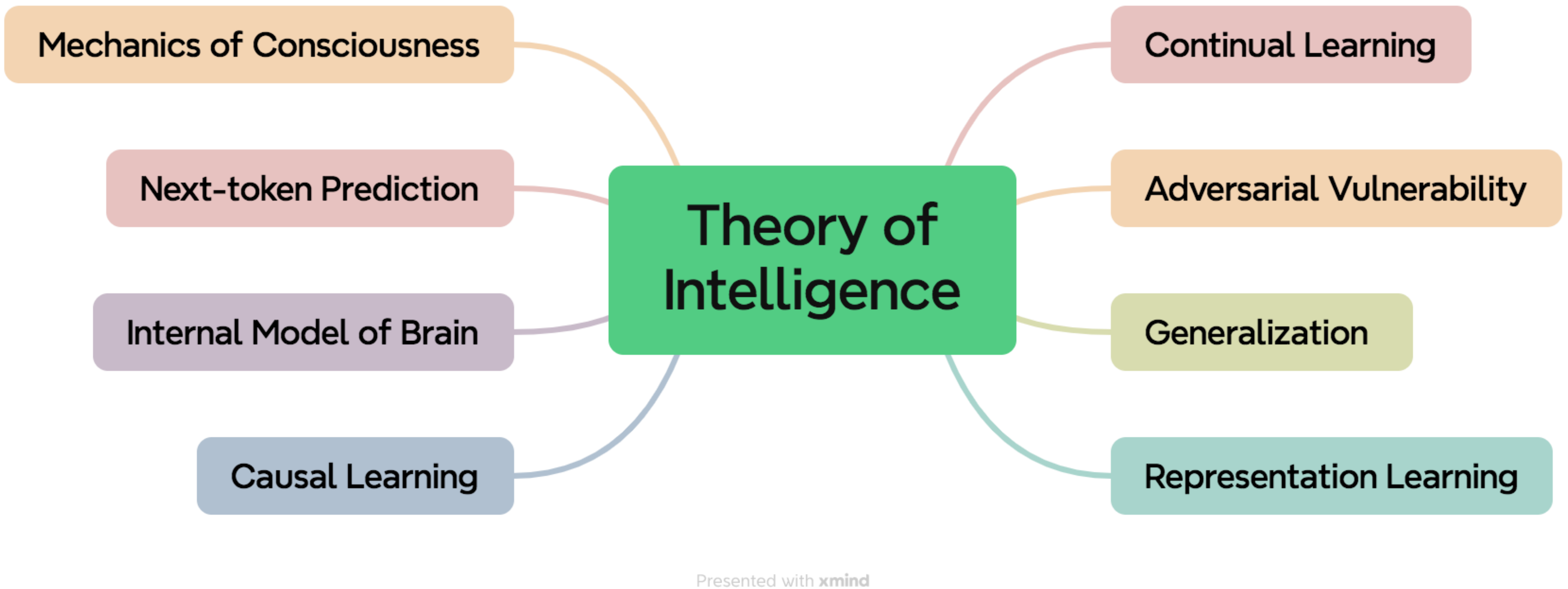
# Order parameters for emergence of consciousness?

Haiping Huang
Sun Yat-sen University
Apr 28, 2024

“Rigor of theoretical physics provides guidelines and testable predictions for realistic applications.”
—Giulio Biroli et.al

“Rigor of theoretical physics provides guidelines and testable predictions for realistic applications.”
—Giulio Biroli et.al

**Rigor helps Clarity**




https://arxiv.org/abs/2306.11232

# Challenges in understanding consciousness

When we perceive, think and act, there is a whir of causation and information processing . . . . . . There is also an internal aspect; there is something it **feels** like to be a cognitive agent. This internal aspect is conscious experience.

