How the Relationship Between Information Theory and Thermodynamics Can Contribute to Explaining Brain and Cognitive Activity: An Integrative Approach

Guillem Collell
Research Unit in Cognitive Neuroscience, Dptm. Psychiatry & Forensic Medicine
Universitat Autònoma de Barcelona
Bellaterra (Cerdanyola del Vallès), Spain
g.collell@student.maastrichtuniversity.nl

Jordi Fauquet
Dptm. Psychobiology & Methodology of Health Sciences
Universitat Autònoma de Barcelona
Bellaterra (Cerdanyola del Vallès), Spain
Jordi.Fauquet@uab.cat

***Abstract*— The brain is both a thermodynamic system and an information processor. Cognition is described well in terms of information-based models and brain activity as a physical process, is accurately addressed via a thermodynamic approach. A connection between information theory and thermodynamics in neuroscience is currently lacking in the literature. The aim of this paper is to propose an integrative approach regarding information and energy as two related magnitudes in the brain, and to discuss the main connections between information theory and thermodynamics that may be helpful for understanding brain activity. In this sense, the link between both approaches is based on the concepts of entropy and negentropy, the Boltzmann formula, the Landauer's Principle and the energetic cost for the observation of information proved by Szilard. This set of connections enables us to show that information and energy are two strongly related and interchangeable magnitudes in the brain with the possibility of making this relationship explicit, as well as the possibility of translating the quantities from one to the other. This view also contributes to a better understanding of the fundamental relationship between cognition and physical brain activity. Finally, we propose new conjectures and future lines to work concerning the study of spontaneity of the brain activity from this integrative perspective.**

Keywords-entropy; negentropy; brain thermodynamics; cognitive models; information theory.

I. INTRODUCTION

The brain is both a thermodynamic system and an information processor. Hence, it is indeed subjected to the main constraints of both theories, i.e., the second law of thermodynamics as well as Shannon's source coding theorem, which states that a message cannot be compressed below its entropy bound [1]. It is well known that there exists a set of connections between information theory and thermodynamics. These are essentially its respective entropies and negentropies, as well as the Boltzmann formula, the Landauer's Principle [1][11] and the energetic cost for the observation of information proved by Szilard [2].

Several attempts have been made to find a tenable explanation for brain activity from the framework of thermodynamics, as well as from the information theory perspective [3]. The first perspective describes well the physical brain activity, and the second is able to explain several cognitive aspects (perception, learning, action, active inference, etc.) [4][5]. On the one hand, we consider the contribution of Prigogine [6] to the field of the thermodynamics of the open dissipative systems, in which the brain fits well. In this sense, the concept of *negentropy* is essential [7], and its flow can be expressed by an ordinary differential equation that describes the energetic income/outcome of the brain [8]. Following the thermodynamic stream, La Cerra [9] proposed a physically-principled model of the mind, embedded in an energetic framework. It regards the brain as the machine in charge of ensuring by means of adaptive behavioral responses the optimal ratio of costs/benefits in energetic terms for the entire system. From this perspective, the second law of thermodynamics is considered the main principle that rules human brain activity. On the other hand, Friston [4] presents an information-based brain model embedded in Bayesian variational analysis and hierarchical generative dynamics. This model describes well the learning, perception and inference cognitive processes. From this viewpoint, it is legitimate to expect that if there exists a set of connections between thermodynamics and information theory, these must preserve the validity of the statements and principles that underlie each of the above models, and will show, indeed, a high degree of consistency when we transform the quantities and magnitudes from one model to the other by using such links. That is to say, if a certain model predicts the change of the system in one direction basing its prediction on energetic measures, we expect to find the same outcome in an information-based model if we translate these energetic quantities into its equivalent information magnitudes.

In Section II, we expose the main connections between thermodynamics and information theory that may be useful for understanding the exchange between energy and information in the brain. In Section III, we introduce, analyze and connect some of the most relevant brain and cognitive models formulated from the scope of thermodynamics as well as from the information theory viewpoint. Thereafter, we give insight for a new integrative approach to the brain activity considering the points commented before. Finally, we make some conjectures referred to the possibility of using this new view of brain modeling in order to study the spontaneity of the cognitive and brain processes.

