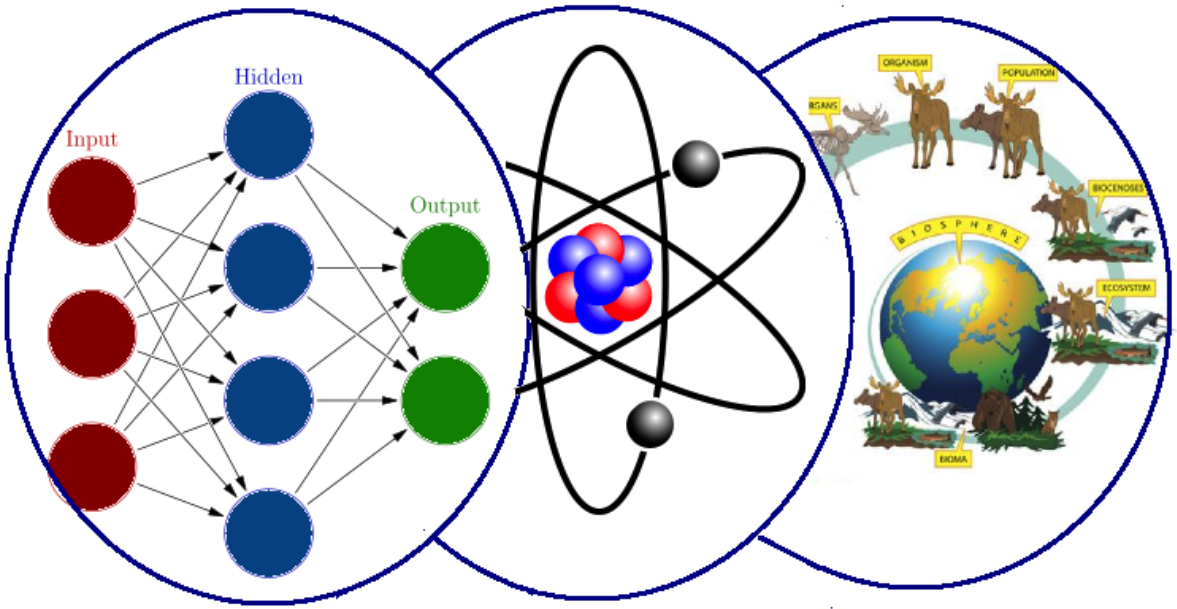
EVOLUTION AS LEARNING AND A THERMODYNAMIC ORIGIN OF LIFE



Vitaly Vanchurin

National Institutes of Health, USA
Research Professor
http://NIH.gov

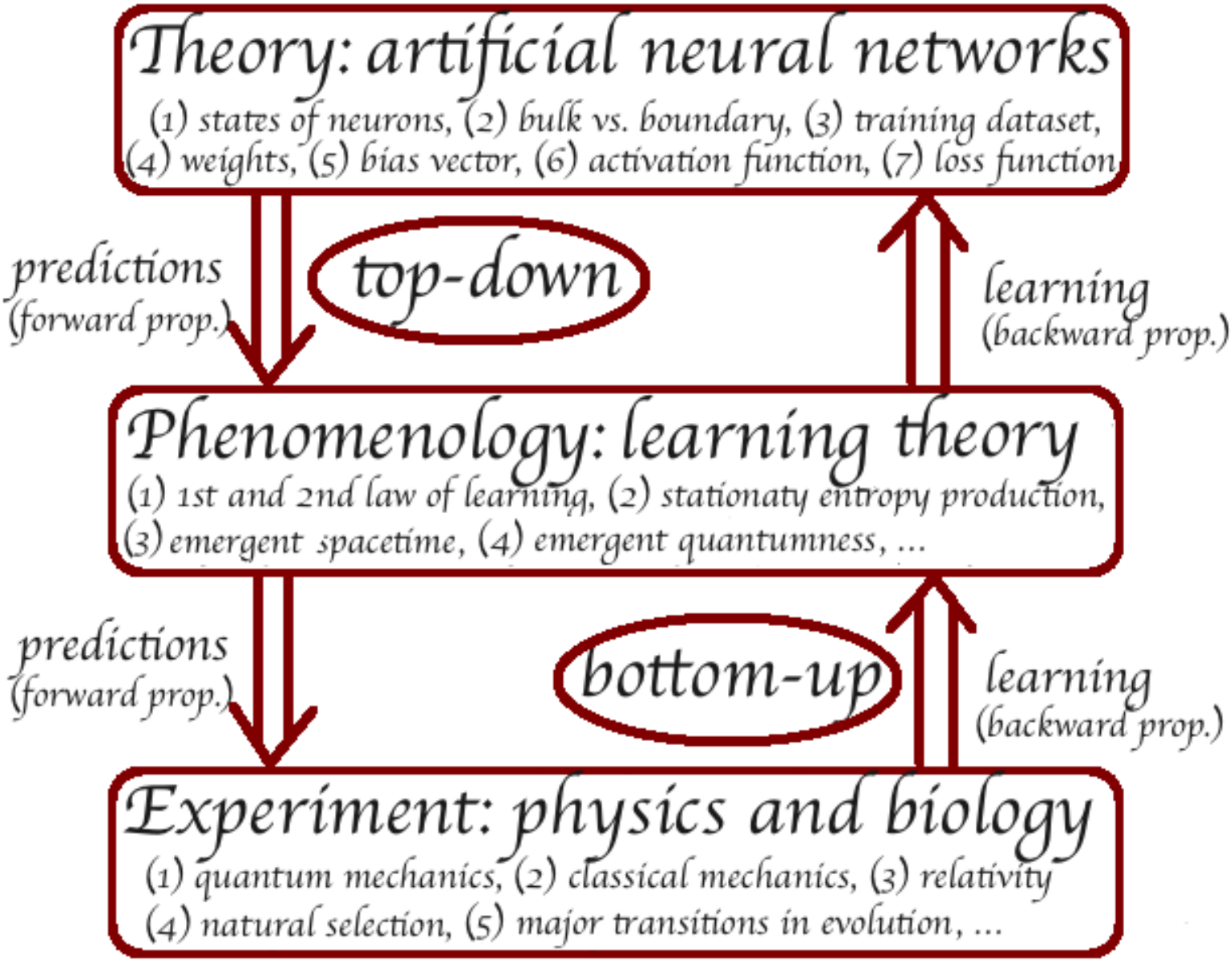
⇒

Artificial Neural Computing, USA
Founder & CEO
http://ArtificialNeuralComputing.com