Chalmers, D. J. [1997] The Conscious Mind: In Search of a Fundamental Theory

**Subjective experience relates to feelings associated with stimuli**

**Arxiv: 2401.13690**

**“大脑用行动来检验其假设”**
**Buzsáki，The brain from inside out**

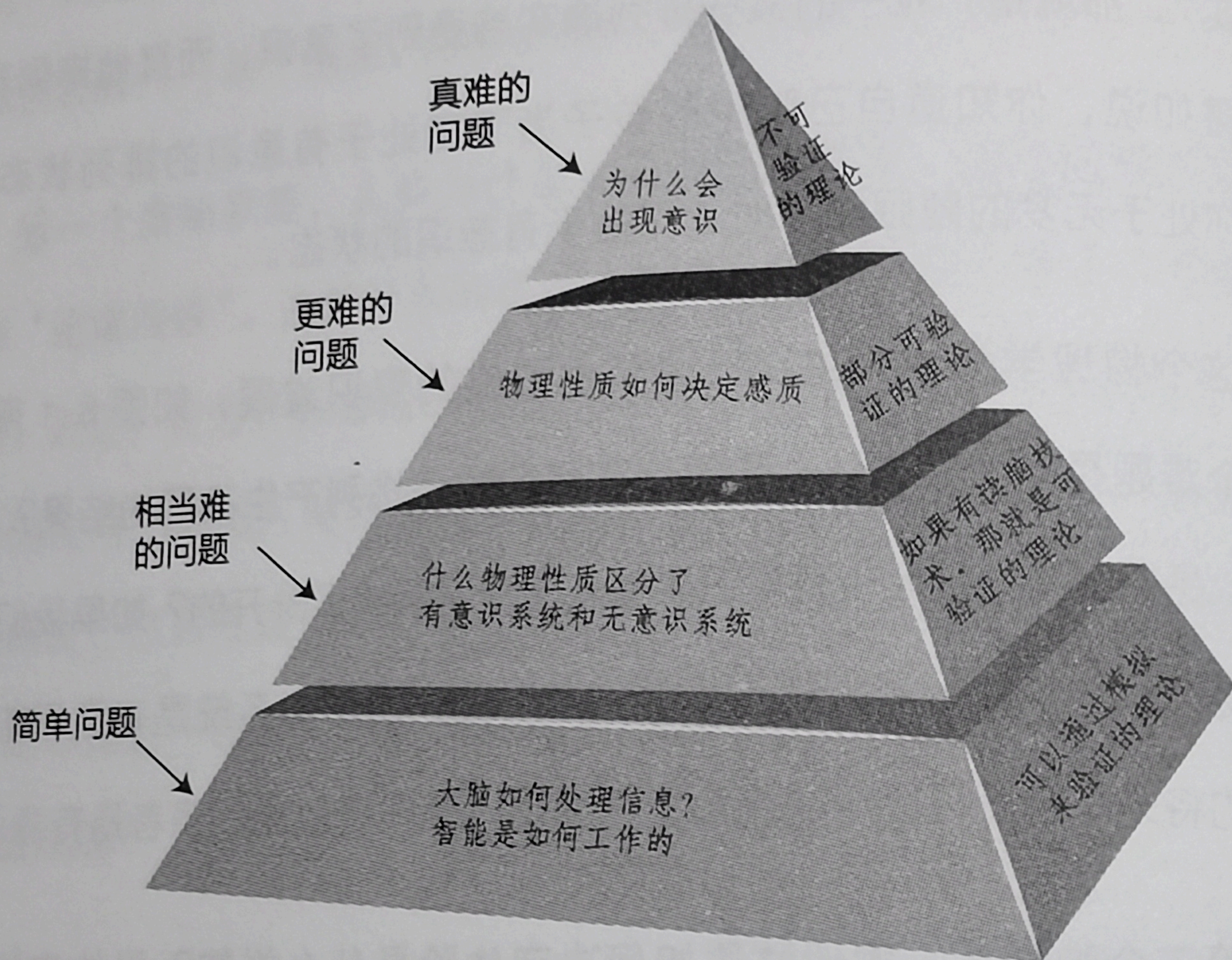
Science typically studies the third person aspect, you know, I can take a brain of a mouse and poke it, put it in the scanner and record from the individual neurons, etc. But consciousness is about how the system feels from the inside. So that's always been the big, the big challenge.

Consciousness, which we can define as experience -- your experience both of yourself and of the world around you -- is one of the hardest to pin down

Christof Koch

How neural interactions support conscious experience?

deep sleep, anesthesia, death




信息处理：
有意识（高级）
（注意力）
无意识（简单）

某种对称性破缺?

Life 3.0

图 8-1 三个彼此独立的意识难题

注：对心智的理解涉及几个层次的问题。大卫·查尔默斯所谓的“简单问题”可以不提到主观体验。一些但不是全部物理系统是有意识的，这个事实提出了三个不同的问题。如果有一个理论可以回答“相当难的问题”，那它就可以用实验来检验。如果检验成功的话，我们就可以以它为基础来解决上层那些更棘手的问题。