 ISBN: 978-1-61208-340-7

II. CONNECTIONS BETWEEN INFORMATION THEORY AND THERMODYNAMICS

A selection of the connections between information theory and thermodynamics that may be helpful for understanding brain activity is presented in this section.

A. The thermodynamic and Shannon's entropies and negentropies

A core concept appears in both theories, the so-called *entropy*. The meaning of the thermodynamic entropy and Shannon's entropy -from information theory- is not exactly the same, but the intuitive ideas behind both are very similar, as Boltzmann formula evidences.

The meaning of the *physical entropy* S is essentially the degree of disorder of the system, namely the energy dissipated in the form of molecular vibration that cannot be used to produce work. Its difference ΔS between two states a_1 and a_2 can be computed as follows [10]

$$\Delta S = S_1 - S_2 = \int_{a_2}^{a_1} \frac{\delta Q}{T} \quad (1)$$

where Q is the heat function and T the absolute temperature of the system.

From entropy, its complementary concept appears naturally, namely *negentropy* or *free energy* [7]. It corresponds to the amount of energy that can be used to produce work. For instance, there is negentropy present in a system that has an electrical or a pressure gradient, like a neuron before spiking. In the physical sense, it can be described as the Helmholtz free energy F_H [6]

$$F_H = U - TS \quad (2)$$

where U is the internal energy of the system.

Before describing Shannon's entropy, we must first define the *information I* contained in one character x_i [11]:

$$I(x_i) = -\log_2 p(x_i) \quad (3)$$

where $p(x_i)$ is the probability of the occurrence of the character x_i.

The information expected in one message X composed of one character from the set of all possible $\{x_i\}$, i=1,...,n is known as *Shannon's entropy H:*

$$H(X) = -\sum p(x_i) \log_2 p(x_i) \quad (4)$$

B. The Boltzmann formula

The first statistical formulation of thermodynamic entropy was provided by Boltzmann in 1877, giving an absolute interpretation of this quantity (no longer as a measure of the entropy difference between two states). Historically, this work was done about 70 years before Shannon's entropy was introduced. In particular, the latter was thought of an abstraction of the former.

The physical entropy S of a system can be computed as follows [10]

$$S = k \cdot \ln(W) \quad (5)$$

where W is the number of possible microstates of the system (assuming that all are equiprobable) and *k* is the Boltzmann constant (approximately $1.38 \cdot 10^{-23}$).

C. Landauer's principle and the information observation

Szilard [2] proved that there is a minimum energetic cost of $T \cdot k \cdot ln(2)$ J (or equivalently, an increase of physical entropy of $k \cdot ln(2)$ J/K) that every system must pay in order to observe one bit of information, namely if we want to know whether it is a "0" or a "1". Here, *k* is the Boltzmann constant. Brillouin generalized this principle stating that negentropy can always be transformed into information and vice versa [12].

Likewise, Laundauer's principle states that the same minimal quantities are to be paid for encoding (or erasing) the same bit of information [1]. This energetic cost for both procedures is due to the very nature of them, which are, in turn, the most optimal possible.

III. THE THERMODYNAMIC AND THE INFORMATION-BASED MODELS OF THE BRAIN AND COGNITION

A selection of thermodynamic models is considered for our work. These can be regarded together as a simple set of equations and principles that describes brain activity in energetic terms. Thereafter, we briefly illustrate the main ideas of Friston's model [4], which describes cognition in terms of information theory. Since it is not the aim of this paper, we will leave out the mathematical details of this theory and will focus on the intuitive concepts behind it. Nevertheless, we want to stress that the mathematical treatment of this model is fundamental in order to connect the two theories. Moreover, an important part our current work is based on the mathematical definitions of the respective entropies and free energies present in both theories.

