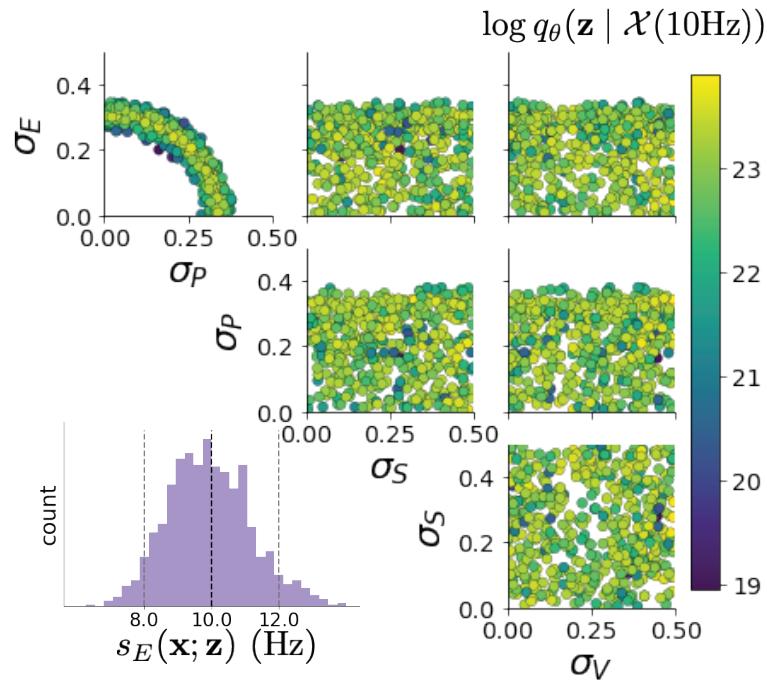


Figure 3-figure supplement 1: EPI inferred distribution for $\mathcal{X}(10\text{Hz})$.

1275 5.4.3 Sensitivity analyses

1276 In Figure 3E, we visualize the modes of $q_{\boldsymbol{\theta}}(\mathbf{z} \mid \mathcal{X})$ throughout the σ_E - σ_P marginal. At each local
 1277 mode $\mathbf{z}^*(\sigma_P)$, where σ_P is fixed, we calculated the Hessian and visualized the sensitivity dimension
 1278 in the direction of positive σ_E .

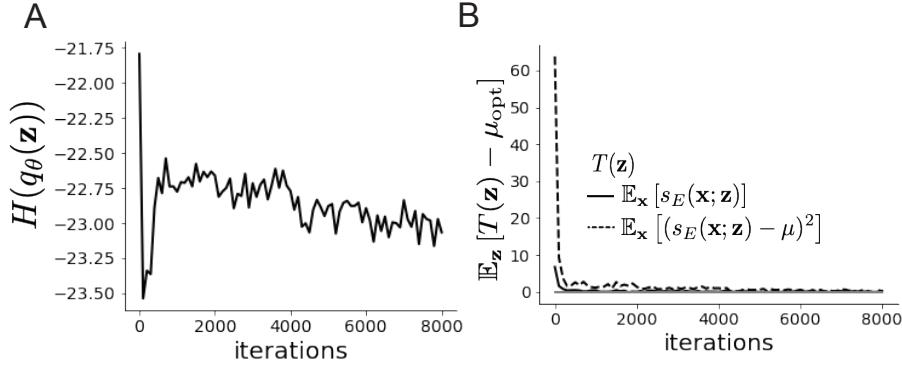


Figure 3-figure supplement 2: EPI optimization $q_\theta(\mathbf{z} | \mathcal{X}(5\text{Hz}))$ **A.** Entropy throughout optimization. **B.** The emergent property statistic means and variances converge to their constraints at 8,000 iterations following the fourth augmented lagrangian epoch.

1279 **5.4.4 Testing for the paradoxical effect**

1280 The paradoxical effect occurs when a populations steady state rate is decreased (or increased)
1281 when an increase (decrease) in current is applied to that population [12]. To see which, if any,
1282 populations exhibited a paradoxical effect, we examined responses to changes in input (Figure 3-
1283 figure supplement 4). Input magnitudes were chosen so that the effect is salient (0.002 for E and P,
1284 but 0.02 for S and V). Only the P-population exhibited the paradoxical effect at this connectivity
1285 W and input \mathbf{h} .

1286 **5.4.5 Primary visual cortex: Mathematical intuition and challenges**

1287 The dynamical system that we are working with can be written as

$$\begin{aligned} dx &= \frac{1}{\tau}(-x + f(Wx + h + \epsilon))dt \\ d\epsilon &= -\frac{dt}{\tau_{\text{noise}}} \epsilon + \frac{\sqrt{2}}{\sqrt{\tau_{\text{noise}}}} \Sigma_\epsilon dW \end{aligned} \tag{81}$$

1288 Where in this paper we chose

$$\Sigma_\epsilon = \tau_{\text{noise}} \begin{bmatrix} \tilde{\sigma}_E & 0 & 0 & 0 \\ 0 & \tilde{\sigma}_P & 0 & 0 \\ 0 & 0 & \tilde{\sigma}_S & 0 \\ 0 & 0 & 0 & \tilde{\sigma}_V \end{bmatrix} \tag{82}$$