# Consciousness is supported by near-critical slow cortical electrodynamics

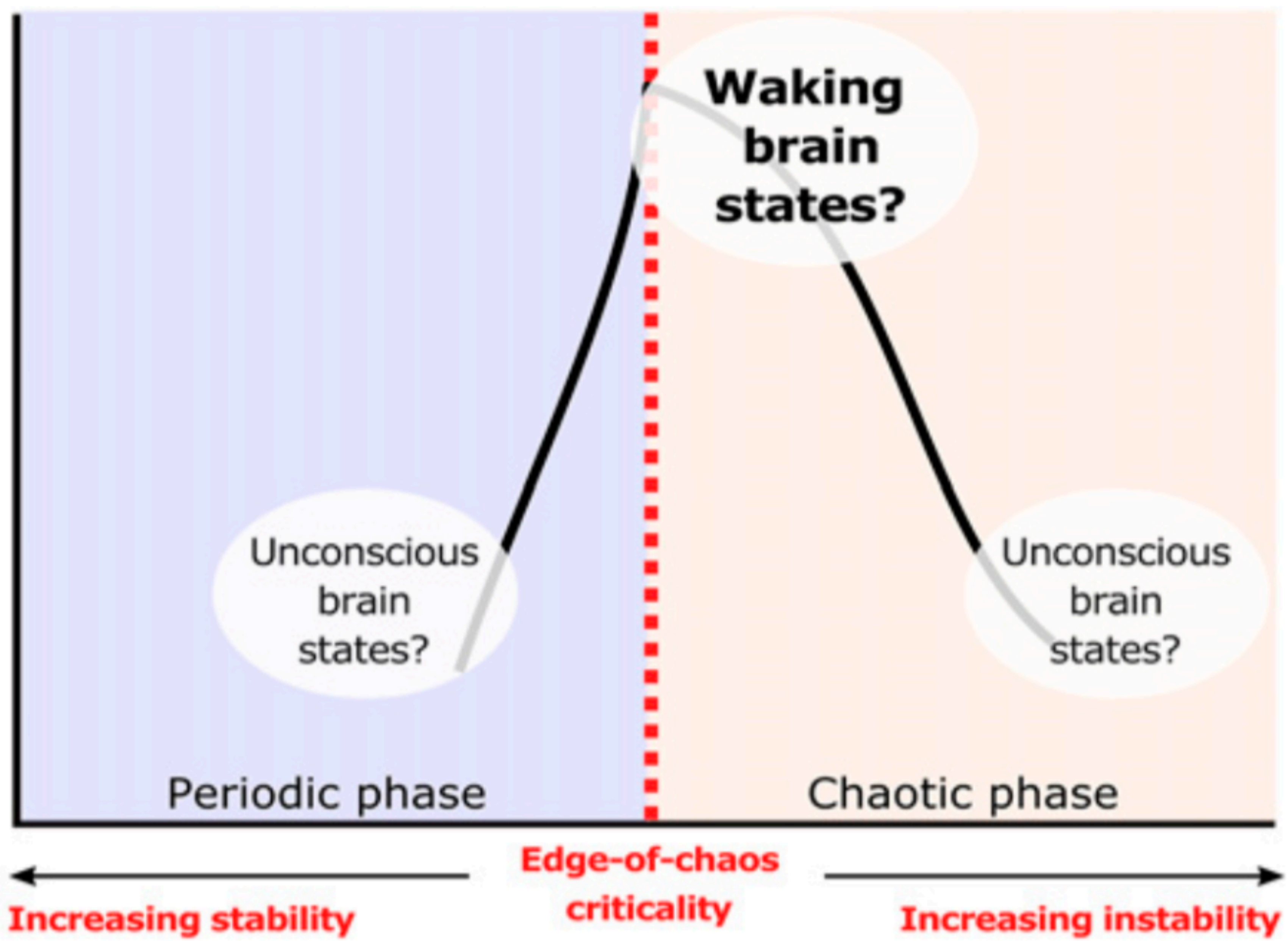
PNAS, 2022

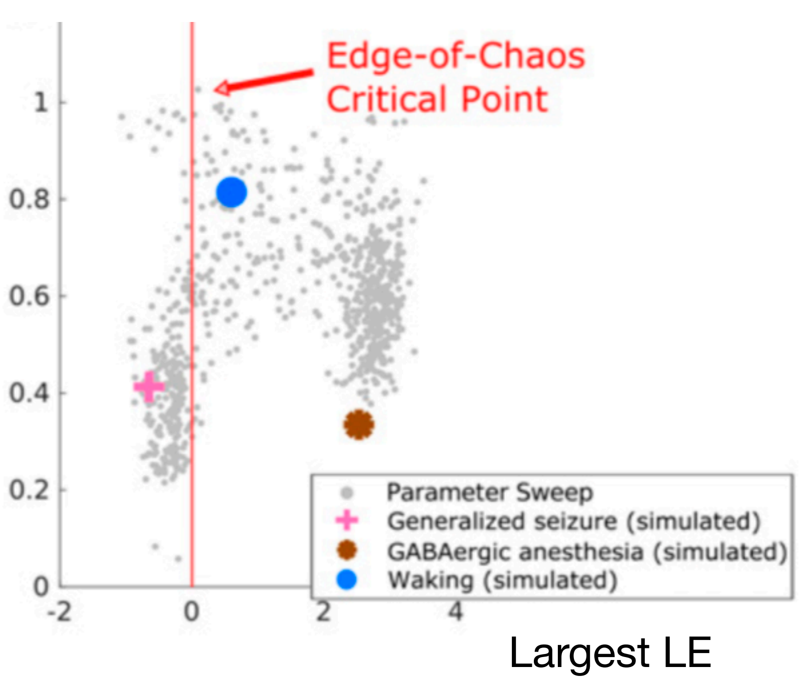
Daniel Toker[a,1], Ioannis Pappas[b,c,d], Janna D. Lendner[b,e], Joel Frohlich[a], Diego M. Mateos[f,g,h], Suresh Muthukumaraswamy[i], Robin Carhart-Harris[j,k], Michelle Paff[l], Paul M. Vespa[m], Martin M. Monti[a,m], Friedrich T. Sommer[b,n], Robert T. Knight[b,c], and Mark D'Esposito[b,c]

