These from Bits

Yutaka Shikano

Research Center of Integrative Molecular Systems (CIMoS),
Institute for Molecular Science,
Okazaki 444-8585, Japan.
and
Schmid College of Science and Technology, Chapman University,
Orange, CA 92866, USA.
Email: yshikano@ims.ac.jp

October 31, 2013

1 Operational Derivation of Physical Laws

When answering the question of what properties a material has, a theoretical physicist may ask

> "What is its Hamiltonian?" or "What is its Lagrangian?"

Most physicists seem to believe that every physical property of a material can be predicted once the Hamiltonian or Lagrangian of some physical phenomena are known. This is often called "physics imperialism." In the 20th century, we perhaps benefited too much from practical developments in physics—semiconductors, lasers, and magnetometers. To reinforce its position, the 20th century saw physics expanding the boundaries of various physical phenomena, from the sub-nanometer to the cosmological scale.

On the other hand, when somebody asks the same question to a non-expert physicist, they may try to break open the object with a hammer, for example, or measure its electrical properties. That is, to reveal the attributes of this material, they take a step-by-step approach. We can think of this as *operational* thinking. This method is very powerful when it comes to understanding unknown physical phenomena. Further, operational thinking is a natural process for all experimentalists. To reveal a material's physical properties, experimentalists construct their experimental setup, start the detection by flicking a switch, measure something, and then analyze the experimental data. Obviously, before the experimental setup has been constructed, we cannot collect experimental data. This is essentially a step-by-step (operational) process. In order to naturally understand physical properties via such a process, it seems to be necessary to reconstruct all physical laws from an operational point of view.

Operational thinking has been formalized as information theory. Historically, as recounted in the book "Science and Information Theory," Leon Brillouin tried to apply this theory to physical laws [1]. His book aims to capture various physical phenomena from the information-theoretical idea initiated by Claude Elwood Shannon. In particular, he tried to derive the entropy of physical systems from the information-theoretical quantity known as the Shannon entropy. From an information theory standpoint, the Shannon entropy can be thought of as the averaged rate of the optimal data compression [2]. This seems to fit the concept of John Archibald Wheeler's famous quote:

> "It from Bit."

However, as shown in the next section, information-theoretical concepts cannot be applied to a single event. In this essay, we show that this quote should in fact be rewritten as:

> "These from Bits."

2 Individuals and Information Theory

First of all, how should we evaluate the quantity of information? For example, the abbreviation "IMS"[1] has the following ASCII binary code:

IMS ⇒ 010010010100110101010011

Thus, "IMS" has a 24-bit string. However, nobody would claim that the Shannon entropy of "IMS" is 24. Furthermore, the 24-bit string alone has no meaning. For example, another abbreviation, "MIT," can be converted to

MIT ⇒ 010011010100100101010100

[1] "IMS" stands for "Institute for Molecular Science," which is the author's working institute.

This also has a 24-bit string, but the meaning of the two abbreviations is completely different. Therefore, the amount of information does not reflect the meaning of each word. So what does the amount of information express? Neither word has an information-theoretical meaning. Therefore, we have to define the amount of information for an ensemble of bit strings. For example, we could consider the set of bit strings given by "CIT," "NIT," and "TSU," and evaluate the probability distribution of the bit string pattern, e.g., the ratio of the number of 1's. However, this probability distribution cannot be evaluated from just a single event. Therefore, we require an ensemble containing a large number of samples. Then, for a sufficiently large number of samples, the probability distribution becomes the "true" probability distribution. In this case, each bit string is called a typical sequence.

For a typical sequence of N bits, Shannon analytically showed that the optimal data compression rate could be written as

$$\tilde{N} = NH(p), \tag{1}$$

where $\tilde{N}$ is the averaged number of the optimally compressed bits[2], and $H(p)$ is the Shannon information for the bit string, which is given by

$$H(p) = -p \log_2 p - (1-p) \log_2 (1-p), \tag{2}$$

where p is the ratio of the number of 1's in the bit string. Therefore, on applying information theory to physical laws, macroscopic systems, such as those of thermodynamics and statistical mechanics, are needed. Information theory cannot be applied to Newtonian mechanics and electromagnetism, as the theory breaks down for small data sets or a single event. However, in our physical experiences and daily life, such phenomena or events are commonly encountered. We must therefore construct a relevant description of information theory on this scale.

3 Equilibrium Thermodynamics and Statistical Mechanics from an Operational Viewpoint

In the previous section, we showed that information theory can only be applied to physical systems with a macroscopically large number of samples.

[2]Shannon originally showed that there exists some lower bound of the (reversible) compression process such that $NH(p) \leq \tilde{N} < NH(p) + 1$ for any N-bit string.

