

LETTERS

Entropy Revisited, Gorilla and All

The article by Elliott H. Lieb and Jakob Yngvason (PHYSICS TODAY, April, page 32) was indeed a fresh look at entropy and the second law of thermodynamics, but I believe it did not do justice to Constantin Carathéodory who, in his axiomatic development of thermodynamics,¹ was the first to replace traditional statements of the second law (heat engines, cyclic processes). He is mentioned once in the article, but in a context that does not recognize his seminal role in laying an axiomatic foundation for thermodynamics (beginning more than 90 years ago!). The mathematical complexities of his formulation of the second law not only obscured the physical simplicity of his idea, but for some time they were a significant impediment to the use of his methods by physicists. Hans Buchdahl's exposition² of the Carathéodory formulation played an extremely important role in popularizing the axiomatic development begun by Carathéodory, so that by the late 1960s, there were several undergraduate textbooks^{3,4} incorporating his approach. Although the authors do give Buchdahl credit, there is even more reason to credit Carathéodory: He started the whole movement!

References

1. C. Carathéodory, *Math. Ann.* **67**, 355 (1909); *Sitzungsber. K. Preuss. Akad. Wiss.* **39** (1925).
2. H. A. Buchdahl, *Amer. J. Phys.* **17**, 41, 44, 212 (1949).
3. H. A. Buchdahl, *The Concepts of Classical Thermodynamics*, Cambridge U. Press, London (1966).
4. C. J. Adkins, *Equilibrium Thermodynamics*, McGraw-Hill, London (1968).

DAVID SIMINOVITCH
(siminovitch@hg.uleth.ca)
The University of Lethbridge
Lethbridge, Alberta, Canada

In connection with the interesting article on entropy by Lieb and Yng-

Letters submitted for publication should be sent to Letters, PHYSICS TODAY, American Center for Physics, One Physics Ellipse, College Park, MD 20740-3843, or by email to pletter@aip.org (using your surname as "Subject"). Please include your affiliation, mailing address, and daytime phone number. We reserve the right to edit letters.

vason, I should like to make a few comments to broaden the background and possibly to provide a slight change of emphasis.

▷ On page 34, the authors rightly emphasize extensivity as normally essential for the second law of thermodynamics. One must not overlook, however, the possibility of gravitational interactions. Their effect is that, when the system is doubled, its energy does not double, because the gravitational interaction enters as an "extra." The same applies to the entropy.

Furthermore, for large enough masses, objects become unstable and collapse (see, for example, Chandrasekhar's limit for white dwarfs), which incidentally causes a problem with the concept of the "thermodynamic limit." Also, similar systems can merge with a resulting decrease of entropy.¹

▷ On page 36, the positivity of the specific heats is noted. One might observe here that the effect of gravitation can lead to negative specific heats. See for example, ref. 2.

These effects are reminders that "systems of particles interacting by long-range forces ought to be ruled out on page 1 of any book on ('normal') thermodynamics."² Relativistic thermodynamics is of course finding a way around these difficulties. But "normal" thermodynamics rules these possibilities out (usually implicitly).

▷ Regarding the symbol for adiabatic accessibility, appropriately stressed on page 34 of the Lieb and Yngvason article, some credit might be given to Constantin Carathéodory for an early use of this concept. His work need no longer be avoided on grounds of mathematical complication, since Carathéodory's principle can now be related to Kelvin's principle, mentioned on page 33 of the article, rather simply.⁴ Proceedings of the first international conference on thermodynamics⁵ provide further citations.

▷ To the (perhaps rhetorical) question raised on page 32 of the article as to whether statistical mechanics is essential to the second law, the answer is presumably "No," since the main work of Sadi Carnot and Rudolph Clausius preceded that of Willard Gibbs.