**Lempel–Ziv Complexity**

The amount of non-redundant information in a signal

Information-richness of cortical electrodynamics



## Metaphor (toy model) of brain dynamics

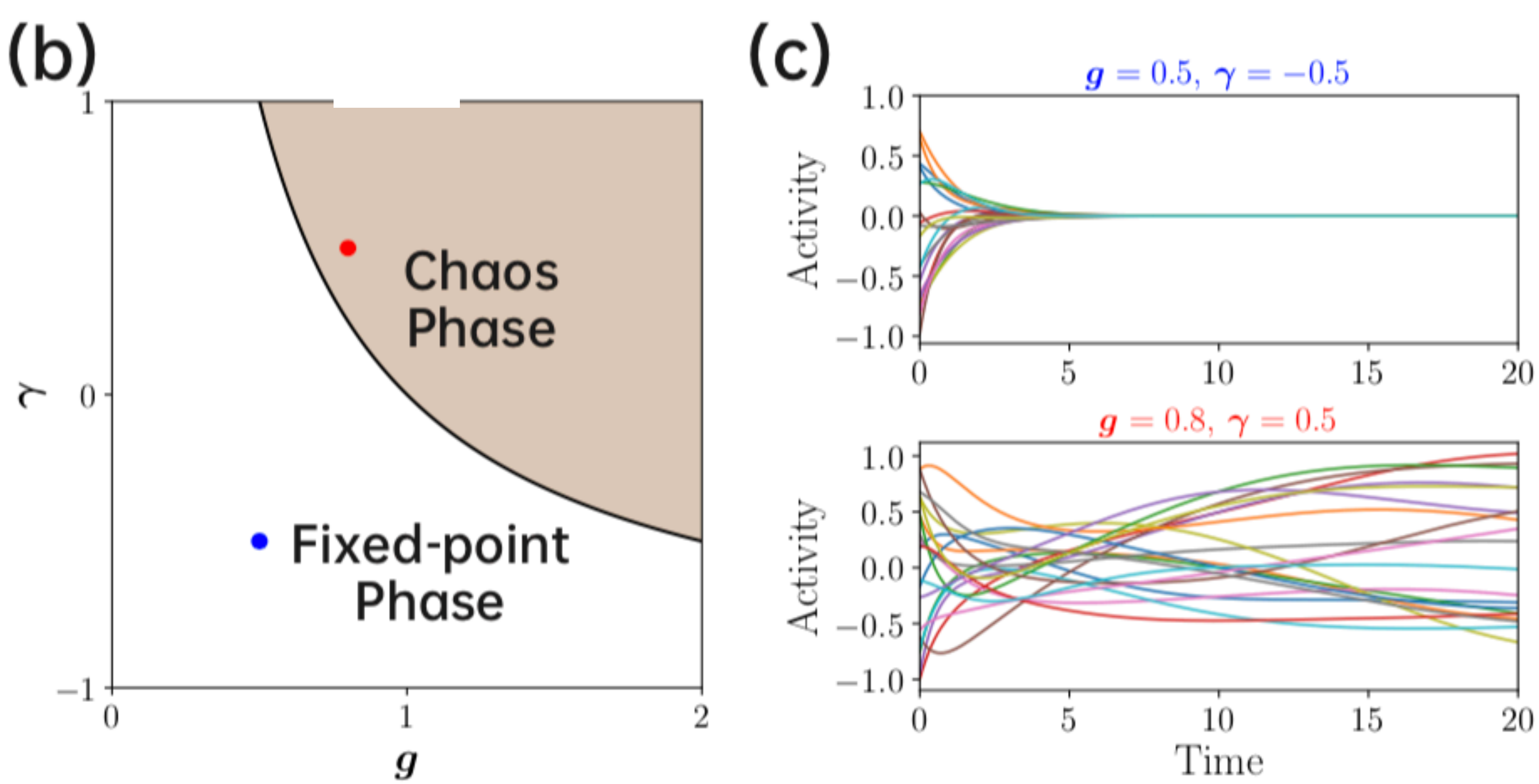
**Synaptic currents**

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -x_i + \sum_{j=1}^{N} J_{ij}\phi(x_j),$$

**Connectivity**

$$\left\langle (J_{ij})^2 \right\rangle = \frac{g^2}{N},$$
$$\langle J_{ij}J_{ji} \rangle = \frac{g^2}{N}\gamma,$$

**Characteristics:**
**Asymmetric, correlated synapses, non-linear dynamics [Non-gradient dynamics]**




Daniel Marti et.a., PRE 18'

# Equilibrium limit

$$\frac{dx_i(t)}{dt} = -x_i(t) + g\sum_{j=1}^{N} J_{ij}x_j(t) + \sigma\xi_i(t), \qquad J_{ij} = J_{ji}$$

$$\mathcal{H}(\boldsymbol{x}) = -\frac{1}{2}\sum_i x_i^2 - \frac{1}{2}g\sum_{i\neq j} J_{ij}x_ix_j, \qquad F = -\nabla_x \mathcal{H}(x)$$

**Symmetric, Linear dynamics:** $$P(\boldsymbol{x}) \sim \exp\left(-\frac{\mathcal{H}(\boldsymbol{x})}{T}\right), \quad T = \sigma^2/2$$

Lyapunov Function decreasing over dynamics

arxiv: 2305.08459

# More complex cases

$$\frac{d\mathbf{x}}{dt} = F(\mathbf{x}) + \zeta$$

**Probability conservation**

$$\partial P / \partial t + \nabla \cdot \mathbf{J} = 0,$$

**Non-equilibrium potential**

**Steady state:** $P_{ss}(x) = e^{-U(\mathbf{x})} \longrightarrow \mathbf{J}_{ss} = \mathbf{F}P_{ss} - \nabla \cdot (\mathbf{D}P_{ss})$

**Curl force**

$$\mathbf{F} = \mathbf{J}_{ss}/P_{ss} - \mathbf{D}\nabla U + \nabla \cdot \mathbf{D}$$

