doi:10.20944/preprints202205.0134.v1

WILL ARTIFICIAL INTELLIGENCE BECOME CONSCIOUS? CAN THERMODYNAMICS EXPLAIN THE EVOLUTION OF INTELLECT?

Eva Deli*

University of Debrecen, Department of Anatomy, Histology and Embryology
3420 East Anderson drive, Phoenix, AZ 85032

(507) 279-8370

* Corresponding author

eva.kdeli@gmail.com

0000-0003-1126-1970

Abstract:

Deep neural networks (DNNs), founded on the brain's neuronal organization, can extract higher-level features from raw input. However, complex intellect via autonomous decision-making is way beyond current AI design. Here we propose an autonomous AI inspired by the thermodynamic cycle of sensory perception, operating between two information density reservoirs. Stimulus unbalances the high entropy resting-state and triggers a thermodynamic cycle. By recovering the initial conditions, self-regulation generates a response while accumulating an orthogonal, holographic potential. The resulting high-density manifold is a stable memory and experience field, which increases future freedom of action via intelligent decision-making.

Keywords: deep neural networks, artificial intelligence, intelligent computation, orthogonal transformation, entropy, Carnot cycle

 doi:10.20944/preprints202205.0134.v1

1. Introduction

Although all life processes are based on interaction with the environment, their fundamental characteristic is homeostasis, a dynamic equilibrium. Sensory abilities lend environmental insight for purposeful behavior to even the most primitive animals. The sensory cycle revolves between an information-rich source (the environment) and an information-hungry sink (the brain). Synaptic complexity, based on interaction with the environment, evolves into an abstract mirroring. Spontaneous meaning generation is a fundamental character of intellect. Because the Carnot engine can represent sensory perception and response (Deli, et al., 2018; Deli, et al., 2021), response to a stimulus emerges from the brain's self-regulation to restore the resting state.

The processing of the stimulus generates memory, which lends the predictive ability to cognition (Bubic, et al., 2010; Fingelkurts, et al., 2015). The principle of least action ensures a minimal energy conformation when moving in space, but intelligent systems optimize their action repertoire between the past and the future (Deli, et al., 2020). Intellect then is a temporal orientation—increasing future freedom of action. Therefore, intellect is a highly personal consideration of short and long-term consequences.

AI systems transform inputs into outputs by manipulating the data as it moves through the layers (Maass, et al., 2019). Like the human mind, the deep neural network (DNN) accumulates intellect from experience by extracting higher-level features from the raw input. However, noisy data can fool the neural network's internal representation. Therefore, the network's "black box" requires deeper understanding (Alain, et al., 2017).

2. Discussion

3.1. Holographic considerations (the field representation of complexity)

The second law of thermodynamics is a fundamental law of physics. Entropy maximization might be a general phenomenon in cosmology, computer science, and animal behavior (Wissner-Gross, et al., 2013; Cerezo, et al., 2018; Cerezo, et al., 2018). Entropy maximization may also be behind intellect's ability to increase future freedom of action (Deli, 2020.a; Wissner-Gross, et al., 2013; Ryan, et al., 2017).

Response to a stimulus depends on an intimate understanding of the environment. Therefore, mental abstraction of the physical environment is an abstract mirroring; the mind is 'about' the world by being foremost about itself. The brain forms the mind by adopting the physical laws of the environment. Information compression (Tsao, 2018; El-Kalliny, et al., 2019) generates a holographic, high fidelity temporal manifold (Saaty, et al., 2017). It is a holographic, orthogonal transformation (Libby, et al., 2021) that generates a temporal orientation, leading to the predictive nature of intellect (Fingelkurts, et al., 2015; Buzsaki, et al., 2013). Thus, the synaptic map represents a stable field, which accumulates memory (Figure 1).

The brain forms a resting state, the basis of the constancy of the sense of self. The same organization must also characterize AI. In the computer brain, like in the organic brain, the intelligent response depends on continuous energy balance changes to recover equilibrium. Because transmission strengthens connections between units, subsequent activations require less energy, giving rise to segregated, hierarchical, and modular structures. In the following, we approach artificial networks from thermodynamic principles (Deli, et al., 2020; Deli, et al., 2021).