References

1. P. T. Landsberg, R. B. Mann, *Class. Quantum Grav.* **10**, 2373 (1993).
2. R. D. Sorkin, R. M. Wald, Z. Z. Jiu, *Gen. Rel. Grav.* **13**, 1127 (1981). See also P. T. Landsberg, R. P. Woodward, *J. Stat. Phys.* **73**, 361 (1993).
3. P. T. Landsberg, in *The Study of Time III*, J. T. Fraser, N. Lawrence, D. Park, eds., Springer, New York (1978), p. 118. P. T. Landsberg, *Seeking Ultimates*, Institute of Physics, Bristol, UK (2000), p. 121.
4. P. T. Landsberg, *Nature* **201**, 485 (1964).
5. P. T. Landsberg, *Pure Appl. Chem.* **22**, 215 (1970). Also published as a separate volume, *Proceedings of the International Conference on Thermodynamics*, Butterworths, London (1970).

PETER T. LANDSBERG
(ptl@maths.soton.ac.uk)
University of Southampton
Southampton, UK

The article by Lieb and Yngvason discusses aspects of entropy and the second law of thermodynamics that are not well known or mentioned in the commonly used texts. Most of the article is devoted to a complicated presentation of the properties of entropy, concluding with the statement that they have developed an axiomatic foundation for thermodynamics and eliminated the intuitive but hard-to-define terms such as "hot," "cold," and "heat" from its development. This desirable goal, however, was achieved in a much simpler and more transparent manner by an old friend, the late Herbert B. Callen of the physics department at the University of Pennsylvania. His book¹ shows how a postulational foundation of thermodynamics makes entropy a necessary variable for the understanding of thermal, chemical, mechanical, and electrical processes. It is interesting that the authors credit Peter T. Landsberg for beginning the movement to use the second law as a pillar of physics in its own right, but no interaction between Callen and Landsberg is mentioned in the article.

To show the simplicity and beauty of Callen's approach, a brief summary of his ideas follows. He first points out that conservation of energy is not sufficient to explain many physical phenomena: A body at a uniform temperature does not spontaneously develop a gradient nor does a homogeneous chemical system separate

into components. An extensive variable, the entropy S , is needed to explain why. Callen states that

- ▷ Entropy depends on the internal energy U , volume τ , number of constituents N , and charge q of a system through a relation of the general form $S = S(U, \tau, N, q)$, which may be inverted, in principle, to give $U = U(S, \tau, N, q)$.
- ▷ Entropy is defined only for equilibrium states.
- ▷ The entropy of a system for which all differences in temperature, pressure, concentration, and so forth are allowed to equalize will go to a maximum.

Differentiating the expression for U then leads to the absolute temperature, the negative of the pressure, and the chemical and electrostatic potentials. With those definitions in hand, the first law of thermodynamics follows, and includes thermal, mechanical, chemical, and electrostatic energy. It is then a simple matter to show that conservation of energy combined with the entropy-maximizing postulate leads to well-known equilibrium conditions, such as uniformity of temperature, chemical potential, or electrostatic potential. If N and q are constant, then $dU = T dS - P d\tau$, which we may rewrite as $dU = dQ - dW$, where $+dQ$ is the amount of heat that enters a body and $+dW$ is the amount of work done by the body. This is the primitive form of the first law, and the relation $dS = dQ/T$ can be applied to a Carnot cycle to verify that all changes in an isolated system lead to an increase in the entropy.²

As the entropy of a system increases, so does a measure of its organization known as the disorder W . The connection between entropy and disorder can be expressed as

$$S = k \log_e W$$

where k is Boltzmann's constant. An ingenious single combinatorial expression by Cowan,³ plus the use of distinguishability and the exclusion principle, shows that maximizing S with Lagrange multipliers leads to the appropriate statistical distribution: Maxwell-Boltzmann, Fermi-Dirac, or Bose-Einstein. I agree with the authors, then, that the Gibbs-Boltzmann approach is clearly the wrong direction to go, but it also follows that entropy can be described and explained without all the complex mathematics.