As is well known, the macroscopic theory of physics is described by thermodynamics and statistical mechanics. Let us first consider the structure of thermodynamics. Equilibrium thermodynamics itself has an operational perspective, and, further, it can be axiomatized by a specific operational process, namely the adiabatic process [3][3]. Therefore, the long history of thermodynamics can be placed into an information-theoretical context. The famous parallel between thermodynamics and information theory is the paradox of Maxwell's demon [4], explained as follows. Consider a molecular gas inside a box. The box contains a partition that divides it into two regions, and the partition has a window that can be either open or shut. The demon operates this window. When the demon sees molecules moving at higher speeds, he guides them to the left side of the box via the window. Similarly, the demon guides molecules moving at lower speeds to the right side of the box. The demon repeats this process repeatedly. Eventually, the temperature in the left of the box increases, and vice versa. This seems to violate the second law of thermodynamics, and was taken as the paradoxical issue. However, Rolf William Landauer pointed out that the mind of the demon retains the memory of the molecular speed, and further that the erasure of this memory must incur some cost [5]. This cost is equivalent to the gain from the physical system. Therefore, by considering not only the thermodynamical cycle but also the information cycle, the second law of thermodynamics is not violated. Further developments on the resolution of the Maxwell's demon paradox have been contributed by various researchers, particularly Charles Henry Bennett [6]. However, there remains an unsolved problem of the relationship between the thermodynamical entropy of the physical system and the Shannon entropy of the demon. In Ref. [7], we pointed out the equivalence between these entropies when the cleverest Maxwell's demon operates the physical and information-theoretical processes in a specific context. These physical processes do not incur any cost from the operation of the partition, the window, or the measurement. We can also ensure that the information-theoretical processes do not incur any computational cost in the demon's memory. Only when the cleverest Maxwell's demon applies the optimal data compression to his memory before the erasure does the Shannon entropy equal the thermodynamical entropy. Therefore, if all molecules in the box are measured by the cleverest demon, the thermodynamical entropy in all of the thermodynamical processes can be characterized by the Shannon entropy in the information-theoretical context. Hence, "These (thermodynamical processes) from Bits."

[3]The same authors recently showed that nonequilibrium thermodynamics cannot, in general, be defined in the same way [9].

Next, let us consider another macroscopic physical theory: statistical mechanics. In equilibrium statistical mechanics, we conventionally discuss a derivation of the ground state of a sufficiently large number of spins and a phase transition from liquid to solid, for example. Equilibrium statistical mechanics does not have an operational structure. Therefore, to pursue our idea that any physical process can be reformulated from an operational viewpoint, we must construct some operational scenarios. For simplicity, consider a physical system with N two-level atoms. Somebody, who we symbolize as Maxwell's demon in the following, measures each two-level atom. First, Maxwell's demon measures the N-ary physical system. The demon's memory stores the bit-string of the excited state (1) or the ground state (0), and so the demon incurs the optimal erasure cost[4] given by

$$W_{era}(p) = N\, H(p)\, k_B T \ln 2 \tag{3}$$

where p denotes the ratio of the number of excited states, k_B is the Boltzmann constant, and T is the temperature of the heat bath in the physical erasure model. From Landauer's well-known principle, the averaged cost of the erasure process is $k_B T \ln 2$. We can also determine the cost of exciting the physical system from the ground state for all two-level atoms as

$$W_{phys}(p) = Np\,\epsilon \tag{4}$$

for the two-level energy difference ϵ. Then, we define the cost function $F(p)$ as

$$F(p) := W_{phys}(p) - W_{era}(p). \tag{5}$$

Intuitively, one of the essential properties of the equilibrium state is its robustness against small perturbations to the physical system. In our operational context, we define the equilibrium state as the robustness of the cost function $F(p)$ under a small change to the physical system:

$$\frac{dF(p)}{dp} = 0 \tag{6}$$

for sufficiently large N [8]. Thus, we can derive the Maxwell–Boltzmann distribution as

$$\frac{\text{the number of } 1's}{\text{the number of } 0's} = \frac{p}{1-p} = \exp\left(-\frac{\epsilon}{k_B T}\right). \tag{7}$$

[4]We consider the optimal erasure cost because equilibrium thermodynamics can be equated to equilibrium statistical mechanics.

To conclude, we derive the Maxwell–Boltzmann distribution, which is the conventional derivation of the equilibrium state in statistical mechanics from an operational statistical process with optimal data compression and erasure processes[5]. Once again, therefore, we have "These (physical systems to satisfy statistical physics) from Bits."

4 Concluding Remarks

Following in Brillouin's footsteps, we tried to reformulate some physical theories from an operational viewpoint. However, as information theory is not currently applicable to situations where there are only a small number of samples, we could only consider macroscopic physical theories: equilibrium thermodynamics and equilibrium statistical mechanics. The optimal information-theoretical process corresponds to the equilibrium macroscopic system, and its essence is a sufficiently large number of samples. Therefore, Wheeler's famous slogan should be changed to "These from bits." To revive the original "It from bit," we must extend information theory to small-number samples or non-typical sequences. I believe that microscopic physical theories, such as Newtonian mechanics, can play a great part in the development of information theory. At such a time, "It develops Bit," and we will surely acquire "It from Bit."

References

[1] L. Brillouin, *Science and Information Theory* (Dover, Mineola, N.Y., [1956, 1962] 2004).

[2] C. E. Shannon, Bell System Technical Journal **27**, 379 (1948); 623 (1948).

[3] E. H. Lieb and J. Yngvason, Phys. Rep. **310**, 1 (1999).

[4] J. C. Maxwell, *Theory of Heat* (Longmans, Green, London) pp. 308-309 (1871).

[5] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961).

[6] C. H. Bennett, IBM J. Res. Dev. **17**, 525 (1973); Int. J. Theor. Phys. **21**, 905 (1982).

[5]Our approach is completely different from that of Jaynes [10], as seen in Ref. [8, Appendix B].

[7] A. Hosoya, K. Maruyama and Y. Shikano, Phys. Rev. E **84**, 061117(2011).

[8] A. Hosoya, K. Maruyama and Y. Shikano, arXiv:1301.4854.

[9] E. H. Lieb and J. Yngvason, Proc. Roy. Soc. A **469**, 20130408 (2013).

[10] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).