REFERENCES:

- ON THEORY OF LEARNING:
  - V. V., "Towards a theory of machine learning", MLST 2 (3), 035012 (2020)
- ON EVERYTHING AS LEARNING:
  - V. V., "The world as a neural network", Entropy 22 (11), 1210 (2020)
- ON PHYSICS AS LEARNING:
  - QUANTUM: M.I.Katsnelson, V.V., "Emergent Quantumness in Neural Networks", Foundations of Physics 51 (5), 1-20 (2021)
  - CRITICAL: M. I. Katsnelson, V.V., T.Westerhout, "Self-organized criticality in neural networks", arXiv:2107.03402
  - GRAVITATIONAL: V.V., "Towards a theory of quantum gravity from neural networks", Entropy 24 (1), 7 (2021)
- ON BIOLOGY AS LEARNING:
  - THEORY: V.V., Y.I. Wolf, M. I. Katsnelson, E.V. Koonin, "Towards a Theory of Evolution as Multilevel Learning", PNAS 119 (6) (2022)
  - PHENOMENOLOGY: V.V., Y.I. Wolf, E.V. Koonin, M. I. Katsnelson, "Thermodynamics of Evolution and the Origin of Life", PNAS 119 (6)
  - NUMERICS: A.Grabovsky, V.V., "Bio-inspired machine learning: programmed death and replication", (to appear on arXiv tonight)
  - EXPERIMENT: In progress.

FLOWCHART OF THEORETICAL MODELING



THE PROBLEM OF LEARNING

- During learning the loss function is minimized with respect to trainable variables

$$\mathbf{q} = \left(\mathbf{q}^{(c)}, \mathbf{q}^{(a)}, \mathbf{q}^{(n)}\right)$$

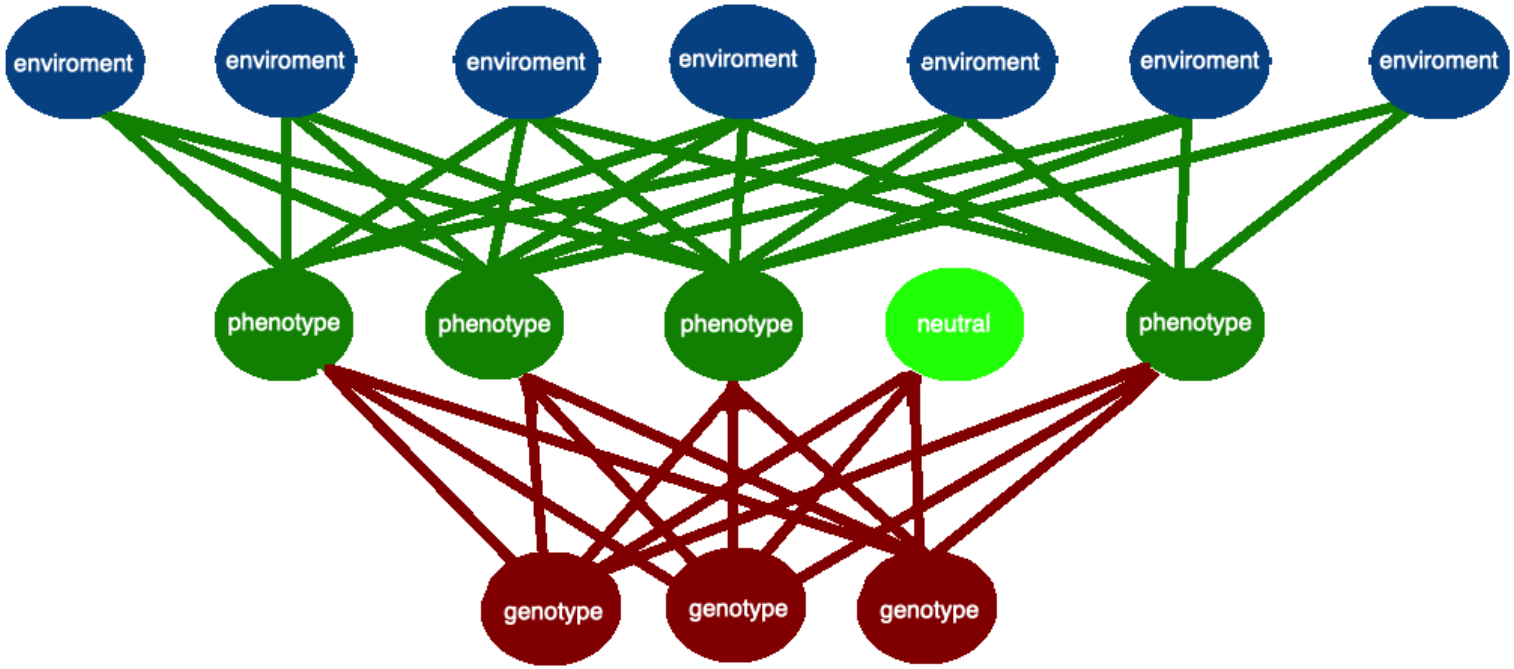
for a given training dataset of non-trainable variables,

$$\mathbf{x} = \left(\mathbf{x}^{(o)}, \mathbf{x}^{(e)}\right)$$

Easy to remember as C-A-N-O-E.