References

1. H. B. Callen, *Thermodynamics*, Wiley, New York (1960).
2. G. Joos, *Theoretical Physics*, G. E.

Stechert, New York (1934).

3. R. D. Cowan, *Am. J. Phys.* **25**, 463 (1957).

ALLEN NUSSBAUM

(nussbaum@ece.umn.edu)

University of Minnesota, Minneapolis

I enjoyed the article on entropy and the second law of thermodynamics by Lieb and Yngvason. I have tried with little success to tell physics professors that statistical mechanics is unnecessary for understanding the second law. One professor called me a chemist, apparently his idea of a hopeless ignoramus!

Nevertheless, I find Lieb and Yngvason's gorilla model confusing, and remote from practical, physical aspects of thermodynamics. The clearest, most self-evident, and most useful explanation of entropy seems to me to fall out of the temperature-entropy diagrams used routinely by mechanical and chemical engineers. The area below the cycle on such a diagram is simply the portion of process energy that is unavailable for conversion to mechanical work. And the reason for the unavailability is simple: A thermodynamic process needs an energy inhomogeneity (typically a temperature gradient) to drive it, and most of the process energy is used to maintain the gradient, thus becoming unavailable for conversion to work—for example, by dissipation in cooling water. By dissipating energy, the process tends to destroy the gradient and stop itself, so energy is continuously supplied to the process cycle at a high level to maintain the gradient, and continuously dumped at a low level into a homogeneous energy sea, where there are no gradients to act as process drivers. Entropy is the measure of the diminution of energy inhomogeneity needed to drive processes, and the formalism of thermodynamics enables engineers to calculate the entropy change of a particular process.

The "disorder" of statistical mechanics corresponds to the homogeneity of the sea of dissipated energy, but we don't need to know this to understand the physical reality of entropy and the second law.

RICHARD H. TOURIN

New York City

Lieb and Yngvason's article is indeed a breath of fresh air on a quantity that is so often described unintelligibly by undergraduate textbooks as a measure of "disorder."

I am not convinced, however, that their approach to entropy is "independent of any statistical model—or

even of atoms." Their definition of entropy is not an operational one. To determine which of two systems has the greater entropy, we must imagine an infinite number of hypothetical processes that might transform one into the other. Only if we find such a process do we get a definite answer. If we do not know of the existence of atoms, then it is far from obvious how to compare the entropies of a grasshopper and a flower.

To operationalize Lieb and Yngvason's definition, we would have to find a generic strategy for comparing entropies, rather than creating methods ad hoc. Such a strategy exists, but it depends on the existence of atoms: We can take advantage of the additive property of entropy to subdivide a system into smaller and smaller parts until we arrive at subsystems simple enough that we can find their entropy by inspection. Only the existence of atoms gives us reason to believe that this process of subdivision will terminate with subsystems that contain small enough numbers of atoms to be analyzed trivially.

It would perhaps be more accurate to refer to Lieb and Yngvason's definition as a metadefinition that states the general properties that should be possessed by a true operational definition of entropy. It would then be a non-trivial task to arrive at the appropriate operational definition in the case of, say, a truly continuous system.

BENJAMIN CROWELL

(crowell00@lightandmatter.com)

Fullerton College

Fullerton, California

LIEB AND YNGVASON REPLY: We are happy that our article attracted the interest of so many letter writers and we appreciate their comments. Yes, we would have wished to include more references, but space limitations gave us little room for scholarship. However, our long paper¹ discusses the history and relationships of the various formulations of the second law; we hoped that readers would consult that paper. The recent paper of Jos Uffink² is also valuable.

As mentioned by Siminovitch and Landsberg, and by us in our article, Carathéodory is one of the founding fathers, and it is true that he was the first to emphasize the idea of a relation among states of macroscopic systems based on adiabatic accessibility. We regarded his ideas, which date to 1909, as well known, but perhaps the connection could have been clearer. However, although his work motivated Landsberg and the others

we mentioned, his conceptual framework is different in important respects from the circle of ideas that led to the work of Giles and then to our work.