A. Thermodynamic models of the brain activity

A fundamental constraint to which all thermodynamic systems are subject is the *second law of thermodynamics*. It states that the entropy for every isolated system can only increase (except for small random fluctuations, according to its probabilistic formulation). Thus, this principle restricts the *spontaneous processes* (i.e., those that occur without external help) to only one possible direction, namely the one which implies a gain of entropy, or, equivalently $dS/dt \geq 0$. By using the above links, the second law of thermodynamics can be reformulated into informational terms as: "our information about an isolated system can never increase (only by measurement can new information be obtained). Reversible processes conserve, irreversible ones (mostly spontaneous) loose information." [13]

The brain is a thermodynamic device characterized by being an open dissipative system, isothermal, isobaric and with a constant flow of negentropy [8]. In order to study the entropy exchange between the inside and outside of the

brain (and the same can be applied to a single neuron or every biological system), we must split the dS into two terms, making the distinction between the entropy produced within the system $d_i S$, and the transfer of entropy across its boundaries $d_e S$ [6]. Therefore, since the second law of thermodynamics must hold, after every cognitive task $d_i S > 0$ is always obtained. This initial increase of entropy within the system is always followed by a removal of entropy through its frontiers in order to preserve the structure and functionality [3][9]. Moreover, so as to recover the capacity of producing work (e.g., to transmit an electric impulse) there must be an inflow of negentropy (adenosine triphosphate (ATP), for instance). Hence, this energy/entropy flow for the brain can be expressed by the following ordinary differential equation [8] obtained by differentiating (2).

$$\frac{dF_H}{dt} + T\frac{d_i S}{dt} = J_{i1} + J_{c1} - J_{i2} - J_{c2} \quad (6)$$

Here, J_i and J_c denote the radiation and chemical energy flows respectively, and the subscripts 1 and 2 refer to the incoming and out coming flows [8].

From this perspective, the brain activity is driven by the quest to consume free energy in the least possible time [14]. That is to say, the flows of energy (i.e., electric activity) themselves will search for the paths of transduction, selecting those that consume free energy in the least time. The transduction paths are established when an experience, encoded in some sort of energy, is recorded. These paths for energy dispersal within the neuronal network constitute memory, i.e., the register of remembrance that can be consolidated or reorganized when a certain path is activated again by an energy flow.

B. Information-based model of cognition

Friston's model is embedded in machine learning framework and regards the brain as an inference machine ruled by the "free energy minimization principle". Here the so-called free energy is defined as an upper bound for what Friston defines as "surprise", i.e., how unlikely it is for the system to receive a certain input. In short and skipping mathematical content, the free energy function F can be expressed conceptually as:

(i) F = Cross Entropy + Surprise

At the moment, we just notice that the "cross entropy" is a positive term (the so-called Kullback-Liebler divergence). Equivalently, by algebraic rearrangement of its mathematical formulation the following equation is obtained

(ii) F = Expected Energy – Entropy

where "expected energy" corresponds to the surprise of observing an input jointly with its cause, and "entropy" is equivalent to the Shannon's entropy of the variable causes (i.e., what caused input). Interestingly, notice that (ii) resembles the Helmholtz free energy in the physical sense.

Via free energy minimization, neuronal networks [4] always tend to optimize the probabilistic representations of what caused its sensory input, and therefore the prediction error is minimized. This free energy minimization in the system can be performed by changing its configuration to change the way it samples the environment, or by modifying its expectations [4]. Therefore, hypothetically the system must encode in its structure a probabilistic model of the environment, and the brain uses hierarchical models in order to construct its prior expectations in a dynamic context. In this sense, this model is capable of explaining a wide variety of cognitive processes.

C. Connections between the two approaches

A close look at Friston's and La Cerra's models yielded a finding of the following functional similarities:
(1) The system (brain) must avoid phase transitions, i.e., drastic changes in its structure and properties, and (2) the system encodes a probabilistic relationship between: the internal state of the organism, specific sensory inputs, behavioral responses, and the registered adaptive value of the outcomes of these behaviors.