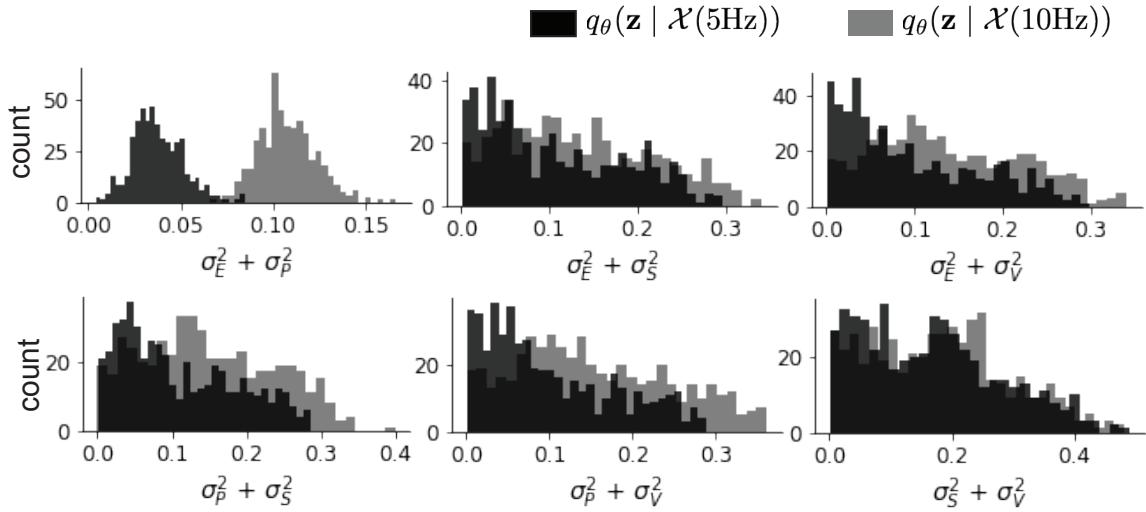


Figure 3-figure supplement 3: EPI predictive distributions of the sum of squares of each pair of noise parameters.

1289 where $\tilde{\sigma}_\alpha$ is the reparameterized standard deviation of the noise for population α from Equation
 1290 70.

1291 In order to compute this covariance, we define $v = \omega x + h + \epsilon$ and $S = I - \omega f'(v)$, to re-write Eq.
 1292 (81) as an 8-dimensional system:

$$d \begin{pmatrix} \delta v \\ \epsilon \end{pmatrix} = - \begin{pmatrix} S & -\frac{\tau_{\text{noise}} - \tau}{\tau \tau_{\text{noise}}} I \\ 0 & \frac{1}{\tau_{\text{noise}}} I \end{pmatrix} \begin{pmatrix} \delta v \\ \epsilon \end{pmatrix} dt + \begin{pmatrix} 0 & \frac{\sqrt{2}}{\sqrt{\tau_{\text{noise}}}} \Sigma_\epsilon \\ 0 & \frac{\sqrt{2}}{\sqrt{\tau_{\text{noise}}}} \Sigma_\epsilon \end{pmatrix} d\mathbf{W} \quad (83)$$

1293 Where $d\mathbf{W}$ is a vector with the private noise of each variable. The $d\mathbf{W}$ term is multiplied by a
 1294 non-diagonal matrix is because the noise that the voltage receives is the exact same than the one
 1295 that comes from the OU process and not another process. The solution of this problem is given by
 1296 the Lyapunov Equation [59, 66]:

$$\begin{pmatrix} S & -\frac{\tau_{\text{noise}} - \tau}{\tau \tau_{\text{noise}}} I \\ 0 & \frac{1}{\tau_{\text{noise}}} I \end{pmatrix} \begin{pmatrix} \Lambda_v & \Lambda_c \\ \Lambda_c^T & \Lambda_\epsilon \end{pmatrix} + \begin{pmatrix} \Lambda_v & \Lambda_c \\ \Lambda_c^T & \Lambda_\epsilon \end{pmatrix} \begin{pmatrix} S^T & 0 \\ -\frac{\tau_{\text{noise}} - \tau}{\tau \tau_{\text{noise}}} I & \frac{1}{\tau_{\text{noise}}} I \end{pmatrix} = \begin{pmatrix} \frac{2}{\tau_{\text{noise}}} \Lambda_\epsilon & \frac{2}{\tau_{\text{noise}}} \Lambda_\epsilon \\ \frac{2}{\tau_{\text{noise}}} \Lambda_\epsilon & \frac{2}{\tau_{\text{noise}}} \Lambda_\epsilon \end{pmatrix} \quad (84)$$

1297 To obtain an equation for Λ_v , we solve this block matrix multiplication:

$$S \Lambda_v + \Lambda_v S^T = \frac{2 \Lambda_\epsilon}{\tau_{\text{noise}}} + \frac{\tau_{\text{noise}}^2 - \tau^2}{(\tau \tau_{\text{noise}})^2} \left(\left(\frac{1}{\tau_{\text{noise}}} I + S \right)^{-1} \Lambda_\epsilon + \Lambda_\epsilon \left(\frac{1}{\tau_{\text{noise}}} I + S^T \right)^{-1} \right) \quad (85)$$