- Near equilibrium, the first derivative of loss function with respect to q_i's is small, but the second derivative can either be large for *core* variables, $\mathbf{q}^{(c)}$; small for *adaptable* variables, $\mathbf{q}^{(a)}$; or near zero, for *neutral* variables $\mathbf{q}^{(n)}$.
- From the biological perspective, this is equivalent to optimizing the state of an organism $\mathbf{x}^{(o)}$ with respect to the state of the environment $\mathbf{x}^{(e)}$ by adjusting the biological traits of organism, or equivalently the trainable degrees of freedom $\mathbf{q}$.
- On the time scale τ of life time of an organism, adaptable variables $\mathbf{q}^{(a)}$ are the phenotypic traits that quickly react to environmental changes $\mathbf{x}^{(e)}$, whereas the core variables $\mathbf{q}^{(c)}$ are the genomic sequences that change minimally if at all.

ORGANISM MODELED AS A NEURAL NETWORK



The "canoe":

- $\mathbf{q}^{(c)}$ trainable genotype variables (red nodes/links)
- $\mathbf{q}^{(a)}$ trainable adaptive phenotypic variables (dark green nodes and links)
- $\mathbf{q}^{(n)}$ neutral variables (light green node)
- $\mathbf{x}^{(o)}$ non-trainable organism variables (red and green nodes)
- $\mathbf{x}^{(e)}$ non-trainable environmental variables (blue nodes)

NEURAL NETWORK THEORY

- Consider a learning system represented as a neural network, with the state vector described by trainable variables $\mathbf{q}$ (e.g. weight matrix $\hat{w}$ and bias vector $\mathbf{b}$) and non-trainable variables $\mathbf{x}$ (e.g. state vector of individual neurons).
- Non-trainable variables are updated in discrete time-steps

$$x_i(t+1) = f_i\left(\sum_j w_{ij}x_j(t) + b_i\right) \tag{1}$$

where $f_i(y)$'s are some non-linear activation functions (e.g. hyperbolic tangent).
- Trainable variables are updated according to (stochastic) gradient descent

$$q_i(t+1) = q_i(t) - \gamma\frac{\partial H(\mathbf{x}(t), \mathbf{q}(t))}{\partial q_i} \tag{2}$$

where γ is the learning rate parameter and $H(\mathbf{x}, \mathbf{q})$ is the loss function.
- For example, "boundary" loss function is

$$H_e(\mathbf{x}, \mathbf{q}) \equiv \frac{1}{2}\sum_i \left(x_i^{(e)} - f_i^{(e)}\left(\mathbf{x}^{(o)}, \mathbf{q}\right)\right)^2 \tag{3}$$

and "bulk" loss function is

$$H(\mathbf{x}, \mathbf{q}) = \frac{1}{2}\sum_i \left(x_i - f_i\left(\mathbf{x}^{(o)}, \mathbf{q}\right)\right)^2 + V(\mathbf{x}, \mathbf{q}). \tag{4}$$

FITNESS FUNCTION

- Bulk loss function

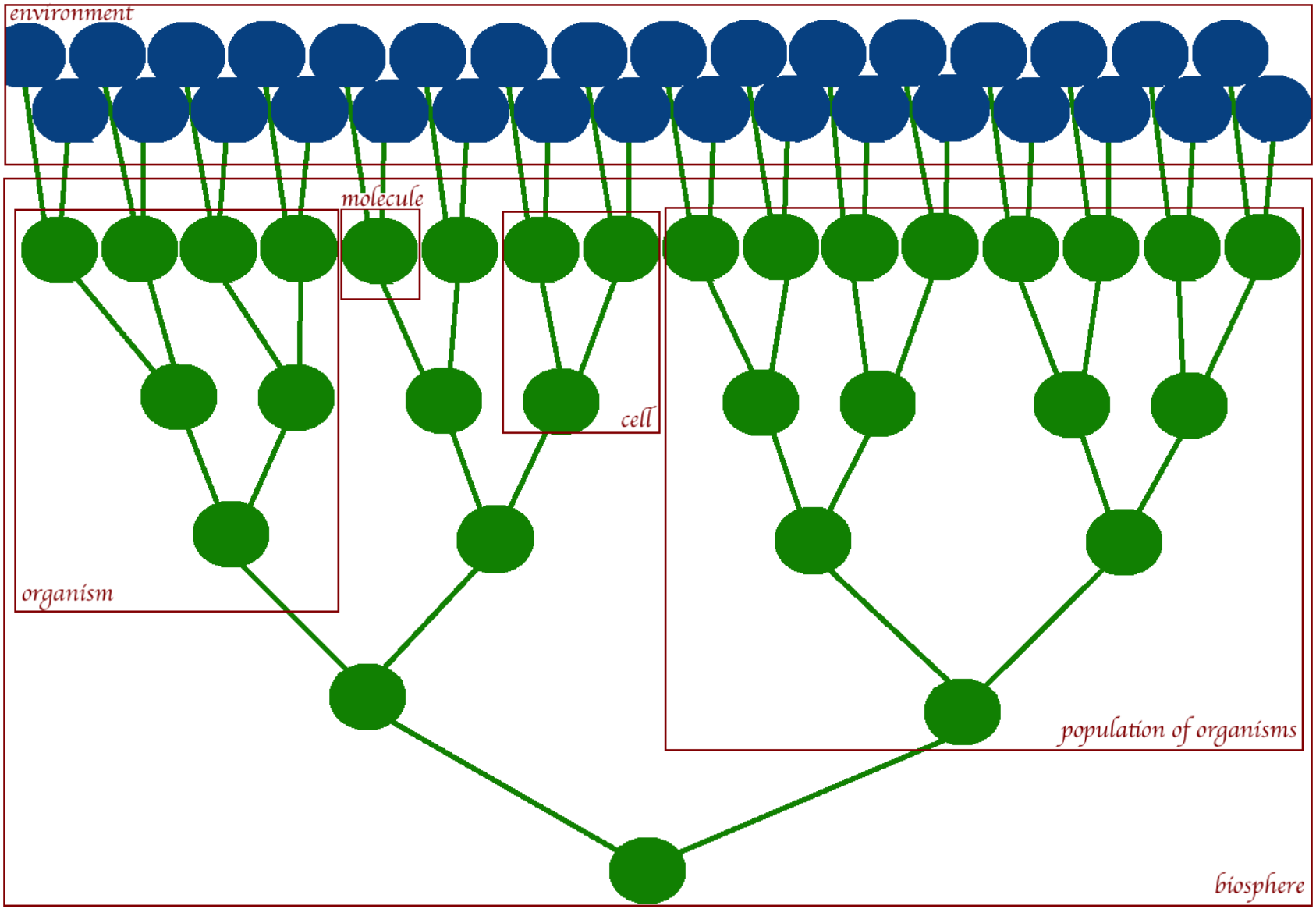
$$H(\mathbf{x}, \mathbf{q}) = \frac{1}{2} \sum_i \left(x_i - f_i\left(\mathbf{x}, \mathbf{q}\right)\right)^2 + V(\mathbf{x}, \mathbf{q}) \tag{5}$$