**Divergent free flux:** $\nabla \cdot \mathbf{J}_{ss} = 0$

**Diffusion matrix of stochastic noise e.g., background activity**

But: hard to get U(x), and steady state probability generally unknown

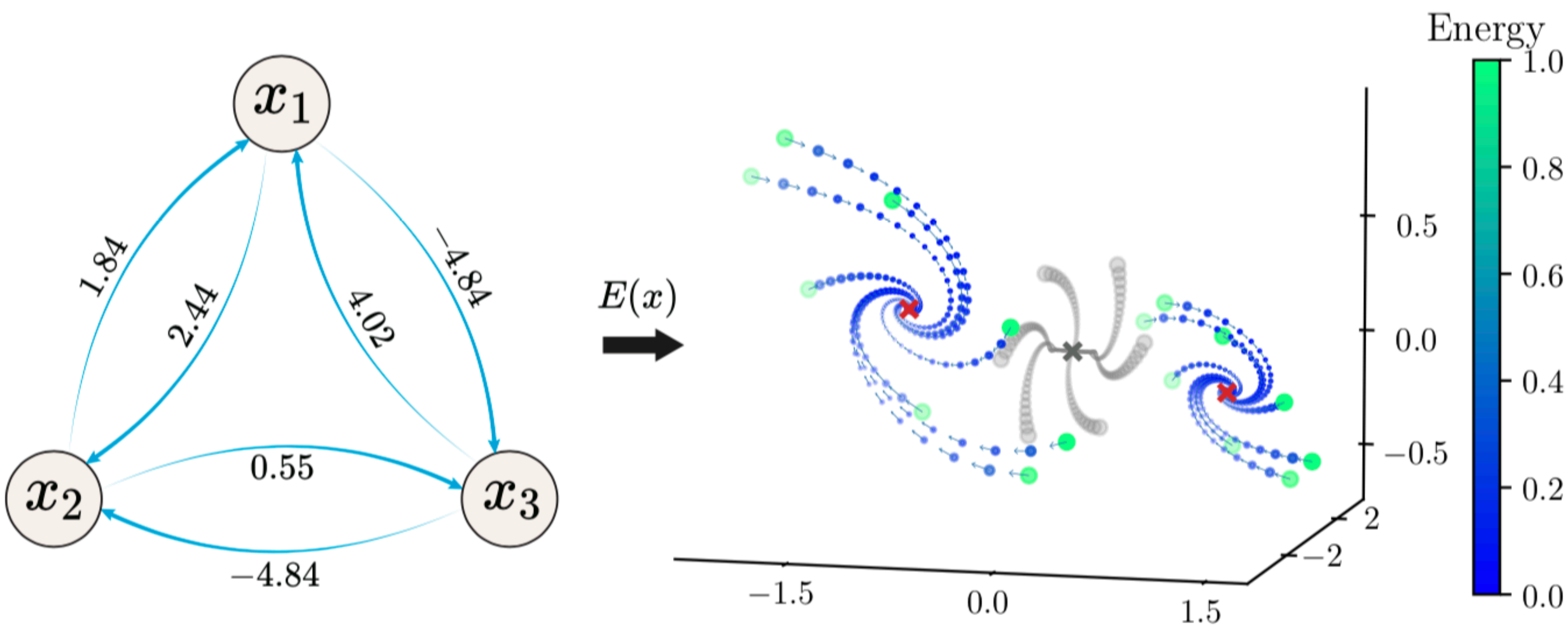
Jin Wang and his colleagues, PNAS 2013

# Intuitive illustration

**The phase space is partitioned into different regions of different speeds**

(quasi-potential) of the dynamics,

$$E(\mathbf{x}) = \frac{1}{2}\sum_i \left(-x_i + \sum_{j=1}^{N} J_{ij}\phi(x_j)\right)^2 + \eta\|\mathbf{x}\|^2, \qquad \textbf{(4)}$$



# Rationale underlying the intuition

According to Eq. (4), we can write down the following stochastic gradient dynamics (or Langevin dynamics)

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = -\nabla_{\mathbf{x}} E(\mathbf{x}) + \sqrt{2T}\boldsymbol{\epsilon}, \tag{5}$$

where $\boldsymbol{\epsilon}(t)$ is a time-dependent white noise, whose statistics is given by $\langle \epsilon_i(t) \rangle = 0$, $\langle \epsilon_i(t)\epsilon_j(t') \rangle = \delta_{ij}\delta(t-t')$. The temperature $T$ is used to tune the energy level, playing the same role as in the equilibrium Boltzmann measure. We can write the gradient (force) in Eq. (5) in the component wise,

$$F_i \equiv -x_i + h_i - \phi'(x_i) \sum_{j:j\neq i} J_{ji}(h_j - x_j), \tag{6}$$

Impact on neighbors

The force is not a gradient of a potential, but approaches the true dynamics in the steady state limit!

**Therefore:**

!!We can formulate the steady state behavior as the canonical ensemble of stationary fixed points!!

$$P(\mathbf{x}) = \frac{1}{Z} e^{-\beta E(\mathbf{x})},$$

**Order parameters for non-gradient dynamics**

# Free energy for non-equilibrium steady state/NESS

$$-\beta f = \lim_{n \to 0} \frac{\ln\langle Z^n \rangle}{Nn}$$

**Quenched disorder over J**

$$Z^n = \int d\mathbf{x} e^{-\beta\left(\frac{1}{2}\sum_{ia}\left(-x_i^a + \sum_j J_{ij}\phi(x_j^a)\right)^2 + \eta\sum_a ||\mathbf{x}^a||^2\right)}$$

**Constraining norm of activity**

**Replicated dynamics states**

# Order parameters

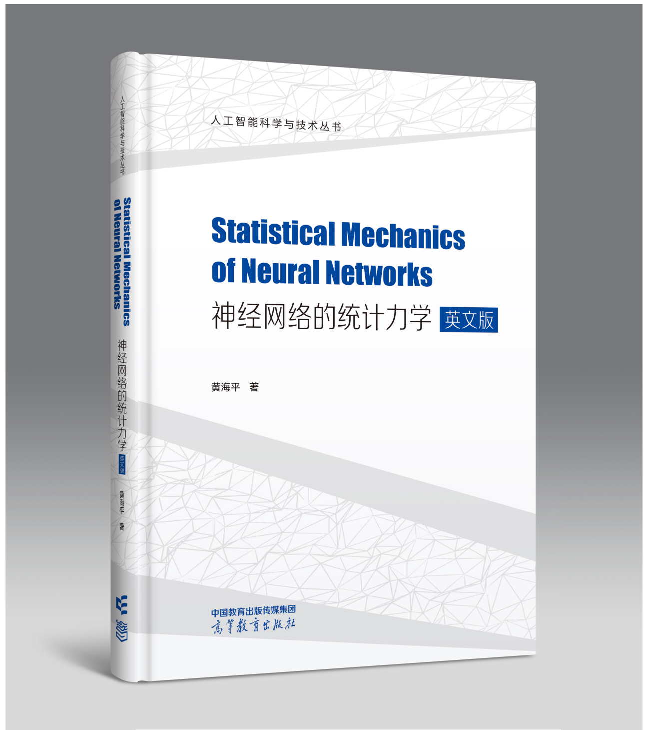
$$Q^{ab} = \frac{1}{N}\sum_i \phi(x_i^a)\phi(x_i^b),$$