Our work is logically divided into two parts. The first shows that entropy comes out of little more than the list of pairs of states X and Y such that one can go from X to Y without doing more to the surroundings than moving a weight. It is this fact that speaks for our approach (and Giles's book). Quasistatic paths in state space are not needed; calculus and the sophisticated mathematics of differential forms are also unnecessary in this part. It is not even necessary to parametrize equilibrium states by coordinates such as energy and volume. Nevertheless, entropy emerges together with a specific formula that determines the entropy function uniquely, except for the choice of units. Carathéodory, on the other hand, makes essential use of coordinates and differentials in state space, and entropy and temperature appear to be more delicate constructs than they naturally are. Likewise, we do not have to introduce extraneous physical principles and heat engines, as in the older approach of Carnot and successors.

What we must introduce in the first part is the notion of adiabatic accessibility, and since we do not have the help of coordinates, we must do this empirically and without mentioning heat and temperature. This is where the gorilla enters as a metaphor for the rest of the universe and its possible action on the system under discussion. Unlike Tourin, we do not find this confusing. Rather, we consider it to be an essential clarification of the kinds of processes with which the second law deals.

Many formulations of the second law have been made, and choosing among them is largely a question of taste. We submit, however, that our assumptions are easier to understand than those of Kelvin, Planck, Clausius, and Carathéodory. Callen's approach, mentioned by Nussbaum, is excellent (as is the work of Tisza and ultimately Gibbs, on which he relies), but Callen starts where we leave off. In his approach one postulates the existence and basic properties of entropy and then runs with the ball. We seek to answer the prior question: Where does entropy—and its properties, especially its temporal increase—come from? Our answer is that it comes from a relation among macroscopic equilibrium states

whose simple properties are, except for one, so obvious that they are taken for granted and rarely mentioned explicitly.

The one nonobvious property of equilibrium states is the "comparison hypothesis," which states that, given any two states X and Y of a system, either X is adiabatically accessible from Y or Y is adiabatically accessible from X . The second half of our work, which goes beyond Giles, Landsberg, and the others we mentioned, turns this hypothesis into a fact, with the aid of some reasonable assumptions. Here, for the first time, we describe states by means of energy and volume, in the usual way, and make some contact with Carathéodory's approach. We introduce an important assumption that is similar to, but significantly narrower than, Carathéodory's main one: For every state X there is another state, Y , somewhere, and not necessarily nearby, that is adiabatically accessible from X , but not the other way around. Otherwise, our approach is mathematically and physically different from Carathéodory's. It is not necessary to derive the entropy function afresh, because that was done in the first part. We simply have to derive the simpler, nonquantitative fact that any two states are comparable.

The uniqueness of entropy implies that it can be measured conveniently without using the original formula that established its existence. This answers the objections of Crowell and Tourin; you can find the difference of entropy between two states in the old-fashioned way by measuring specific heats, compressibilities, and so on. The uniqueness guarantees that all experiments will answer the question in the same way. Once we have the existence and uniqueness of entropy, all the techniques of the mechanical and chemical engineers mentioned by Tourin are at our disposal.

We disagree with Crowell's comment that one can find the entropy of a macroscopic system by cutting it into tiny subsystems. If, by this, he means that reduction to submesoscopic sizes will simplify thermodynamics, we have to demur. Clearly, additivity and scaling do not hold down to the atomic level because of surface effects. Even if, at some sufficiently small size, entropy can be computed from Boltzmann's formula, the law of its temporal increase may have to be replaced by an as yet unknown version of the law, perhaps

a statistical one. This is a fascinating and largely unexplored area; see refs. 3 and 4 for additional remarks.