- The kinetic term reflects the ability of organisms (or learning subsystems) to predict the changes in the state of the given environment over time, whereas the potential term reflects its compatibility with a given environment.
- In the context of biological evolution, Malthusian fitness φ is defined as the expected reproductive success of a given genotype, that is, the rate of change of the prevalence of the given genotype in an evolving population.
- In the context of the theory of learning (as we shall see) the more relevant function is additive fitness $\log \varphi$ which is related to the loss function through

$$H(\mathbf{x}, \mathbf{q}) = -T \log \varphi(\mathbf{x}, \mathbf{q}). \tag{6}$$

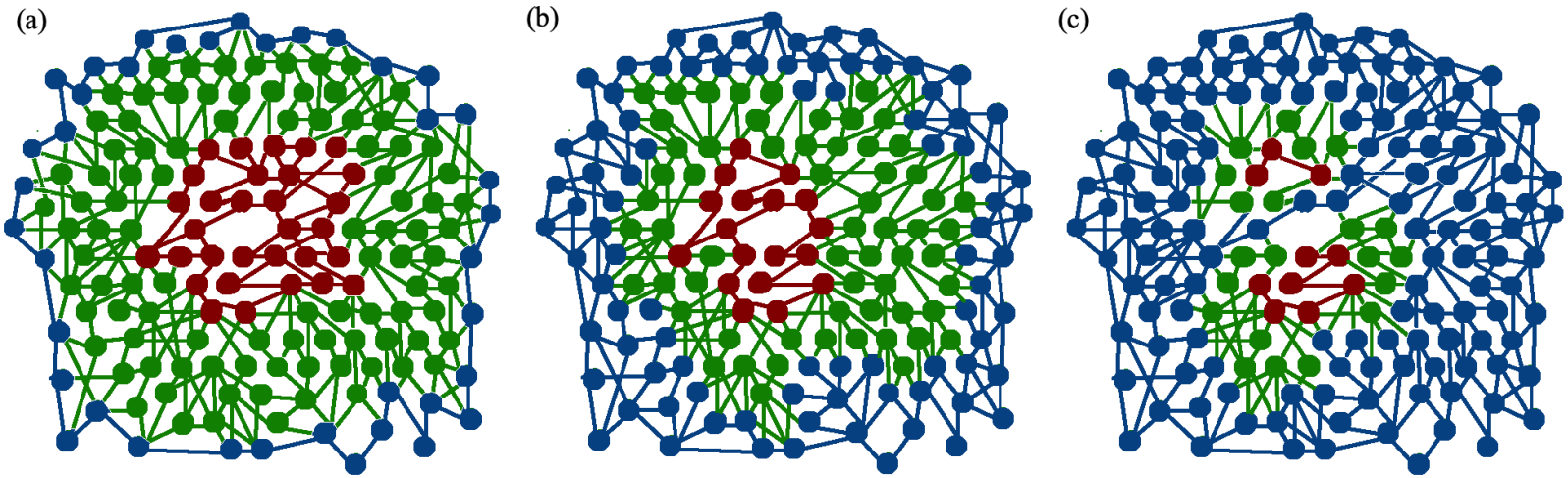
- At the level of microscopic description of learning, the proportionality constant T is unimportant, but at the level of statistical ensembles, $\beta = T^{-1}$ is the Lagrange multiplier which imposes constraint on the average loss function.