Activity level of the dynamic system

$$R^{ab} = \frac{1}{N}\sum_i \hat{x}_i^a \phi(x_i^b),$$

Response property of the dynamic system

**A response field inspired by DMFT [2305.08459]**




**Technical note:**
**Q and R emerges naturally from disorder average!**

# Theory summary

$$
\begin{aligned}
-\beta f =& \frac{1}{2}Q\hat{Q} - q\hat{q} + R\hat{R} - r\hat{r} - \ln\sigma + \frac{1}{2}\beta g^2\gamma\left(r^2 - R^2\right) \\
&+ \int (DuDv)\ln I,
\end{aligned}
\tag{11}
$$

where $\sigma \equiv \sqrt{1 + g^2\beta(q - Q)}$, and $\hat{Q}$, $\hat{q}$, $\hat{R}$ and $\hat{r}$ are conjugated order parameters. The integral $I \equiv \int dx e^{\mathcal{H}(x)}$, where

$$
\begin{aligned}
\mathcal{H}(x) \equiv& -\beta\eta x^2 + \frac{1}{2}\left(2\hat{q} - \hat{Q}\right)\phi^2(x) \\
&+ \left(\sqrt{\frac{g^2\beta Q\hat{Q} - \hat{R}^2}{g^2\beta Q}}u + \frac{\hat{R}}{g\sqrt{\beta Q}}v\right)\phi(x) \\
&- \frac{1}{2\sigma^2}\left(g\sqrt{\beta Q}v + (\hat{r} - \hat{R})\phi(x) - \sqrt{\beta}x\right)^2.
\end{aligned}
\tag{12}
$$

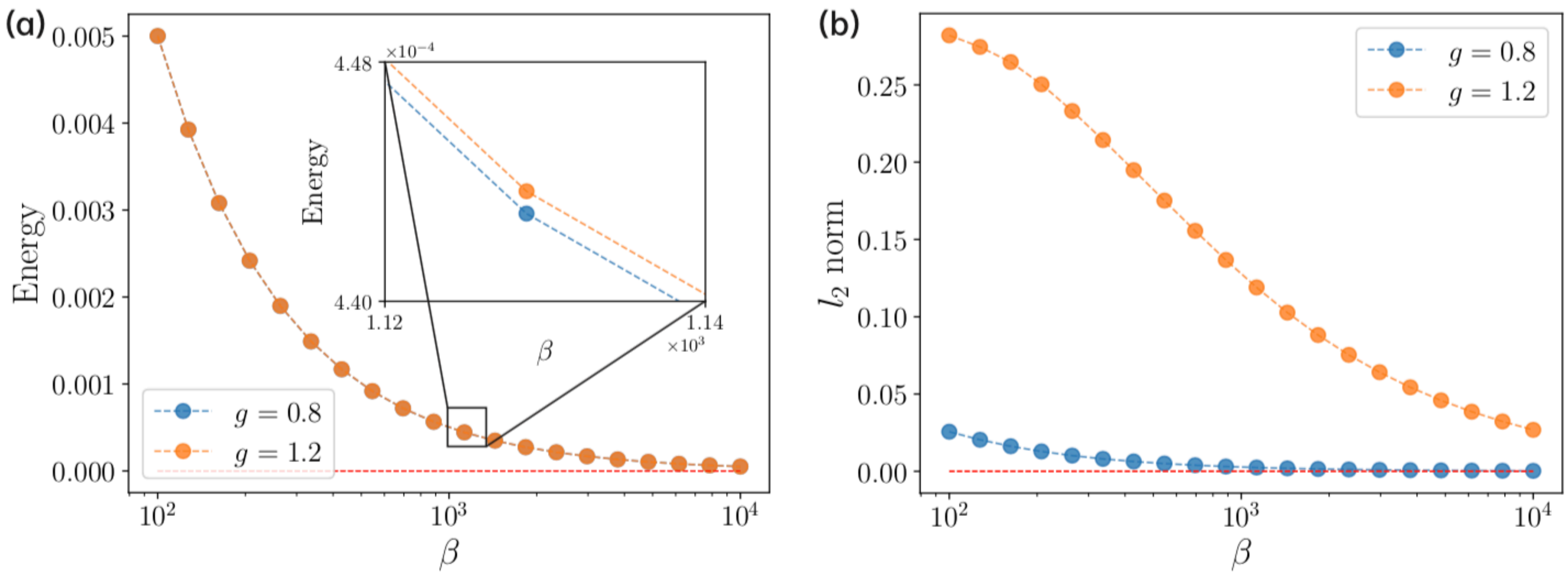
**Single variable effective Hamiltonian**

$$
\begin{aligned}
q =& \left[\langle\phi^2\rangle\right], \\
Q =& \left[\langle\phi\rangle^2\right], \\
\hat{q} =& -\frac{gk}{2} + \frac{g^2k^2Q}{2} + \frac{gk}{2\sigma^2}\left(\hat{r} - \hat{R}\right)f(1,1,-2)\left[\langle\phi^2\rangle\right] \\
&+ \frac{gk}{\sigma^2}\left(\hat{r} - \hat{R}\right)f(0,-1,1)\left[\langle\phi\rangle^2\right] + \frac{k^2}{2}(1 - 2gkQ)\left[\langle x^2\rangle\right] \\
&+ \frac{gk\sqrt{\beta}}{\sigma^2}f(0,1,-2)[\langle x\rangle\langle\phi\rangle] + \frac{gk\sqrt{\beta}}{\sigma^2}f(-1,0,2)[\langle x\phi\rangle] \\
&+ gk^3Q\left[\langle x\rangle^2\right], \\
\hat{Q} =& g^2k^2Q + \frac{2gk}{\sigma^2}\left(\hat{r} - \hat{R}\right)f(0,1,-1)\left[\langle\phi^2\rangle\right] \\
&+ \frac{gk}{\sigma^2}\left(\hat{r} - \hat{R}\right)f(1,-3,2)\left[\langle\phi\rangle^2\right] - 2gk^3Q\left[\langle x^2\rangle\right] \\
&- \frac{2gk\sqrt{\beta}}{\sigma^2}f(1,-2,2)[\langle x\rangle\langle\phi\rangle] + \frac{2gk\sqrt{\beta}}{\sigma^2}f(0,-1,2)[\langle x\phi\rangle] \\
&+ k^2(1 + 2gkQ)\left[\langle x\rangle^2\right], \\
r =& -\frac{1}{\sigma^2}f(1,0,-1)\left[\langle\phi^2\rangle\right] + \frac{1}{\sigma^2}f(0,1,-1)\left[\langle\phi\rangle^2\right] \\
&+ \frac{\sqrt{\beta}}{\sigma^2}(1 - gkQ)[\langle\phi x\rangle] + \frac{\sqrt{\beta}}{\sigma^2}gkQ[\langle\phi\rangle\langle x\rangle], \\
R =& -\frac{1}{\sigma^2}f(0,1,-1)\left[\langle\phi^2\rangle\right] - \frac{1}{\sigma^2}f(1,-2,1)\left[\langle\phi\rangle^2\right] \\
&- \frac{\sqrt{\beta}}{\sigma^2}gkQ[\langle x\phi\rangle] + \frac{\sqrt{\beta}}{\sigma^2}(1 + gkQ)[\langle x\rangle\langle\phi\rangle], \\
\hat{r} =& \beta g^2\gamma r, \\
\hat{R} =& \beta g^2\gamma R,
\end{aligned}
$$