3.2. Thermodynamic considerations

In physics, exothermic processes make endothermic ones possible (Cox, 1979). Intelligent computation belongs to the second category. Endothermic processes control the future by boosting mental abilities (Deli, et al., 2018; Fry, 2017; Wissner-Gross, et al., 2013). Thus, intellect is the evolution of the synaptic map through the thermodynamic regulation of the cortical brain.

Information science reinterpreted entropy as the "amount of information" contained in a variable. Nevertheless, information consumption embeds the brain within the environment's energy-information cycle. Sensory perception is an automatic and involuntary process. For example, reading road signs is instinctive because sensory stimuli push the brain's energy balance out of equilibrium. Subsequently, perception (response) results from self-regulation, restoring the high entropy (resting) state. Therefore, resting homeostasis is an entropic requirement. Because the evoked states build on the resting potential, they can be analyzed using thermodynamics (Deli, et al., 2018; Deli, et al., 2020; Deli, et al., 2021).

The sensory system compresses spatial stimulus into a holographic representation. Data compression increases information density (Gao, et al., 2017; Shwartz-Ziv, et al., 2017; Baumeler, et al., 2019). The orthogonal transformation (Deli, 2020.a; Deli, 2020.b; Libby, et al., 2021) integrates temporally distant identities into a subjective observer state (Bubic, et al., 2010; Fingelkurts, et al., 2015). Therefore, while temporal variability affords high degrees of freedom, the transformation of sensory data by an orthogonal morphing generates stability for experience and memory (Libby, et al., 2021), independent from sensory disturbance.

Recently an offline sleep-like phase was introduced in a neural network (Tadros, et al., 2020). Similar to biological brains, sleep modifies synaptic weights according to spike-timing-dependent plasticity rules (Sakai, 2020). It enables so-called forward transfer learning when a task shares some properties with the previously trained tasks. Introducing sleep into artificial neural networks could counter catastrophic forgetting by recovering forgotten tasks. Similarly, resting dynamics strengthen basic network features. Furthermore, high entropy oscillations' ability to access microstates (generating configurations with equal likelihood) increases the degrees of freedom or the freedom of thought.

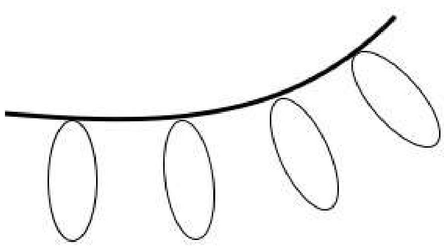


Figure 1. Mental field evolution The thermodynamic cycle (indicated by ovals) represents a fluid, chaotic and reversible process. The energy field (solid line) represents the increasing complexity of the neural system.

3.3. Intelligent computation via the reversed Carnot cycle:

The Carnot cycle is an idealized theoretical process between two heat reservoirs with vanishing net entropy production. The theoretical framework of the reversed Carnot cycle can explain the brain's evoked cycle and the "black box" of the DNN (Figure 2). The reversed Carnot cycle absorbs heat (information) from a low-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2022 doi:10.20944/preprints202205.0134.v1

temperature reservoir (environment) and delivers it to a hot reservoir (the internal organization of the DNN), which requires work input. The product ΔTΔH (temperature and entropy) is the cycle's area and the energy E required to complete a processing cycle. With increasing temperature T, decisions become more random, whereas lower temperature allows deterministic decision functions of measured data. The efficiency is determined by the average connection strength between the states; therefore, learning takes many cycles.

(A-B) Attention compresses incoming sensory information (adiabatic compression): The information input is an isothermal heat transfer via a preferred gradient. Because the average energy transferred per interaction is the connection strength times the frequency, the average connections' strength (information transfer density) between neurons correlates with temperature. The frequencies reflect the computational limit and the energy need of synaptic changes. Evoked activation channels compress information, reducing the input size and creating a high-density manifold (Saaty, et al., 2017).

(B-C) Rejection of inconsequential information (isothermal compression): Data compression rejects noise (waste heat), which collapses entropy. The synaptic map condenses information into a holographic, temporal representation. Memory formation is an orthogonal transformation of information. Landauer's principle determines the erasure cost (the work value of information) (Deli, 2021).