BIOSPHERE MODELED AS A NEURAL NETWORK



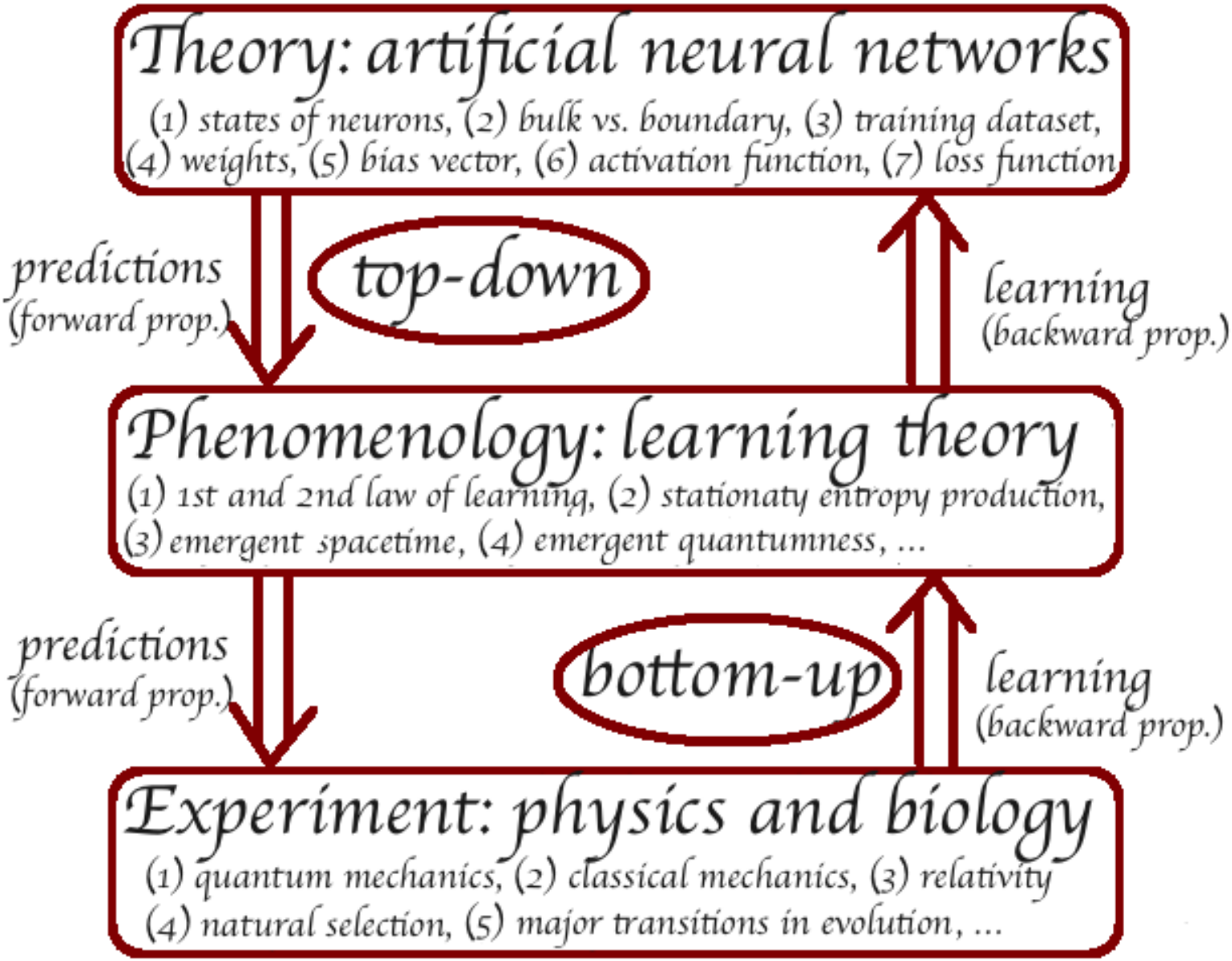
MULTILEVEL LEARNING

- Evidently, depending on the time-scale τ the same degree of freedom might be described using different variables (e.g. $\mathbf{q}^{(c)}$, $\mathbf{q}^{(a)}$, $\mathbf{q}^{(n)}$, $\mathbf{x}^{(o)}$, $\mathbf{x}^{(e)}$,).
- It is useful to partition all these variables into three classes depending on how fast they change with respect to τ, i.e. considered time scale:
  1. slow-changing variables are the already well trained and effectively constant degrees of freedom $\mathbf{q}^{(c)}$ that only change on time scales $\gg \tau$.
  2. intermediate-changing variables are either adaptable $\mathbf{q}^{(a)}$ or neutral $\mathbf{q}^{(n)}$ variables that change on time scales $\sim \tau$.
  3. fast-changing variables are the non-trainable variables that characterize an organism ($\mathbf{x}^{(o)}$) and its environment ($\mathbf{x}^{(e)}$), and change on time scales $\ll \tau$.



smaller τ larger τ

FLOWCHART OF THEORETICAL MODELING



STATISTICAL MECHANICS OF EVOLUTION

- **Maximum entropy principle:** distribution of any quantity is given by the highest entropy distribution subject to the relevant constraints.
  - For example, constraint imposed on the average loss function

$$\int d^N x \, H(\mathbf{x}, \mathbf{q}) p(\mathbf{x}|\mathbf{q}) = U(\mathbf{q}) \tag{7}$$

  - Prob. distribution over non-trainable variables (a.k.a. canonical ensemble)

$$p(\mathbf{x}|\mathbf{q}) \propto \exp\left(-\beta H(\mathbf{x}, \mathbf{q})\right) \tag{8}$$

  - Corresponding partition function (or macroscopic counterpart of fitness)

$$\mathcal{Z}(\beta, \mathbf{q}) \equiv \int d^N x \, e^{-\beta H(\mathbf{x}, \mathbf{q})} \tag{9}$$

- *Free energy* encodes everything there is to know about the system

$$F(\beta, \mathbf{q}) \equiv -\beta^{-1} \log \mathcal{Z}(\beta, \mathbf{q}) \tag{10}$$

  - average loss function

$$U(\beta, \mathbf{q}) = \frac{\partial}{\partial \beta} \left(\beta F(\beta, \mathbf{q})\right), \tag{11}$$

  - entropy of non-trainable variables (e.g. environment)

$$S(\beta, \mathbf{q}) = \beta^2 \frac{\partial}{\partial \beta} F(\beta, \mathbf{q}) \tag{12}$$

THERMODYNAMICS OF EVOLUTION

- From the first and second laws of learning/thermodynamics:

$$dF = dU - TdS + \mathbf{Q} \cdot d\mathbf{q} = 0, \tag{13}$$

- *Biological* temperature is defined as

$$T = \beta^{-1} \tag{14}$$

where β is a Lagrange multiplier which imposed a constraint on the average loss.