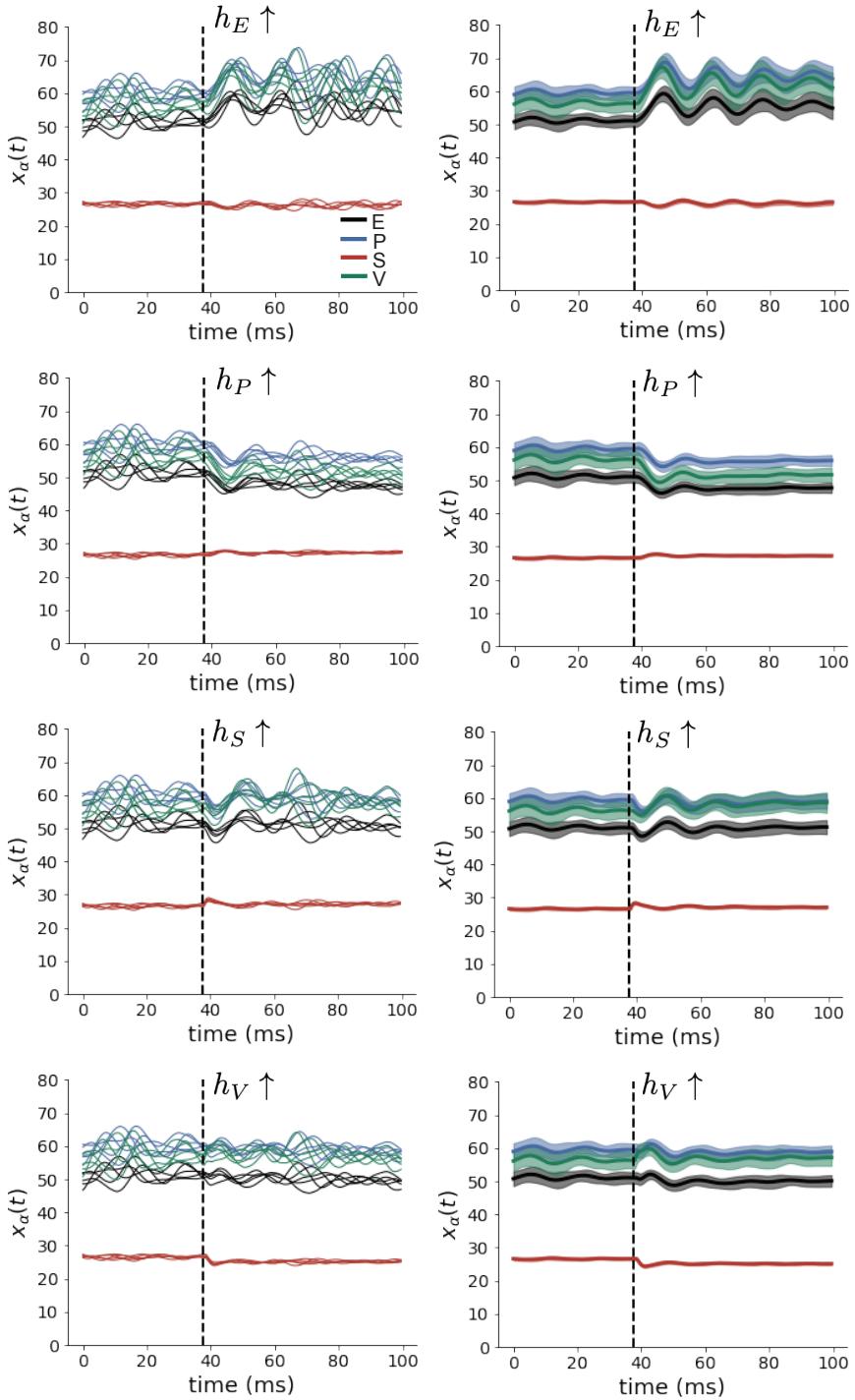


Figure 3-figure supplement 4: (Left) SSSN simulations for small increases in neuron-type population input. (Right) Average (solid) and standard deviation (shaded) of stochastic fluctuations of responses.

Which is another Lyapunov Equation, now in 4 dimensions. In the simplest case in which $\tau_{\text{noise}} = \tau$, the voltage is directly driven by white noise, and Λ_v can be expressed in powers of S and S^T . Because S satisfies its own polynomial equation (Cayley Hamilton theorem), there will be 4 coefficients for the expansion of S and 4 for S^T , resulting in 16 coefficients that define Λ_v for a given S . Due to symmetry arguments [66], in this case the diagonal elements of the covariance matrix of the voltage will have the form:

$$\Lambda_{v_{ii}} = \sum_{i=\{E,P,S,V\}} g_i(S) \sigma_{ii}^2 \quad (86)$$

1298 These coefficients $g_i(S)$ are complicated functions of the Jacobian of the system. Although expres-
 1299 sions for these coefficients can be found explicitly, only numerical evaluation of those expressions
 1300 determine which components of the noisy input are going to strongly influence the variability of ex-
 1301 citatory population. Showing the generality of this dependence in more complicated noise scenarios
 1302 (e.g. $\tau_{\text{noise}} > \tau$ as in Section 3.4), is the focus of current research.

1303 5.5 Superior colliculus

1304 5.5.1 SC model

1305 The ability to switch between two separate tasks throughout randomly interleaved trials, or “rapid
 1306 task switching,” has been studied in rats, and midbrain superior colliculus (SC) has been show to
 1307 play an important in this computation [67]. Neural recordings in SC exhibited two populations of
 1308 neurons that simultaneously represented both task context (Pro or Anti) and motor response (con-
 1309 tralateral or ipsilateral to the recorded side), which led to the distinction of two functional classes:
 1310 the Pro/Contra and Anti/Ipsi neurons [48]. Given this evidence, Duan et al. proposed a model
 1311 with four functionally-defined neuron-type populations: two in each hemisphere corresponding to
 1312 the Pro/Contra and Anti/Ipsi populations. We study how the connectivity of this neural circuit
 1313 governs rapid task switching ability.