(C-D) The spreading of the relevant signal (isothermal heat rejection): Lower frequencies represent a response. The slowdown lowers the temperature, which corresponds to a temporal expansion.

(D-A) Resting (isentropic) expansion: The brain's autonomous regulation restores the high entropy resting state, an active memory-enhancing process. The sensory system and the brain are ready for new input. Thus, the resting condition is necessary for thermodynamic reasons because it prepares the neural system for sensory processing.

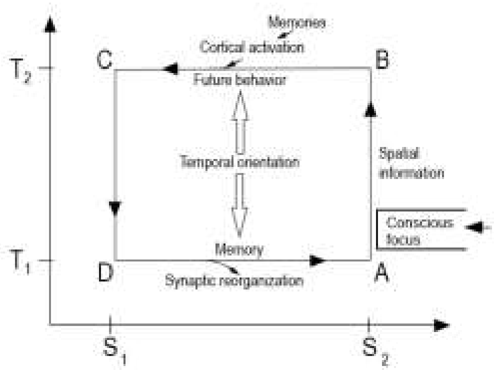


Figure 2. The reversed Carnot cycle representation of the brain's evoked cycle The cycle operates between the high frequency, evoked state (T2), and the resting state (T1); the horizontal axis represents entropy. First, stimulus increases frequencies, compressing information as a function of conscious focus (AB). The electric flow reorganizes the synaptic potentials based on top-down memory (BC). Third, electric flow reversal formulates a response (CD). Finally, self-regulation recovers the high entropy resting state, readying the brain to receive new data (DA).

3.4. The Landauer's principle

Information is physical. Landauer has shown that transformation between energy and information has energetic consequences (Landauer, 1961; Landauer, 1991). It requires energy to erase information from the

 doi:10.20944/preprints202205.0134.v1

computing device; therefore, erasing a bit of information releases a minimum heat. Landauer's principle is the ejection of (all or part of) some correlated information from a controlled, digital form (e.g., a computed bit) to an uncontrolled, non-computational form, i.e., as part of a thermal environment. An irreversible, permanent increase in entropy of log 2 = k ln 2 is an unavoidable and mathematically rigorous consequence of Landauer's principle.

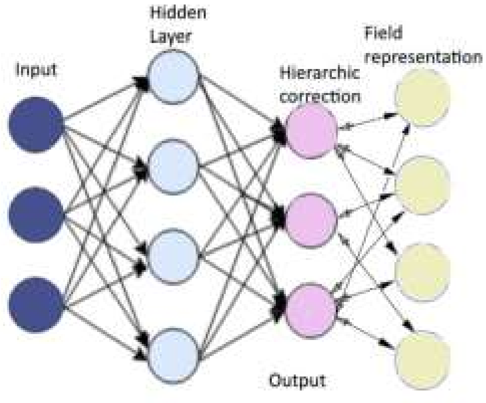
Intelligent computation is the transformation of information into representation and meaning. The erasure cost (the work value of information) is the string's best compression (Baumeler, et al., 2019). Nevertheless, representation depends on a second relaxation step, manifested as an irreversible slowing down (Bennett, 1982). In cyclic processes, the relaxation time is on the order of the operation time.

In gradient descent, slowing down, which represents temporal expansion (Tozzi, et al., 2016), reduces the error. Therefore, the resulting orthogonal transformation may stretch the spatial signal across the time domain. The view here is that this "erasure phase" is a "reset phase" required by the system to acquire new information and begin the next cycle. The transformation enriches mental complexity. Because information storage is analogous to a temperature decrease via "phase transition," the phase energy requirement limits achievable computational efficiency.

The synaptic organization represents a mental model, which makes intelligent decision-making possible. For example, the orthogonal transform might arise via infinitesimal rotations by the hidden layers, which thus condense information into a holographic state (Scheffer, et al., 2009) of the network connectivity. Percolation theory (Broadbent, et al., 1957), where the increasing number of links in a network generates globally connected node clusters, might serve an analogy.