- When the number of variables can vary then, the *grand potential* must vanish

$$d\Omega = dU - TdS - \mu dK = 0 \tag{15}$$

where μ is *evolutionary* potential.

- At the level of the network of information processing units, μ describes evolutionary potential for adding/removing adaptable trainable variables.

POPULATION OF ORGANISMS

- Consider an ensemble of organisms that differ from each other by the values of adaptable variables $\mathbf{q}^{(a)}$, whereas $\mathbf{q}^{(c)}$ are the same for all organisms.
- Ensemble can either represent a Bayesian (subjective) probability distribution over degrees of freedom of a single organism or a frequentist (objective) probability distribution over different organisms.
- In the limit of an infinite number of organisms, the two interpretations are indistinguishable, but in the context of actual biological evolution, the total number of organisms is only exponentially large

$$N_e \propto \exp(bK) \tag{16}$$

- Then to study the state of a learning equilibrium for a grand canonical ensemble of evolving organisms, it is convenient to express the average loss function phenomenologically as

$$U(S,K) = T(S,K)S + \mu(S,K)K \tag{17}$$

where the conjugate variables are, respectively, biological temperature

$$T \equiv \frac{\partial U}{\partial S}, \tag{18}$$

and evolutionary potential

$$\mu \equiv \frac{\partial U}{\partial K}. \tag{19}$$

IDEAL MUTATIONS MODEL

- Consider N_e organisms described by genotypes $\mathbf{q}_1, ..., \mathbf{q}_{N_e}$ that can undergo rare mutations (on time-scales $\sim \tau$) followed by fast fixation (on shorter time-scales $\ll \tau$), but the total number of organisms N_e remains constant [Kimura (1983)]
- Fixation on short time-scales implies that the state of the system is such that all organisms have the same genotype $\mathbf{q}_1 = ... = \mathbf{q}_{N_e} = \mathbf{q}$ and equilibration on the longer time-scales implies that the marginal distribution is given by

$$p(\mathbf{q}) \propto \int \prod_{n=1}^{N_e} d^N x_n \exp\left(-\beta \sum_{n=1}^{N_e} H(\mathbf{x}_n, \mathbf{q})\right) = \exp\left(-\beta N_e F(\mathbf{q})\right) \tag{20}$$

where integration is taken over states of environment $\mathbf{x}_n$ for all organisms.

- This distribution was also considered by Sella and Hirsh in 2005, who interpreted N_e as inverse temperature whereas in our framework it is β.
- The distribution can also be expressed as

$$p(\mathbf{q}) \propto \mathcal{Z}(\mathbf{q})^{N_e} \tag{21}$$

where the partition function $\mathcal{Z}(\mathbf{q}) = \exp(-\beta F(\mathbf{q}))$ is the macroscopic counterpart of fitness $\varphi(\mathbf{x}, \mathbf{q}) = \exp(-\beta H(\mathbf{x}, \mathbf{q}))$.

- Prediction: if such a system evolved from one equilibrium to another then

$$\frac{\log \mathcal{Z}^{(1)}(\mathbf{q})}{\log \mathcal{Z}^{(2)}(\mathbf{q})} = \frac{\beta_1 F(\mathbf{q})}{\beta_2 F(\mathbf{q})} = \frac{\beta_1}{\beta_2} = \frac{T_2}{T_1} \tag{22}$$

must be independent of $\mathbf{q}$, that is, are the same for all organisms in the ensemble.

PHENOMENOLOGICAL MODELING

- Obtained distribution enables us to calculate the average loss function

$$U(K) = \langle H(\mathbf{x}, \mathbf{q}) N_e \rangle \propto \langle H(\mathbf{x}, \mathbf{q}) \rangle \exp(bK), \tag{23}$$

where $\langle H(\mathbf{x}, \mathbf{q}) \rangle$ is the average loss of individual organisms, but the dependence on entropy is not yet explicit.

- In principle, we should be able to reconstruct $U(S, K)$ directly from experiment or simulation, but for the sake of illustration, consider a phenomenological model

$$U(S, K) = \langle H(\mathbf{x}, \mathbf{q}) N_e \rangle = a S^n \exp\left(\frac{b}{S} K\right) \tag{24}$$

- Thus we assume:
  - loss function of individual organisms scales as $\langle H(\mathbf{x}, \mathbf{q}) \rangle \propto S^n$ from some $n > 0$, i.e. the loss is greater in an environment with a higher entropy and
  - the number of adaptable variables scales as $K \propto S \log N_e$, i.e. the large the entropy S in the environment the more variables are required to learn it
- By preforming Legendre transformation of $U(S, K)$ we obtain grand potential

$$\Omega(T, \mu) = -a(n-1)\left(\frac{\mu}{eb}\right)^{\frac{n}{n-1}} \exp\left(\frac{bT}{(n-1)\mu}\right), \tag{25}$$

which can be reconstructed from numerical simulations or observations of time-series of the number of organisms $N_e(t)$ and of their fitness $Z(\mathbf{q}, t)$.