"Extensivity," or additivity, can break down on large scales as well, in the presence of gravitational interactions, as mentioned by Landsberg. This is correct, well known, and worthy of emphasis. However, it lies outside the realm of the laboratory physics under discussion.

Nussbaum's remark about the Gibbs-Boltzmann approach being "the wrong direction" could be misinterpreted to imply that we have an anti-Gibbs-Boltzmann bias. This would be unfortunate since we do value statistical mechanics as does any physicist. We said, and we maintain, that the second law, as understood for equilibrium states of macroscopic systems, does not require statistical mechanics, or any other particular mechanics, for its existence. It does require certain properties of macroscopic systems, and statistical mechanics is one model that, hopefully, can give those properties, such as irreversibility. One should not confuse the existence, importance, and usefulness of the Boltzmann-Gibbs-Maxwell theory with its necessity on the macroscopic level as far as the second law is concerned. Another way to make the point is this: If the statistical mechanics of atoms is essential for the second law, then that law must imply something about atoms and their dynamics. Does the second law prove the existence of atoms in the way that light scattering, for example, tells us what Avogadro's number has to be? Does the law distinguish between classical and quantum mechanics? The answer to these and similar questions is "No," and if there were a direct connection, the late 19th-century wars about the existence of atoms would have been won much sooner. Alas, there is no such direct connection that we are aware of, despite the many examples in which atomic constants make an appearance at the macroscopic level—Planck's radiation formula, the Sackur-Tetrode equation, stability of matter with Coulomb forces, and so on. The second law, however, is not such an example.

As our title read, this is a fresh look at the second law of thermodynamics. It is no more obligatory than any other approach, but happily some readers and colleagues have found it useful.

continued on page 106

References

1. E. H. Lieb, J. Yngvason, *Phys. Rep.* **310**, 1 (1999). Also available at <http://xxx.arXiv.org/abs/cond-mat/9708200>. See especially Section 1B.
2. J. Uffink, in *Studies in History and Philosophy of Modern Physics*, Pergamon, Exeter, UK (in press). Also available at <http://xxx.arXiv.org/abs/cond-mat/0005327>.
3. E. H. Lieb, *Physica A* **263**, 491 (1999).
4. E. H. Lieb, J. Yngvason, in *Visions in Mathematics, Towards 2000*, GAFA, Geom. Funct. Anal. Special Volume (1) 334 (in press). Also available at http://www.ma.utexas.edu/mp_arc-bin/mpa?yn=00-332.

ELLIOTT H. LIEB
(lieb@princeton.edu)
Princeton University
Princeton, New Jersey

JAKOB YNGVASON
(yngvason@thor.thp.univie.ac.at)
University of Vienna
Vienna, Austria

Optimal Vision: Blurring and Aliasing

In his article "Retinal Imaging and Vision at the Frontiers of Adaptive Optics" (*PHYSICS TODAY*, January, page 31), Donald T. Miller shows that it is possible to improve the resolution, contrast, and clarity of retinal images by correcting for defects in the eye's optics. We agree that "the best retinal image quality is obtained with the largest physiological pupil diameter (8 mm) and with full correction of all ocular aberrations." However, we disagree with Miller's suggestion that the quality of vision may be improved similarly (to achieve "supernormal vision") if the eye's optics could be "corrected" with "adaptive optics" to produce the performance of an aberration-free 8-mm lens.

The angular spacing between retinal photoreceptors, as Miller states, "represents a neural limitation to visual resolution." In terms of communication theory, this spacing determines the sampling passband of the eye that, analogous to the bandwidth of a communication channel, sets an upper bound on the highest spatial frequencies that the eye can convey to the higher levels of the brain. The preferred modulation transfer function (MTF)—or spatial frequency response—of the eye's optics relative to this sampling passband is inescapably a compromise between blurring and aliasing. Because the MTF decreases smoothly with increasing frequency, aliasing can be substantially decreased only at

the cost of blurring and vice versa.