1314 The four populations of this model are denoted as left Pro (LP), left Anti (LA), right Pro (RP)
 1315 and right Anti (RA). Each unit has an activity (x_α) and internal variable (u_α) related by

$$x_\alpha = \phi(u_\alpha) = \left(\frac{1}{2} \tanh \left(\frac{u_\alpha - a}{b} \right) + \frac{1}{2} \right), \quad (87)$$

1316 where $\alpha \in \{LP, LA, RA, RP\}$, $a = 0.05$ and $b = 0.5$ control the position and shape of the nonlin-

1317 earity. We order the neural populations of x and u in the following manner

$$\mathbf{x} = \begin{bmatrix} x_{LP} \\ x_{LA} \\ x_{RP} \\ x_{RA} \end{bmatrix} \quad \mathbf{u} = \begin{bmatrix} u_{LP} \\ u_{LA} \\ u_{RP} \\ u_{RA} \end{bmatrix}, \quad (88)$$

1318 which evolve according to

$$\tau \frac{d\mathbf{u}}{dt} = -\mathbf{u} + W\mathbf{x} + \mathbf{h} + d\mathbf{B}. \quad (89)$$

1319 with time constant $\tau = 0.09s$, step size 24ms and Gaussian noise $d\mathbf{B}$ of variance 0.2^2 . These
1320 hyperparameter values are motivated by modeling choices and results from [48].

1321 The weight matrix has 4 parameters for self sW , vertical vW , horizontal hW , and diagonal dW
1322 connections:

$$W = \begin{bmatrix} sW & vW & hW & dW \\ vW & sW & dW & hW \\ hW & dW & sW & vW \\ dW & hW & vW & sW \end{bmatrix}. \quad (90)$$

1323 We study the role of parameters $\mathbf{z} = [sW, vW, hW, dW]^\top$ in rapid task switching.

1324 The circuit receives four different inputs throughout each trial, which has a total length of 1.8s.

$$\mathbf{h} = \mathbf{h}_{\text{constant}} + \mathbf{h}_{\text{P,bias}} + \mathbf{h}_{\text{rule}} + \mathbf{h}_{\text{choice-period}} + \mathbf{h}_{\text{light}}. \quad (91)$$

1325 There is a constant input to every population,

$$\mathbf{h}_{\text{constant}} = I_{\text{constant}}[1, 1, 1, 1]^\top, \quad (92)$$

1326 a bias to the Pro populations

$$\mathbf{h}_{\text{P,bias}} = I_{\text{P,bias}}[1, 0, 1, 0]^\top, \quad (93)$$

1327 rule-based input depending on the condition

$$\mathbf{h}_{\text{P,rule}}(t) = \begin{cases} I_{\text{P,rule}}[1, 0, 1, 0]^\top, & \text{if } t \leq 1.2s \\ 0, & \text{otherwise} \end{cases} \quad (94)$$

1328

$$\mathbf{h}_{\text{A,rule}}(t) = \begin{cases} I_{\text{A,rule}}[0, 1, 0, 1]^\top, & \text{if } t \leq 1.2s \\ 0, & \text{otherwise} \end{cases}, \quad (95)$$