In deep learning, information input increases entropy via Brownian motion. Landauer's principle dictates that information enrichment generates temperature in closed systems, but information processing reduces temperature and drives the cycle. Information erasure requires work on the system, dissipated as heat to the environment. The necessary amount of work is determined by our uncertainty about the system — the more we know about the system, the less it costs to 'erase' it. Although these processes' exact mechanism is not understood, information transformation can explain intelligent computation in animals and AI systems.

A Siamese network consists of two parallel and similar output vectors with the same weights (Chicco, 2020). In contrast, a hierarchic network consists of an input system and a relatively stable (i.e., orthogonal) field (Libby, et al., 2021). The stability of thefield representation can influence recognition and decision-making in a top-down manner. Therefore, the output represents the hidden layer outcome, influenced by the field vector in a top-down manner (Figure 3). Ultimately, the cycle's highly fluid operation engenders a gradual field evolution.



 doi:10.20944/preprints202205.0134.v1

Figure 3. Hierarchic representation of the network The input arrives on the left. Information processing occurs in a highly fluid thermodynamic cycle. The field influences the representation via a top-down regulation. Output on the right represents the response.

3. Conclusion

We sought a deeper understanding of AI's "black box" by examining sensory interaction as an energy-information exchange. As a stimulus activates cortical neurons, local electromagnetic imbalances preserve the global charge neutrality, and response to stimulus recovers the resting equilibrium. Therefore, the resting state readies the system to absorb new information.

The endothermic reversed Carnot cycle describes intelligent computation. The cycle's efficiency depends on the two different information content conditions it operates. Information compression is always followed by removal, which expands the system, restoring the initial conditions. The brain's self-regulatory drive for equilibrium boosts future freedom of action by generating a mental evolution.

Information processing occurs in two hierarchical networks, where the first network's thermodynamic cycle processes information. The input network can update the second network, the so-called field. The dynamic and reversible entrance network and the relatively stable field mutually determine the response. Experience accumulates by the stable field as memory and ensures intelligent decision-making, which increases future freedom of action.

Funding

This research received no external funding.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Bibliography

Alain G. and Bengio Y. Understanding intermediate layers using linear classifier probes [Journal] // ArXiv, abs/1610.01644. - 2017.

Baumeler Ä. and Wolf E. The Free Energy of a General Computation [Journal] // Phys. Rev. - 2019. - 5-1 : Vol. 100. - p. 052115.

Bennett C. H. The thermodynamics of computation [Journal] // Int. J. Theor. Phys. - 1982. - Vol. 21. - p. 905.

Broadbent S. and Hammersley J. Percolation processes I. Crystals and mazes [Journal] // Mathematical Proceedings of the Cambridge Philosophical Society. - 1957. - 3 : Vol. 53. - pp. 629–641.

Bubic A., Yves von Cramon D. and Schubotz R. I. Prediction, cognition, and the brain [Journal] // Front. Hum. Neurosci.. - 2010. - 25 : Vol. 4.

Buzsaki G., Logothetis N. and Singer W. Scaling brain size, keeping timing: Evolutionary preservation of brain rhythms [Journal] // Neuron. - 2013. - pp. 80(4), 751–764.

Cerezo S. H. and B. Guillem D. Fractal AI: A Fragile Theory of Intelligence [Journal] // ArXiv:1803.05049. - 2018.

Cerezo S. H., B. Guillem D. and B. Spiros Solving Atari Games Using Fractals And Entropy [Journal] // ArXiv:1807.01081. - 2018.

Chicco D. Siamese neural networks: an overview [Book Section] // Artificial Neural Networks, Methods in Molecular Biology, 2190 (3rd ed.). - New York City, New York, USA : Springer Protocols, Humana Press, 2020.

Cox R.T. Of inference and inquiry, an essay in inductive logic. [Journal] // In Proceedings of the Maximum Entropy Formalism First Maximized Entropy Workshop. - 1979. - pp. 119–168..

Deli E. and Kisvarday Z. The Thermodynamic Brain and the Evolution of Intellect: The Role of Mental Energy [Journal] // Cognitive Neurodynamics. - 2020. - DOI: 10.1007/s11571-020-09637-y. - pp. 14, 743–756.

Deli E. CAN THE FERMIONIC MIND HYPOTHESIS (FMH) EXPLAIN CONSCIOUSNESS? THE PHYSICS OF SELFHOOD [Journal] // Activitas Nervosa Superior. - 2020.a. - DOI: 10.1007/s41470-020-00070-4.. - pp. 62, 35–47.