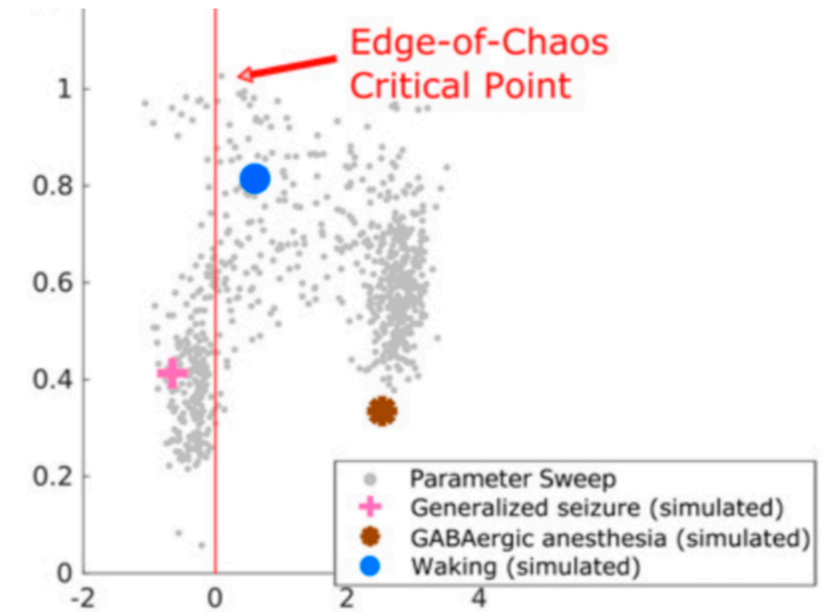
**Replica symmetry Ansatz:**

$$Q^{ab} = q\delta_{ab} + Q(1 - \delta_{ab}) \text{ and } R^{ab} = r\delta_{ab} + R(1 - \delta_{ab})$$