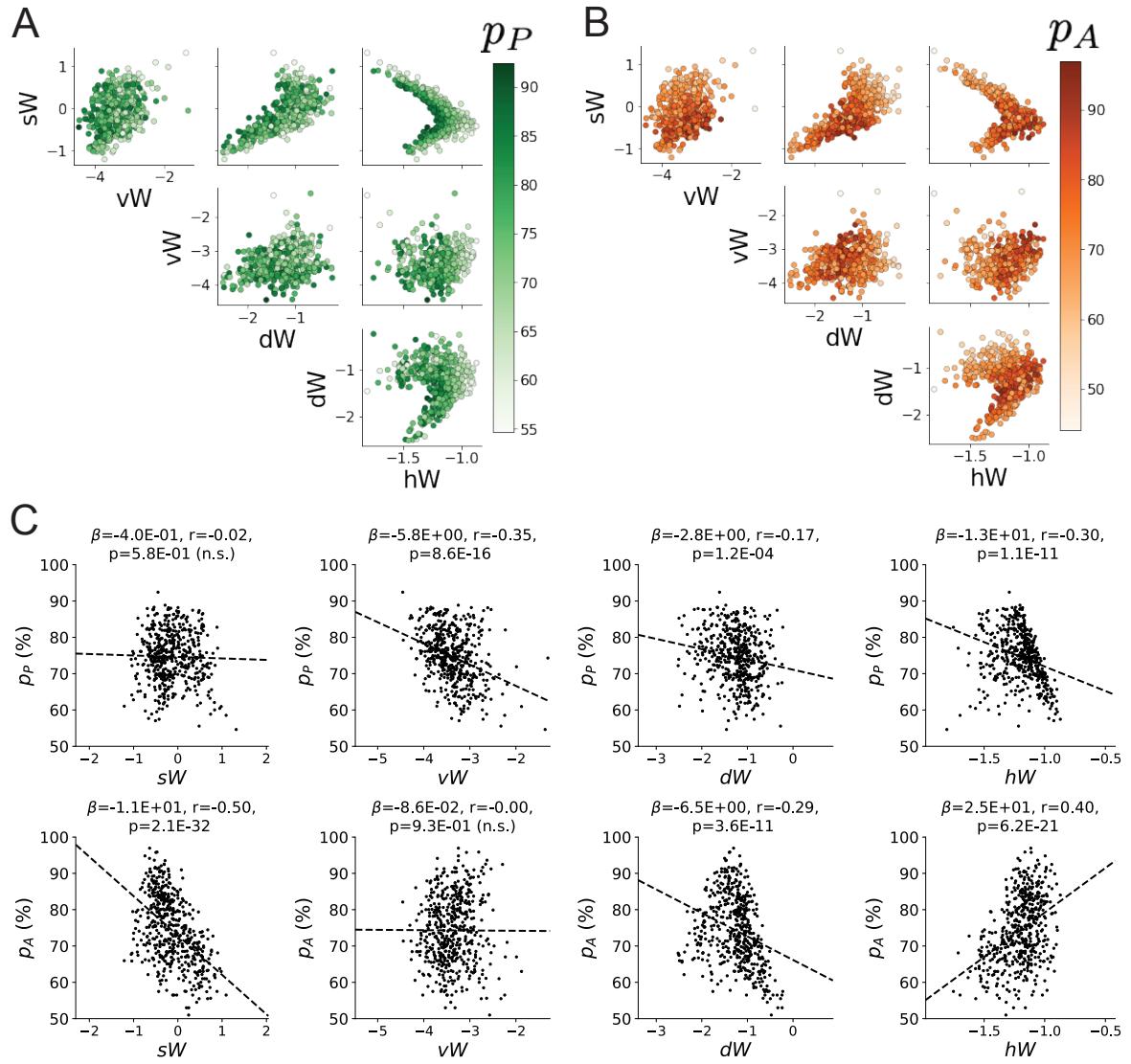


Figure 4-figure supplement 1: **A.** Same pairplot as Figure 4C colored by Pro task accuracy. **B.** Same as A colored by Anti task accuracy. **C.** Connectivity parameters of EPI distributions versus task accuracies. β is slope coefficient of linear regression, r is correlation, and p is the two-tailed p-value.

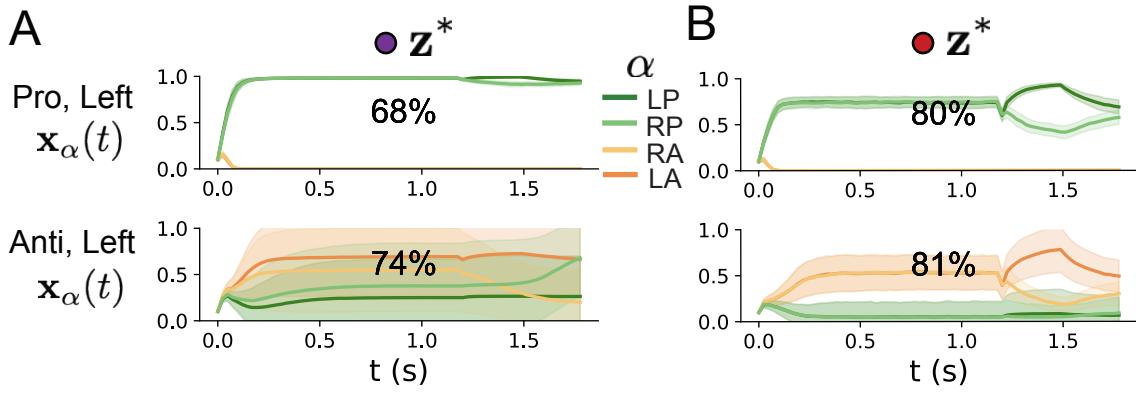


Figure 4-figure supplement 2: **A.** Simulations in network regime 1: $\mathbf{z}^*(sW = -0.75)$. **B.** Simulations in network regime 2: $\mathbf{z}^*(sW = 0.75)$.

1329 a choice-period input

$$\mathbf{h}_{\text{choice}}(t) = \begin{cases} I_{\text{choice}}[1, 1, 1, 1]^\top, & \text{if } t > 1.2s \\ 0, & \text{otherwise} \end{cases}, \quad (96)$$

1330 and an input to the right or left-side depending on where the light stimulus is delivered

$$\mathbf{h}_{\text{light}}(t) = \begin{cases} I_{\text{light}}[1, 1, 0, 0]^\top, & \text{if } 1.2s < t < 1.5s \text{ and Left} \\ I_{\text{light}}[0, 0, 1, 1]^\top, & \text{if } 1.2s < t < 1.5s \text{ and Right} \\ 0, & \text{otherwise} \end{cases}. \quad (97)$$

1331 The input parameterization was fixed to $I_{\text{constant}} = 0.75$, $I_{\text{P,bias}} = 0.5$, $I_{\text{P,rule}} = 0.6$, $I_{\text{A,rule}} = 0.6$,
1332 $I_{\text{choice}} = 0.25$, and $I_{\text{light}} = 0.5$.

1333 5.5.2 Task accuracy calculation

1334 The accuracies of the Pro and Anti tasks are calculated as

$$p_P(\mathbf{x}; \mathbf{z}) = \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x}|\mathbf{z})} [d_P(\mathbf{x}; \mathbf{z})] \quad (98)$$

1335 and

$$p_A(\mathbf{x}; \mathbf{z}) = \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x}|\mathbf{z})} [d_A(\mathbf{x}; \mathbf{z})] \quad (99)$$

1336 where $d_P(\mathbf{x}; \mathbf{z})$ and $d_A(\mathbf{x}; \mathbf{z})$ calculate the decision made in each trial (approximately 1 for correct
1337 and 0 for incorrect choices). Specifically,

$$d_P(\mathbf{x}; \mathbf{z}) = \Theta[x_{LP}(t = 1.8s) - x_{RP}(t = 1.8s)] \quad (100)$$