MAJOR (AND MINOR) TRANSITIONS IN EVOLUTION

- Phase transitions from a gas of non-interacting subsystems defined by

$$\langle N_e \rangle = \bar{N}_e \tag{26}$$

to a gas of interacting subsystems defined by

$$\langle K \rangle = \bar{K} \tag{27}$$

- Mathematically transition is from grand canonical ensemble (e.g. molecules)

$$\Omega_p(\mathcal{T}, \mathcal{M}) \propto \mathcal{T}^{\alpha} \exp\left(\gamma \mathcal{M}/\mathcal{T}\right) \tag{28}$$

to grand canonical ensemble (e.g. organisms)

$$\Omega_b(T, \mu) \propto \mu^{c} \exp\left(bT/\mu\right) \tag{29}$$

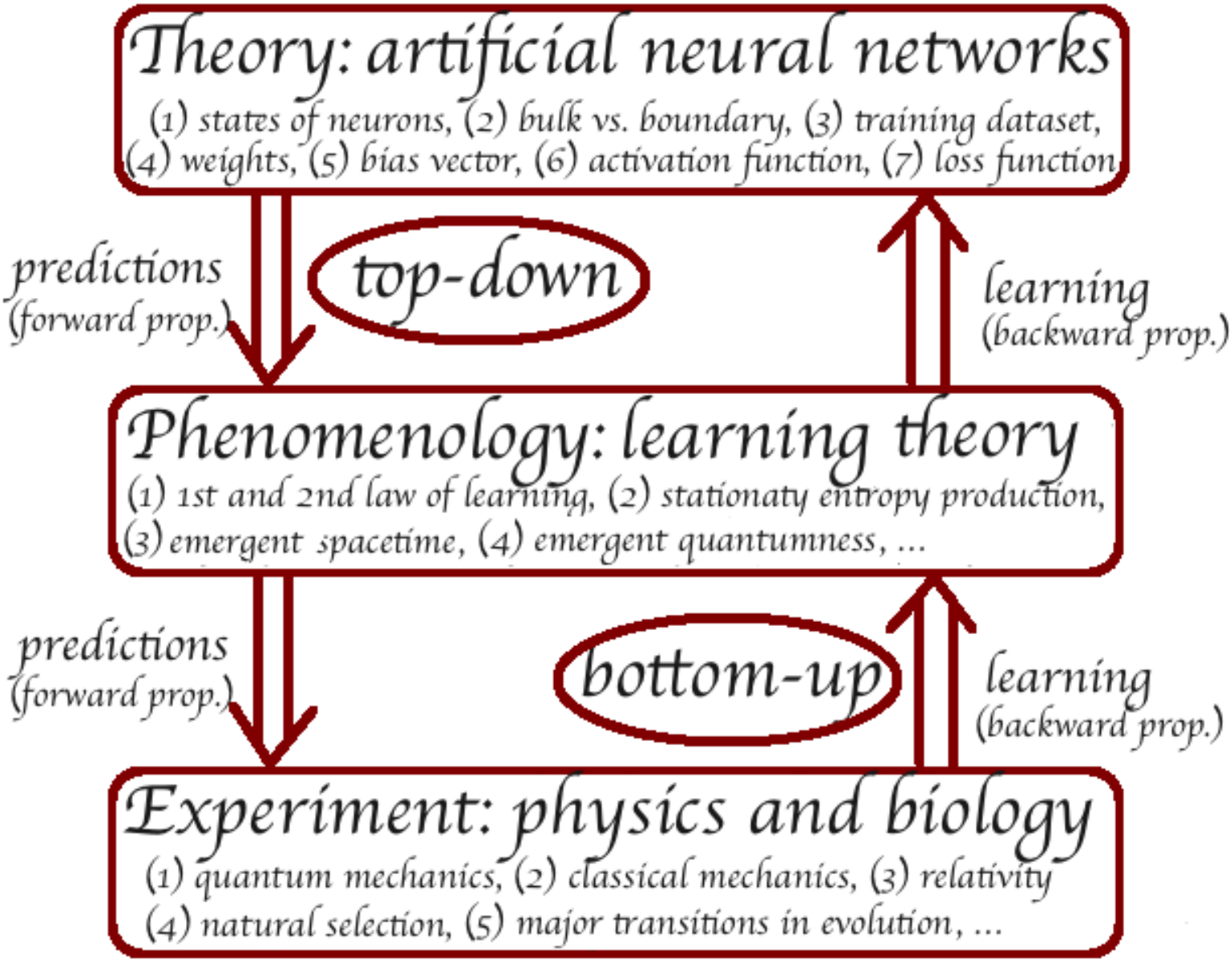
- At the point of phase transition the two potentials are equal

$$\Omega_p(\mathcal{T}_0, \mathcal{M}_0) \propto \mathcal{T}_0^{\alpha} e^{\gamma \mathcal{M}_0/\mathcal{T}_0} = \left(e^{\frac{bT_0}{\alpha \mu_0}}\right)^{\alpha} e^{c \log(\mu_0)} = e^{b\frac{T_0}{\mu_0}} \mu_0^{c} \propto \Omega_b(T_0, \mu_0), \tag{30}$$

where $T_0 = \frac{\alpha}{b} \mu_0 \log(\mathcal{T}_0)$ and $\mathcal{M}_0 = \frac{c}{\gamma} \mathcal{T}_0 \log(\mu_0)$