Interestingly, we notice that the principles that underlie both the thermodynamic model of the brain and the information-based model of cognition are consistent in the sense that they steer the system in the same direction. Let us consider the case where both the Friston's and the thermodynamic free energies are large within a certain neuronal population. In the Friston's case, this indicates that the input received is highly "surprising" (e.g., facing a novel or dangerous situation). According to this model, the system will react by taking some action as well as by updating its perception, yielding to free energy minimization ($\Delta F < 0$). The same will occur under the same situation if we think in terms of thermodynamics models. The physical free energy is also large due to the initial blood and energetic inflow that accompanies a novel situation, but this amount will be rapidly reduced by the activation of neuronal paths that consume free energy in the least possible time ($\Delta F_H < 0$). By doing so, the brain is carrying out a codification task in the sense that a smaller amount of free energy will be necessary to activate the same neuronal path on subsequent occasions. Analogously, Friston's model prescribes a subsequent prediction error reduction (i.e., a better encoding of the causes of the input) following this free energy minimization.

In order to provide some insight into how the information must be encoded, let us suppose that we are set the task of memorizing a random number between 0 and 256. It is obvious that we have to encode it somewhere in our neuronal network. Due to Shannon's source coding theorem, we cannot codify it using fewer bits than the information contained in it. Thus, at least 8 bits ($I = \log_2(256) = 8$ bits), namely 8 "places" in our brain circuitry with 2 possible combinations in each one. Of course, we must consider that some numbers can be compressed by using some heuristics.

 ISBN: 978-1-61208-340-7

The former statement can be generalized to every behavior or cognitive task, considering that these are nothing but an "algorithm" that our brain has to encode in our neuronal network, even if it is not a permanent, but rather, a temporal encoding in our working memory.

Clearly, the Szilard and Landauer principles shed some light on the theoretical bounds for the exchange between information and energy present in the brain activity. Nevertheless, these are not realistic quantities in this context, since the encoding mechanisms in the brain are far from being entirely efficient. Moreover, there are also important noise and redundancy factors [15]. Laughlin has shown that considering a consumption of 10^6 ATP molecules for each neuronal spike and the fact that a sensory spike carries between 1 and 10 bits of information, then, the metabolic cost for processing 1 bit of information is about to 10^5 ATP molecules [16], or equivalently $5 \cdot 10^{-14}$ Joules. This means that the human brain is operating about 10^7 times above the thermodynamic limit of $k \cdot T \cdot ln(2)$ J per each bit encoded, which is still more efficient than the modern computers. We want to highlight the fact that these quantities refer to the energetic cost of visual perceptive processing, and therefore, these cannot be extrapolated neither to all the brain areas nor to all cognitive processes, since the redundancy and noise factors are topographic and population-size dependent [15]. In general, the larger the neuronal ensembles, the greater the number of redundant interactions between neurons. We hypothesize that the exchange between energy and information in the brain is performed at a constant rate, similar to Landauer's limit, but this constant depends on the redundancy factor of a particular neuronal ensemble as well as on the type of neurons it contains. This limit would be given by the very nature of the neuronal encoding procedures. Besides, it is clear that higher level visual areas like for instance the MT (responsible of motion perception) are processing, at most, the whole amount of bits received from the sensory neurons plus some extra information that is already stored in our neuronal network (i.e., our pre-knowledge or expectations about the world) which is necessary to integrate the new input. For our conscious and deliberate thinking (e.g., solving a problem that affects our life), there is no external sensory information to integrate, and therefore all the information processed comes from the internal representations stored in our neuronal network. From this, we can conclude that a measure of the energetic expense during a certain cognitive task obtained by means of electroencephalography (EEG) or functional magnetic resonance (fMRI) (ideally a combination of both) would provide a reliable measure of the processing complexity of this task, as well as about its information content. For the conversion from observed energy into its information content we must have previous knowledge about the redundancy factor of the involved brain areas, as well as its energetic cost of processing (as Laughlin computed for the visual sensory coding). Conversely, the model could also be checked the other way around, that is estimating first the necessary information to compute a certain cognitive operation (e.g., a mathematical mental computation) and then comparing the energy expense seen in fMRI and/or EEG during the task with the predicted quantity. From machine learning and computational approaches, the prediction of the information content of a certain cognitive operation could be obtained by computing an equivalent algorithm which carries out the same process in a computer. Indeed, an integrative model must incorporate the referred energetic-informational correspondence so as to allow making predictions by using both magnitudes. We suggest that the empirical validation of the model should be done in the direction described in the above lines.