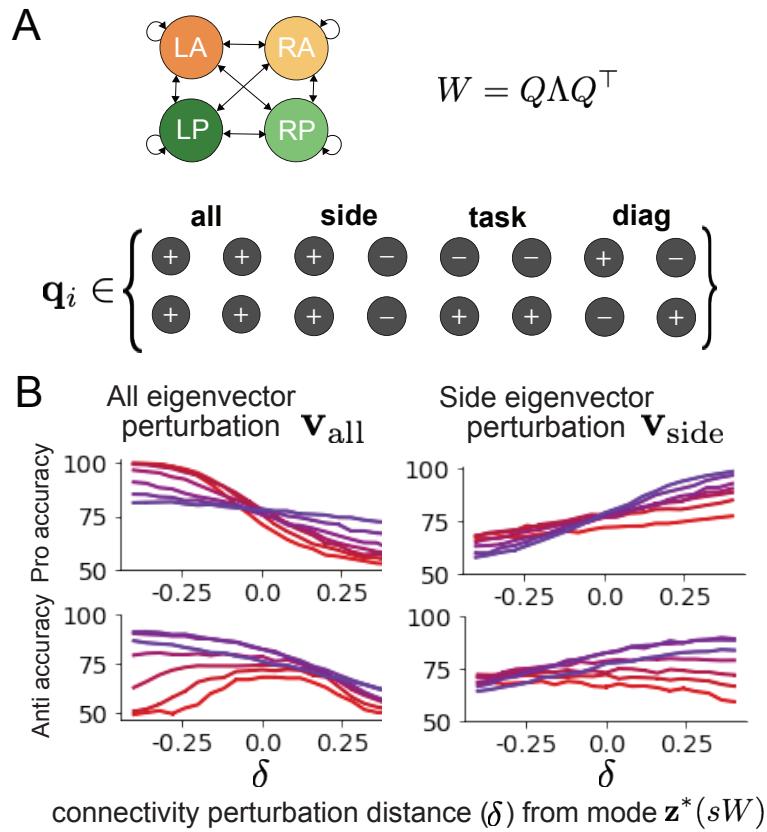


Figure 4-figure supplement 3: **A.** Invariant eigenvectors of connectivity matrix W . **B.** Accuracies for connectivity perturbations when changing λ_{all} and λ_{side} (λ_{task} and λ_{diag} shown in Figure 4D).

1338 in Pro trials where the stimulus is on the left side, and Θ approximates the Heaviside step function.
 1339 Similarly,

$$d_A(\mathbf{x}; \mathbf{z}) = \Theta[x_{RP}(t = 1.8s) - x_{LP}(t = 1.8s)] \quad (101)$$

1340 in Anti trials where the stimulus was on the left side. Our accuracy calculation only considers one
 1341 stimulus presentation (Left), since the model is left-right symmetric. The accuracy is averaged over
 1342 200 independent trials, and the Heaviside step function is approximated as

$$\Theta(\mathbf{x}) = \text{sigmoid}(\beta_\Theta \mathbf{x}), \quad (102)$$

1343 where $\beta_\Theta = 100$.

1344 **5.5.3 EPI details for the SC model**

1345 To write the emergent properties of Equation 9 in terms of the EPI optimization, we have

$$f(\mathbf{x}; \mathbf{z}) = \begin{bmatrix} d_P(\mathbf{x}; \mathbf{z}) \\ d_A(\mathbf{x}; \mathbf{z}) \end{bmatrix} \quad (103)$$

1346

$$\boldsymbol{\mu} = \begin{bmatrix} .75 \\ .75 \end{bmatrix}, \quad (104)$$

1347 and

$$\boldsymbol{\sigma}^2 = \begin{bmatrix} .075^2 \\ .075^2 \end{bmatrix} \quad (105)$$

1348 (see Sections 5.1.3-5.1.4, and example in Section 5.1.5).

1349 Throughout optimization, the augmented lagrangian parameters η and c , were updated after each
 1350 epoch of $i_{\max} = 2,000$ iterations (see Section 5.1.4). The optimization converged after ten epochs
 1351 (Figure 4-figure supplement 4).

1352 For EPI in Figure 4C, we used a real NVP architecture with three coupling layers of affine transfor-
 1353 mations parameterized by two-layer neural networks of 50 units per layer. The initial distribution
 1354 was a standard isotropic gaussian $\mathbf{z}_0 \sim \mathcal{N}(\mathbf{0}, I)$ mapped to a support of $\mathbf{z}_i \in [-5, 5]$. We used an
 1355 augmented lagrangian coefficient of $c_0 = 10^2$, a batch size $n = 100$, and $\beta = 2$. The distribution
 1356 was the greatest EPI distribution to converge across 5 random seeds with criteria $N_{\text{test}} = 25$.

1357 The bend in the EPI distribution is not a spurious result of the EPI optimization. The structure
 1358 discovered by EPI matches the shape of the set of points returned from brute-force random sampling
 1359 (Figure 4-figure supplement 5A) These connectivities were sampled from a uniform distribution over

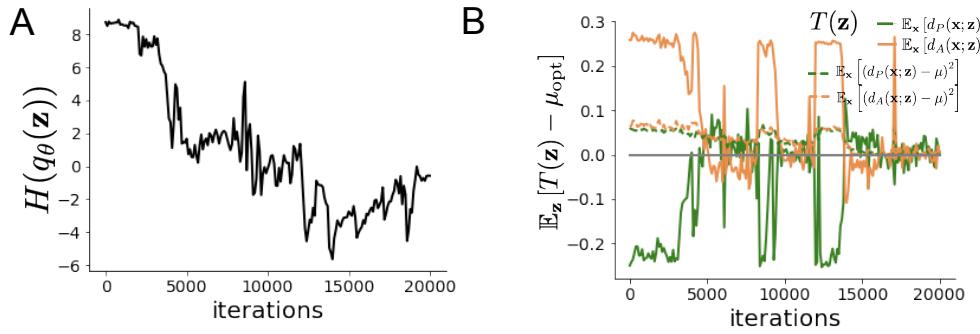


Figure 4-figure supplement 4: EPI optimization of the SC model producing rapid task switching.