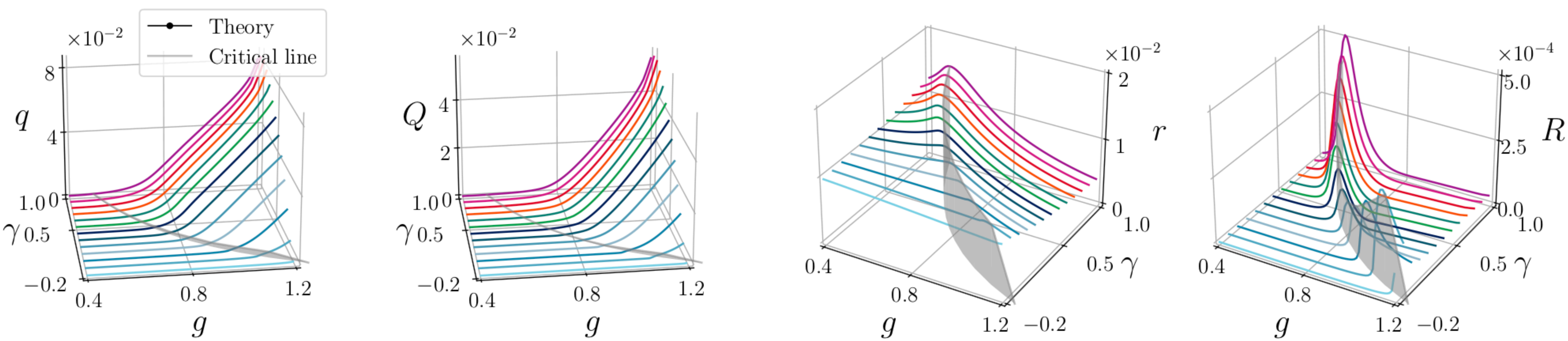
# Result I: energy



# Result II: order parameters



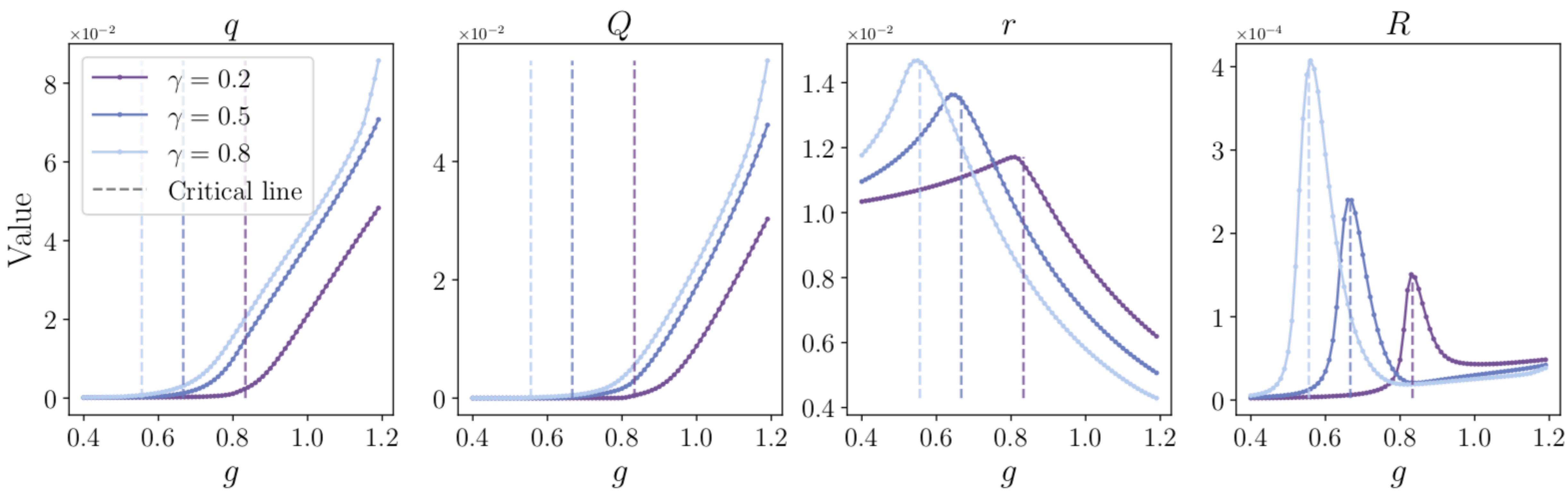

**Continuous transition to chaos!**



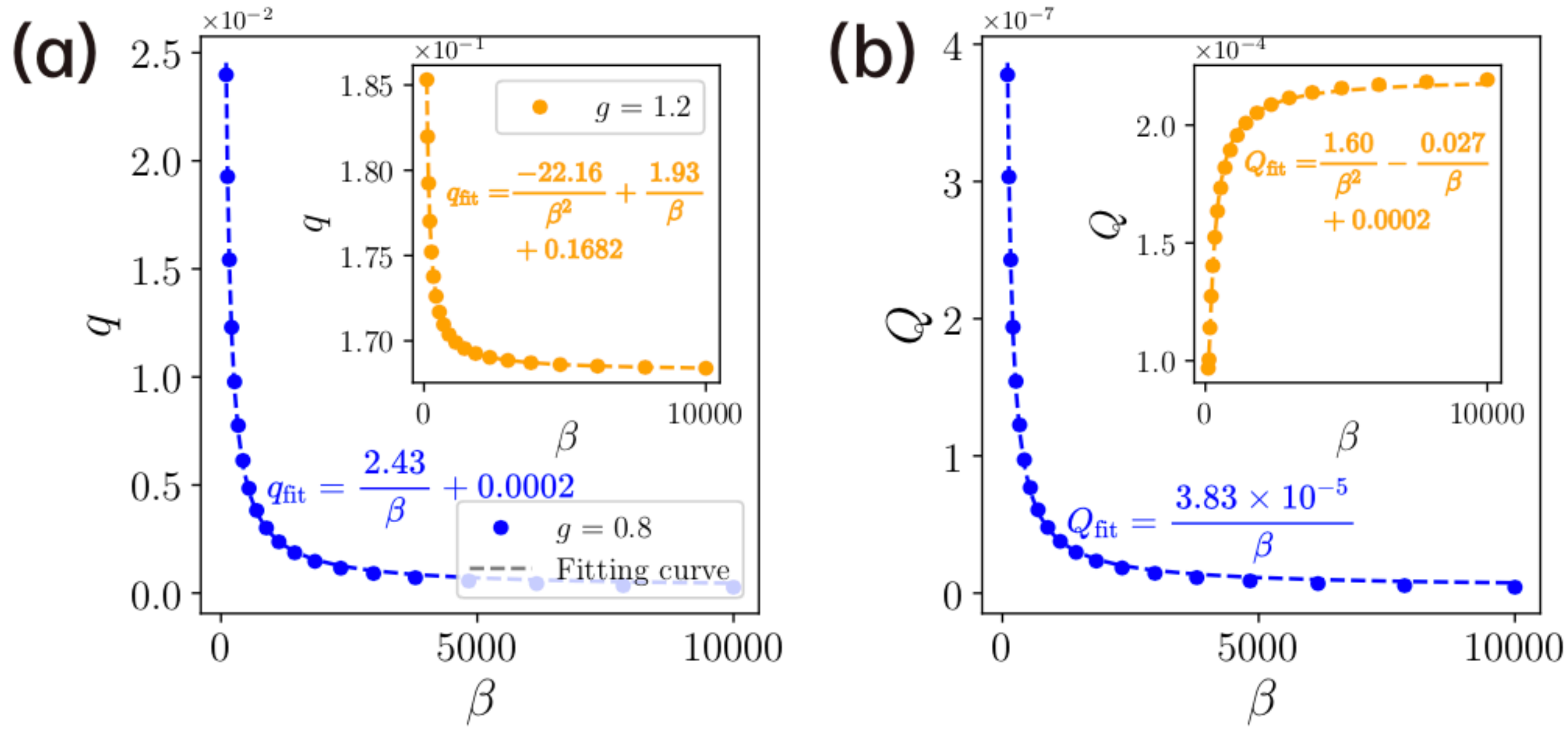

**Response R & r peaked at the transition point**

**!!The steady dynamics is more responsive at the edge of chaos!!**

# Result II: order parameters



# Result III: scaling behavior




$$q - Q \propto \beta^{-1}$$

$$R - r \propto \beta^{-1/2}$$

$$r \propto \beta^{1/2}$$