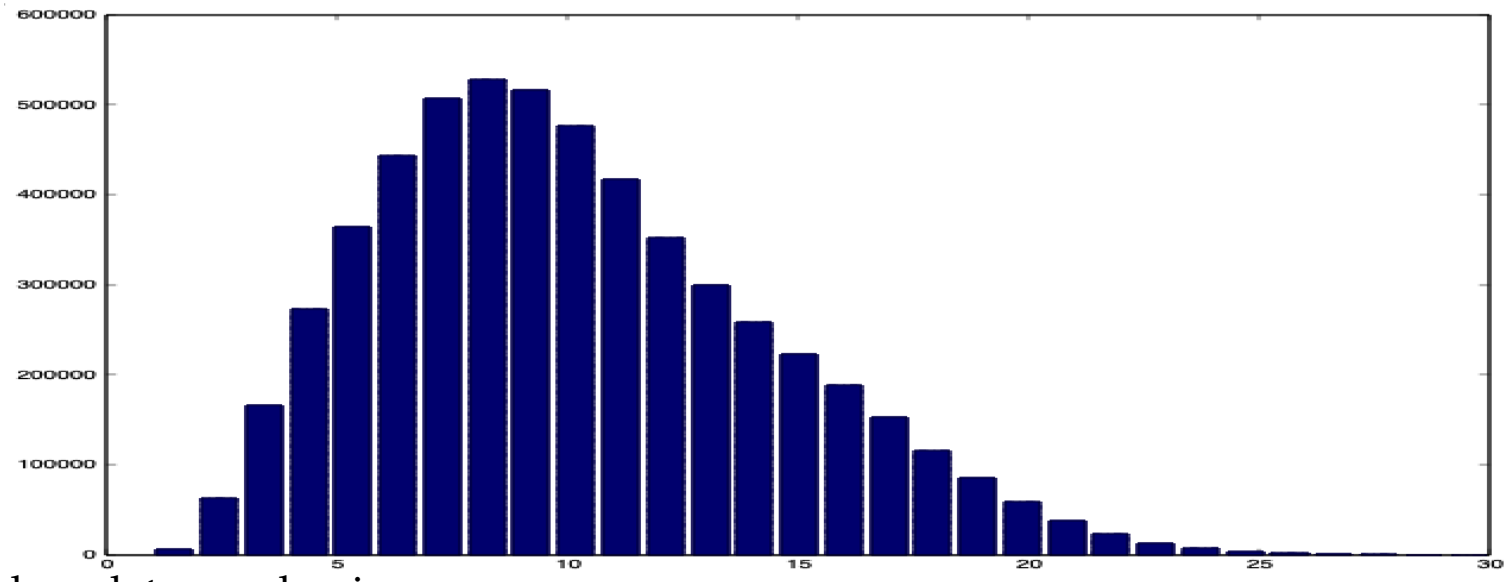
- After the phase transition a new level (i.e. new scale) in the hierarchy is formed.

FLOWCHART OF THEORETICAL MODELING

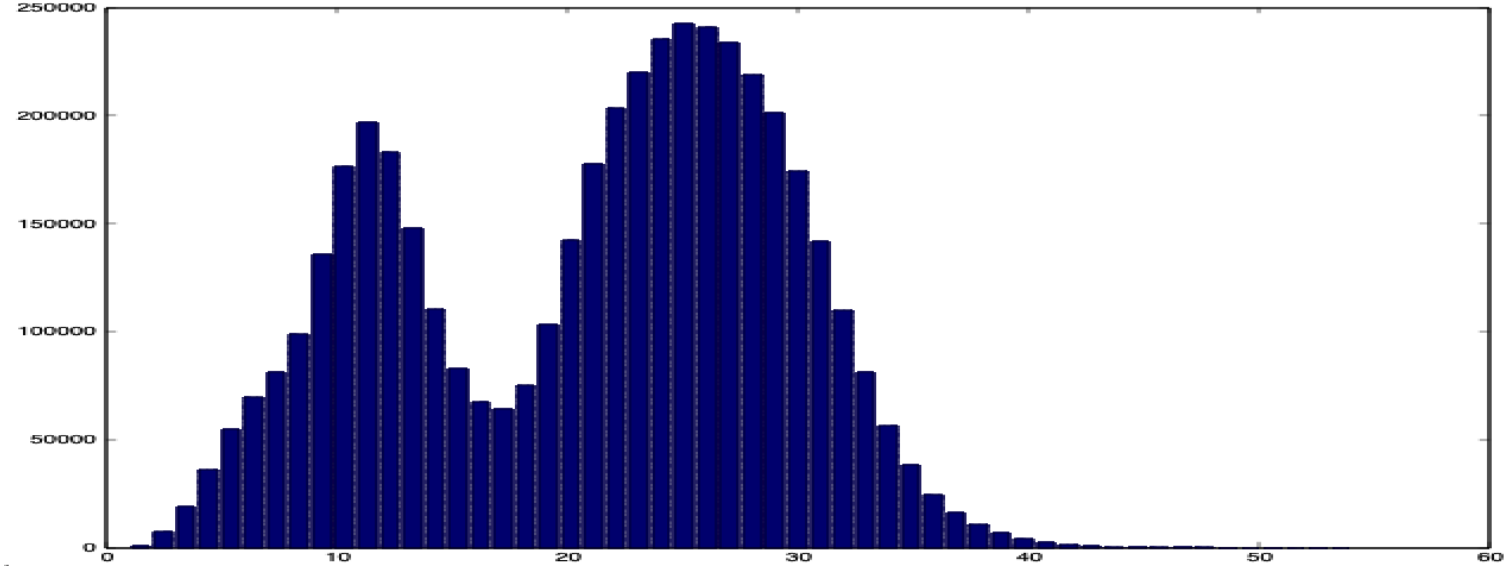


FORMATION OF NEW LEVELS IN EVOLUTION OF CORONAVIRUS

London - early pandemic:



London - late pandemic:



CONCLUSION

1. THEORY: ARTIFICIAL NEURAL NETWORKS
   - major evolutionary phenomena can be modeled using neural networks
   - multilevel learning implies the *same* evolutionary dynamics on all levels
   - generalized central dogma is derived in the context of deep networks
2. PHENOMENOLOGY: LEARNING THEORY
   - biological counterparts of temperature and chem. potential are identified
   - grand potential can be reconstructed phenomenologically from data
   - major transitions in evolution can be described as phase transitions
3. EXPERIMENT: BIOLOGICAL OBSERVATIONS
   - formation of new levels in evolution of coronavirus was observed
   - more observational, experimental and numerical tests are needed
   - e.g. statistical biology, collider biology, artificial biology, etc.

For questions and comments feel free to email me at vitaly.vanchurin@gmail.com