A. Entropy throughout optimization. **B.** The emergent property statistic means and variances converge to their constraints at 20,000 iterations following the tenth augmented lagrangian epoch.

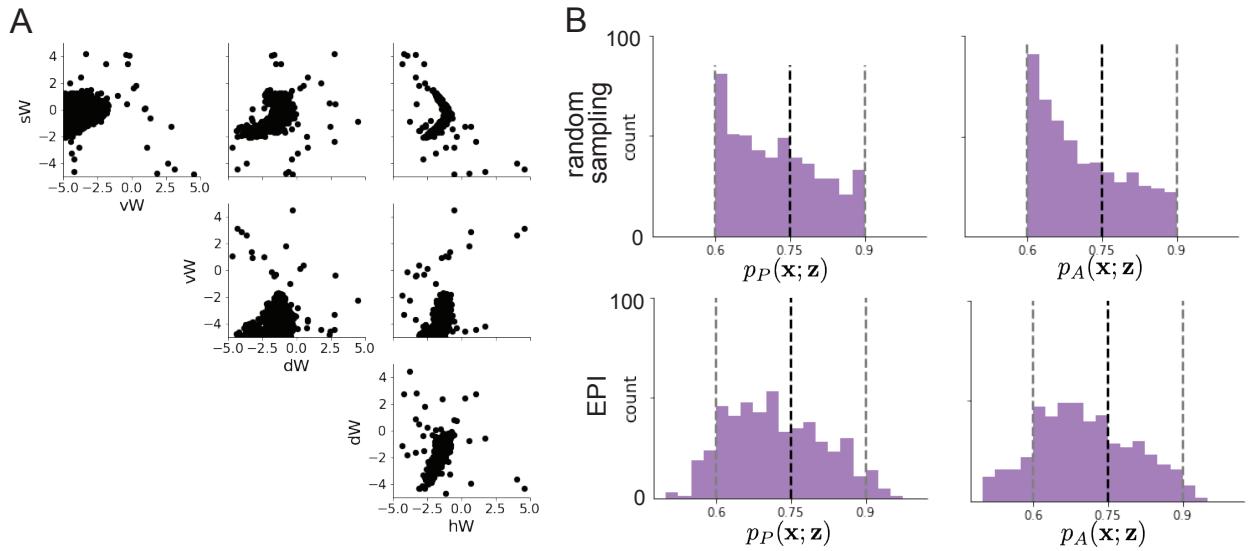


Figure 4-figure supplement 5: **A.** Rapid task switching SC connectivities obtained from random sampling. **B.** Task accuracies of the inferred distributions from random sampling (top) and EPI (bottom).

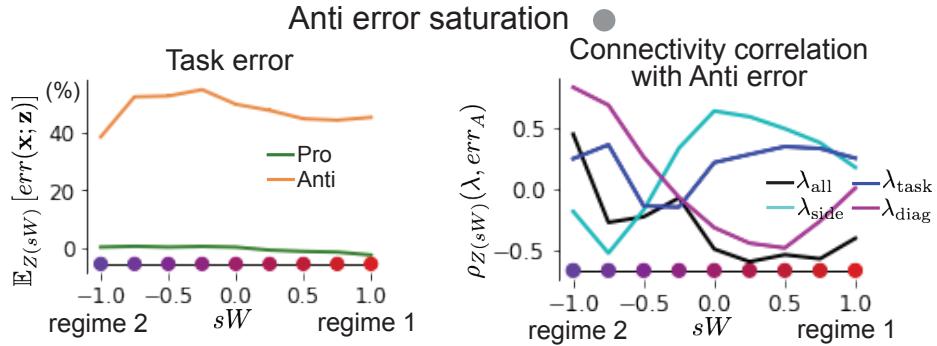


Figure 5-figure supplement 1: (Left) Mean and standard error of Pro and Anti error from regime 1 to regime 2 at $\gamma = 0.85$. (Right) Correlations of connectivity eigenvalues with Anti error from regime 1 to regime 2 at $\gamma = 0.85$.

the range of each connectivity parameter, and all parameters producing accuracy in each task within the range of 60% to 90% were kept. This set of connectivities will not match the distribution of EPI exactly, since it is not conditioned on the emergent property. For example the parameter set returned by the brute-force search is biased towards lower accuracies (Figure 4-figure supplement 5B).

5.5.4 Mode identification with EPI

We found one mode of the EPI distribution for fixed values of sW from 1 to -1 in steps of 0.25. To begin, we chose an initial parameter value from 500 parameter samples $\mathbf{z} \sim q_\theta(\mathbf{z} \mid \mathcal{X})$ that had closest sW value to 1. We then optimized this estimate of the mode (for fixed sW) using probability gradients of the deep probability distribution for 500 steps of gradient ascent with a learning rate of 5×10^{-3} . The next mode (at $sW = 0.75$) was found using the previous mode as the initialization. This and all subsequent optimizations used 200 steps of gradient ascent with a learning rate of 1×10^{-3} , except at $sW = -1$ where a learning rate of 5×10^{-4} was used. During all mode identification optimizations, the learning rate was reduced by half (decay = 0.5) after every 100 iterations.