We consider that a challenging aspect of the neuronal activity to explore is the concept of spontaneity. Namely the direction in which the transformations of the system occur, and which necessary (but not sufficient, due to the non-deterministic inherent character of brain activity) conditions are to be held for the occurrence of these changes. Currently, the Gibbs free energy function G describes this well in terms of thermodynamics, being the spontaneity possible only if $\Delta G < 0$ [10], where $\Delta G = \Delta H - T\Delta S$ and H is the enthalpy. However, with the previous considerations of the interrelationship between information and energy magnitudes, we consider the possibility of enhancing the characterization given by Gibbs into spontaneous processes in the brain by adding information quantities.

IV. CONCLUSIONS AND FUTURE WORK

A first sketch of an integrative approach to studying brain activity was presented here. We provided a set of useful tools, namely the links between information theory and thermodynamics that can be used to connect the cognition described in informational terms with the physical activity of the brain modeled by thermodynamics. Nevertheless, from a theoretical point of view, there are still several points to be connected between the two main models, such as, for instance, the free energy present in Friston's model with the thermodynamic measure equally termed. Our main current work is focused on further elaborating the connections between the thermodynamic and information-based models of the brain, aiming to present the explicit form of the equations that permit us to relate and unify both theoretical approaches in the near future. The possibility of validating the model empirically in the future in the way we indicated above is suggested. We also aim to inspire more researchers to contribute further to this new integrative approach, as well as to empirically studying its consequences.

REFERENCES

[1] L. del Rio, "The thermodynamic meaning of negative entropy," Nature 476, August 2011, pp. 61–63.

[2] L. Szilard, "On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings," Z. Phys., vol. 53, 1929, pp. 840.

[3] L.F. del Castillo, "Thermodynamic Formulation of Living Systems and Their Evolution" JMP, vol.2, 2011, pp. 379-391.

 ISBN: 978-1-61208-340-7

[4] K. Friston, "The free energy principle: a unified brain theory?," Nature Neuroscience, vol. 11, 2010, pp. 227–138.

[5] M. Lee, "How cognitive modeling can benefit from hierarchical Bayesian models," Journal of Mathematical Psychology vol. 55, 2011, pp. 1–7.

[6] I. Prigogine, "Time, Structure and Fluctuations," Science, vol. 201, Issue 4358, September 1978, pp. 777–785.

[7] E. Schrödinger, What is life?, Trinity College, Dublin, February 1943.

[8] J. Kirkaldy, "Thermodynamics of the human brain," Biophys. J. 1965.

[9] P. La Cerra, "The First Law of Psychology is the Second Law of Thermodynamics: The Energetic Evolutionary Model of the Mind and the Generation of Human Psychological Phenomena," Human Nature Rev. vol. 3, 2003, pp.440-447.

[10] R. Feynman, The Feynman Lectures on Physics . 3 vol. 1964.

[11] R. Feynman, Feynman Lectures on Computation, Addison-Wesley Longman Publishing Co., Inc. Boston, US, 1998.

[12] L. Brillouin, "The Negentropy Principle of Information," Journal of Applied Physics, vol. 24, Sep.1953, pp. 338–343.

[13] J. Rothstein "Information, "Measurement, and Quantum Mechanics," Science, vol. 114, 1951, pp. 171–175.

[14] S. Varpula, "Thoughts about Thinking: Cognition According to the Second Law of Thermodynamics," Advanced Studies in Biology, vol. 5, no. 3, 2013, pp. 135 – 149.

[15] S. Nandakumar, "Redundancy and Synergy of Neuronal Ensembles in Motor Cortex," The Journal of Neuroscience, 25, April 2005, pp. 4207–4216.

[16] S. Laughlin, R. d. Ruyter van Steveninck, and J. Anderson, "The metabolic cost of neural computation," Nature Neuroscience, vol. 1, no. 1, 1998, pp. 36–41.

 ISBN: 978-1-61208-340-7