Deli E. THE THERMODYNAMIC IMPLICATIONS OF THE FERMIONIC MIND HYPOTHESIS (FMH) [Journal] // Activitas Nervosa Superior. - 2020.b. - DOI: 10.1007/s41470-020-00074-0. - pp. 62(3), 96-103..

Deli E. Thermodynamic Computation In Self-Evolving Systems [Journal] // Gerlingen: GmBh. doi: 10.13140/RG.2.2.17066.13769. - 2021.

Deli E., Peters J. and Kisvarday Z. The thermodynamics of cognition: A Mathematical Treatment [Journal] // Computational and Structural Biotechnology Journal. - 2021. - Vol. 19. - pp. 784-793.

Deli E., Peters J. and Tozzi A. The Thermodynamic Analysis of Neural Computation [Journal] // J Neurosci Clin Res. - 2018. - p. 3:1.

El-Kalliny MM [et al.] Changing temporal context in human temporal lobe promotes memory of distinct episodes [Journal] // Nat commun. - 2019. - 1 : Vol. 10. - p. 203.

Fingelkurts A. A. and Fingelkurts A. A. Operational architectonics methodology for EEG analysis: theory and results, neuromethods, Ed. Sakkalis V. . [Book Section] // Operational architectonics methodology for EEG analysis. - [s.l.] : springer, 2015.

Fry R. Physical Intelligence and Thermodynamic Computing. [Journal] // Entropy. - 2017. - pp. 19, 107.

Gao X. and Duan LM. E fficient representation of quantum many-body states with deep neural networks [Journal] // Nat commun. - 2017. - Vol. 8. - p. 662.

Landauer R. [Journal] // Physics Today. - 1991. - Vol. 44. - p. 23.

Landauer R. Irreversibility and heat generation in the computing process [Journal] // IBM Journal of Research and Development. - 1961. - 3 : Vol. 5 . - pp. 183–191.

Libby A. and Buschman T.J. Rotational dynamics reduce interference between sensory and memory representations [Journal] // Nat Neuro-sci . - 2021. - 24. - pp. 715–726.

Maass W. [et al.] Brain Computation: A Computer Science Perspective [Book Section] // Steffen B., Woeginger G. (eds) Computing and Software Science. Lecture Notes in Computer Science. - Cham : Springer, 2019.

Ryan R. M. and Deci E. L. Self-Determination Theory: Basic Psychological Needs in Motivation Development and Wellness [Book]. - New York, NY : Guilford Press, 2017.

Saaty T. L. and Vargas L. G. Origin of neural firing and synthesis in making comparisons. [Journal] // European Journal of Pure and Applied Mathematics. - 2017. - pp. 10(4), 602–613.

Sakai J. Core Concept: How synaptic pruning shapes neural wiring during development and, possibly, in disease [Journal] // Proc Natl Acad Sci U S A. - 2020. - Vol. 117. - pp. 16096-16099.

Scheffer M. J. [et al.] Early-warning signals for critical transitions [Journal] // Nature. - 2009. - Vol. 461. - pp. 53–59.

Shwartz-Ziv R. and Tishby N. Opening the black box of deep neural networks via information [Journal] // https://arxiv.org/pdf/1703.00810.pdf. - 2017.

Tadros T. [et al.] Biologically inspired sleep algorithm for increased generalization and adversarial robustness in deep neural networks [Journal] // ICLR. - 2020.

Tozzi A. and Peters J. F. A topological approach unveils system invariances and broken symmetries in the brain [Journal] // Journal of Neuroscience Research,. - 2016. - 5 : Vol. 94. - pp. 351–365.

Tsao A. et al. Integrating time from experience in the lateral entorhinal cortex [Journal] // Nature. - 2018. - Vol. 561. - pp. 57–62.

Wissner-Gross A. D. and Freer C. E. Causal Entropic Forces. [Journal] // Phys. Rev. Lett.. - 2013. - doi: 10.1103/PhysRevLett.110.168702. - p. 110:168702.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2022 doi:10.20944/preprints202205.0134.v1