5.5.5 Sample grouping by mode

For the analyses in Figure 5C and Figure 5-figure supplement 1, we obtained parameters for each step along the continuum between regimes 1 and 2 by sampling from the EPI distribution. Each

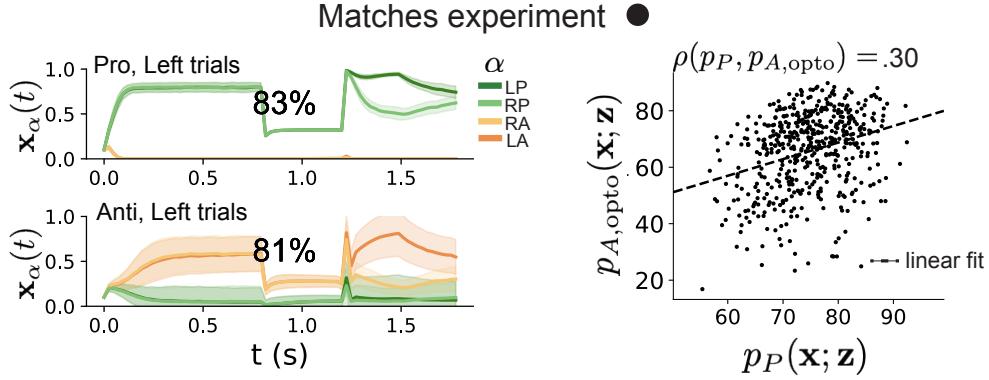


Figure 5-figure supplement 2: (Left) Mean and standard deviation (shading) of responses of the SC model at the mode of the EPI distribution to delay period inactivation at $\gamma = 0.675$. Accuracy in Pro (top) and Anti (bottom) task is shown as a percentage. (Right) Anti accuracy following delay period inactivation at $\gamma = 0.675$ versus accuracy in the Pro task across connectivities in the EPI distribution.

sample was assigned to the closest mode $\mathbf{z}^*(sW)$. Sampling continued until 500 samples were assigned to each mode, which took 2.67 seconds (5.34ms/sample-per-mode). It took 9.59 minutes to obtain just 5 samples for each mode with brute force sampling requiring accuracies between 60% and 90% in each task (115s/sample-per-mode). This corresponds to a sampling speed increase of roughly 21,500 once the EPI distribution has been learned.

5.5.6 Sensitivity analysis

At each mode, we measure the sensitivity dimension (that of most negative eigenvalue in the Hessian of the EPI distribution) $\mathbf{v}_1(\mathbf{z}^*)$. To resolve sign degeneracy in eigenvectors, we chose $\mathbf{v}_1(\mathbf{z}^*)$ to have negative element in hW . This tells us what parameter combination rapid task switching is most sensitive to at this parameter choice in the regime.

5.5.7 Connectivity eigendecomposition and processing modes

To understand the connectivity mechanisms governing task accuracy, we took the eigendecomposition of the connectivity matrices $W = Q\Lambda Q^{-1}$, which results in the same eigenmodes \mathbf{q}_i for all W parameterized by \mathbf{z} (Figure 4-figure supplement 3A). These eigenvectors are always the same, because the connectivity matrix is symmetric and the model also assumes symmetry across hemispheres, but the eigenvalues of connectivity (or degree of eigenmode amplification) change with \mathbf{z} .

1394 These basis vectors have intuitive roles in processing for this task, and are accordingly named the
 1395 *all* eigenmode - all neurons co-fluctuate, *side* eigenmode - one side dominates the other, *task* eigen-
 1396 mode - the Pro or Anti populations dominate the other, and *diag* mode - Pro- and Anti-populations
 1397 of opposite hemispheres dominate the opposite pair. Due to the parametric structure of the connec-
 1398 tivity matrix, the parameters \mathbf{z} are a linear function of the eigenvalues $\boldsymbol{\lambda} = [\lambda_{\text{all}}, \lambda_{\text{side}}, \lambda_{\text{task}}, \lambda_{\text{diag}}]^T$
 1399 associated with these eigenmodes.

$$\mathbf{z} = A\boldsymbol{\lambda} \quad (106)$$

$$A = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{bmatrix}. \quad (107)$$

1400 We are interested in the effect of raising or lowering the amplification of each eigenmode in the
 1401 connectivity matrix by perturbing individual eigenvalues λ . To test this, we calculate the unit
 1402 vector of changes in the connectivity \mathbf{z} that result from a change in the associated eigenvalues
 1403

$$\mathbf{v}_a = \frac{\frac{\partial \mathbf{z}}{\partial \lambda_a}}{\left| \frac{\partial \mathbf{z}}{\partial \lambda_a} \right|_2}, \quad (108)$$

1404 where

$$\frac{\partial \mathbf{z}}{\partial \lambda_a} = A\mathbf{e}_a, \quad (109)$$

1405 and e.g. $\mathbf{e}_{\text{all}} = [1, 0, 0, 0]^T$. So \mathbf{v}_a is the normalized column of A corresponding to eigenmode
 1406 a . The parameter dimension \mathbf{v}_a ($a \in \{\text{all}, \text{side}, \text{task}, \text{and diag}\}$) that increases the eigenvalue of
 1407 connectivity λ_a is \mathbf{z} -invariant (Equation 109) and $\mathbf{v}_a \perp \mathbf{v}_{b \neq a}$. By perturbing \mathbf{z} along \mathbf{v}_a , we
 1408 can examine how model function changes by directly modulating the connectivity amplification of
 1409 specific eigenmodes, which having interpretable roles in processing in each task.

1410 5.5.8 Modeling optogenetic silencing.

1411 We tested whether the inferred SC model connectivities could reproduce experimental effects of
 1412 optogenetic inactivation in rats [48]. During periods of simulated optogenetic inactivation, activity
 1413 was decreased proportional to the optogenetic strength $\gamma \in [0, 1]$

$$x_\alpha = (1 - \gamma)\phi(u_\alpha). \quad (110)$$

1414 Delay period inactivation was from $0.8 < t < 1.2$.