If blurring and aliasing are properly accounted for in terms of their effect on the information rate that the eye conveys to the higher levels of the brain, then it is the MTF of the 3-mm lens rather than that of the 8-mm lens that, in normal daylight, maximizes this rate for the 50 cycles/degree sampling passband of the eye.¹ Hence, communication theory and evolution converge, under appropriate conditions, toward the same optical design. And why not? It seems unlikely that evolution would have missed the opportunity to improve our vision if it could have done so merely by permitting the pupil to be wider than 3 mm during normal daylight conditions.

Reference

1. F. O. Huck, C. L. Fales, D. J. Jobson, Z. Rahman, *Opt. Eng.* **34**, 795 (1995). F. O. Huck, C. L. Fales, Z. Rahman, *Philos. Trans. Roy. Soc. London*, **A354**, 2193-2248 (1996). F. O. Huck, C. L. Fales, Z. Rahman, *Visual Communication: An Information Theory Approach*, Kluwer Academic, Norwell, Mass., 1997.

FRIEDRICH O. HUCK
(f.o.huck@larc.nasa.gov)

CARL L. FALES
NASA Langley Research Center
Hampton, Virginia

MILLER REPLIES: Friedrich Huck and Carl Fales raise valid concerns about realizing supernormal vision. These concerns, however, are also expressed in my article. The neural system will ultimately limit the degree of supernormal vision that may be achieved after the aberrations in the eye are corrected. In my article I state, "In an eye with perfect optics, visual performance becomes constrained by neural factors, specifically the spacing between retinal photoreceptors, which represents a neural limitation to visual resolution that is only slightly higher than the normal optical limit." I go on to say that this would lead to aliasing, which would degrade vision.

Optimal vision then becomes a compromise between blurring and aliasing. But what constitutes optimal vision and what compromise is appropriate for achieving it? These are difficult questions that the vision community continues to address. Currently, our understanding of the limits placed on vision by the retina and visual pathways of the brain are not sufficient to provide universal answers to these questions. The search for answers is further complicated because visual performance is heavily task-dependent. Visual per-

formance for some specialized tasks will probably decline with adaptive optics. For example, when observers viewed a steady point source through adaptive optics, it sometimes appeared green and sometimes red, depending on which photoreceptor type the light was stimulating. With more natural stimuli, however, subjects have regularly experienced a strikingly crisp appearance consistent with the supernormal quality of the retinal image. For everyday vision, the penalty of aliasing may be outweighed by the reward of heightened contrast sensitivity and detection acuity.

Huck and Fales's application of communication theory unfortunately relies on the superficial analogy of the eye as an electronic video camera. It ignores much of the neural processing of the image and does not take into account the type of visual task. A rigorous application of this theory would require a deeper understanding of the visual system than we presently have. It is perhaps for these reasons that the approach of Huck and Fales does not predict the enhanced vision already experienced with adaptive optics. Ultimately, the extent to which vision will be improved by correcting ocular aberrations will be determined in the laboratory.

DONALD T. MILLER
(dtmiller@indiana.edu)
Indiana University, Bloomington

Moore's Law and the Future of Computing

Joel Birnbaum and R. Stanley Williams of Hewlett-Packard comment on Moore's law (*PHYSICS TODAY*, January, page 38) and discuss the projections of its theoretically anticipated validity until 2012, or even 2020. However, it is more a matter of practicable engineering and technology than it is of theoretical limits of the physical theory.

Interestingly, the researchers of Intel Corp see the whole development rather more pessimistically. For instance, David Papworth, Intel Fellow, suggests that Moore's law won't survive beyond 2004-5. He noted at a VLSi Circuits Symposium in Hawaii in 1998 that by using two or three times as many transistors to increase performance 1.8 times, progress continued apace but at cost: Power consumption has doubled or tripled in each generation. Papworth concluded that, after 2004, density increases will slow down.¹

His colleague Paul Packan, com-