Old versus new paradigm simplicity

P. Fraundorf
*Physics & Astronomy/Center for NanoScience, U. Missouri-StL (63121) and Physics, Washington University (63110), St. Louis, MO, USA**

(Dated: June 24, 2011)

Established idea-sets may not update seamlessly. The tension between new and old views of nature is documented even in Galileo's dialogs, and today is present in many fields. One way to reconcile these perspectives is to consider the algorithmic simplicity of paths from *various* starting points to one goal. We illustrate with a look at two such simplifications: The move from **Lorentz-transform to metric-equation** descriptions of space-time, and the move from **classical to statistical thermodynamics** with help from Boltzmann's choice-multiplicity & Shannon's uncertainty. Connections of the latter to **correlation measures** behind available work, model selection, and layered complexity are also explored. This approach may help to constructively channel tension in other areas too.

PACS numbers: 05.70.Ce, 02.50.Wp, 75.10.Hk, 01.55.+b

Contents

I. INTRODUCTION

Evolution of well-worn approaches naturally encounters resistance from experts in the old[1]. For instance Martin Gardner in his book on parity inversion[2] cites Hermann Kolbe's negative reaction to the prediction of carbon's tetrahedral nature by Jacobus van't Hoff (Chemistry's first Nobel Laureate). The hullabaloo[3] about the Nowak et al. paper[4] on models for evolving insect social behavior is a more recent research example, while participants in the content-modernization branch of physics education research (PER) have engaging tales on the education side[5]. For text publishers, however, even funerals may not mark progress since choosers of a course text might understandably like to teach that course the way they learned it, whether they own the strategy or not.

One way to objectively assess new approaches is perhaps to examine the algorithmically-shortest path to quantitative insight from each given starting point. For experts in the old, traditional approaches may be algorithmically-shortest even if they are not shortest for newcomers to a given subject. Differing perceptions, in this context, might thus be put onto a rational footing.

II. PRINCIPLES

From a given starting point, the strategy for putting together a concept map may be to minimize the number of: (i) assumptions and (ii) new concepts needed to make a given set of quantitative predictions possible. Drawings of such paths from various starting points, in this context, might inspire one to evolve one's own starting point in teaching a given class over time.

Note that we are weighing self-consistent approaches for their compactness, portability, and appropriateness in much the same way that different variable-changes in calculus, and coordinate-system choices in analytic geometry, buy more advantage for some tasks and less advantage for others. In that sense, we seek to apply the science of *Bayesian model-selection* to the evolution of what we teach.

Below we illustrate this with a few examples, based on content changes already underway in the evolving physics curriculum. Similar charts for your own approach to a given class, as it relates to the textbook in hand as well as the larger picture, may be worth putting together for sharing with your students and perhaps in electronic collaboration spaces with the larger physics teaching community as well.

III. METRIC-BASED MOTION

The traditional path via *Lorentz transforms*, by using two separate coordinate-frames with their own yardsticks

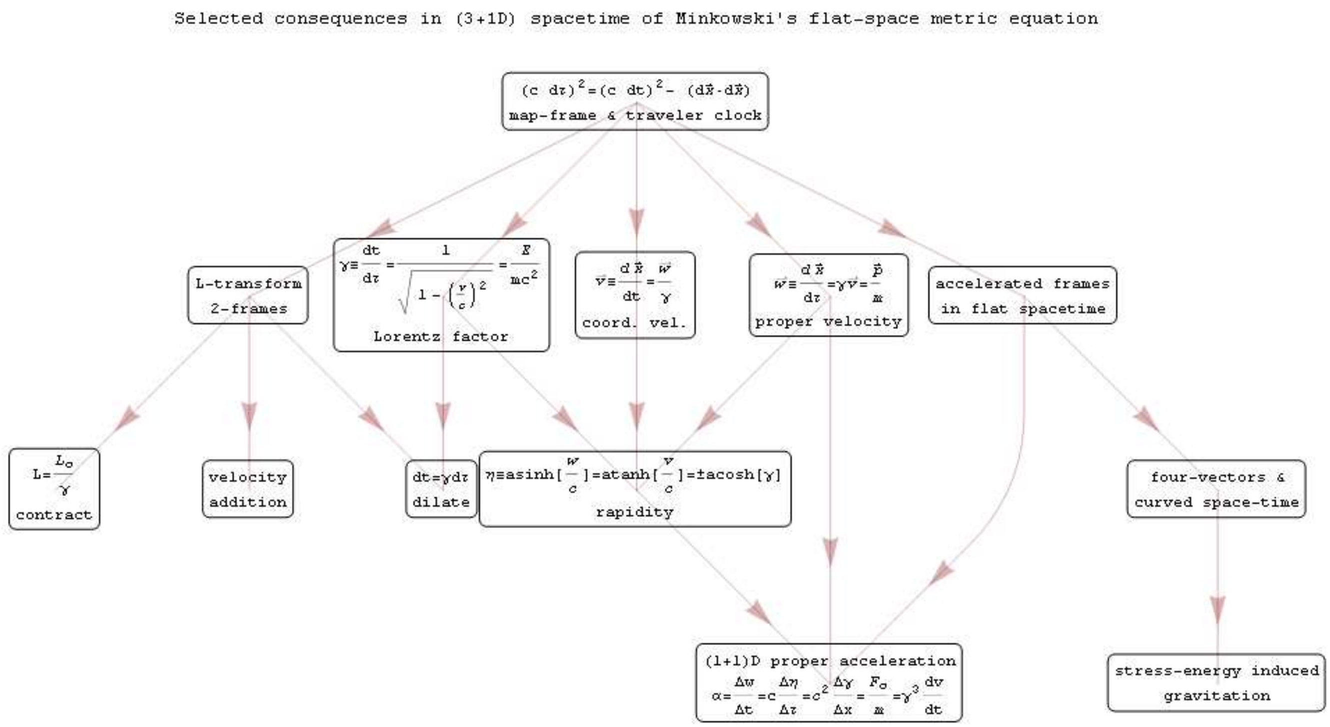

FIG. 1: Some (3+1)D kinematics based on the flat-space metric equation.

and synchronized clocks, likely provides the most direct route to **length contraction** and relativistic **velocity addition**. On the other hand shorter paths to **time dilation**[6], **accelerated motion**[7], and **gravitation**[8] follow by applying the flat-space *metric equation* to a single map-frame of yardsticks & synchronized clocks. The latter are minor issues if one has learned about space-time through Lorentz transforms, and sees relativity as an extension of Newtonian physics to extreme situations.

However if one has been exposed mainly to single map-frame calculations but sees relativity as an explanation for everyday effects like gravitation and magnetism, to which classical physics offers useful low-speed approximation, then the two map-frame approach serves up some cognitive dissonance as well as added complication. The traditional approach for example: (i) emphasizes symmetry between frames even when the home-frame e.g. of a traveling clock or yardstick is quite special, (ii) raises the dissonant spectre[9–12] of relativistic mass, (iii) avoids use of proper-acceleration[13,14] as an integrative complement to geometric-accelerations (affine-connection effects) at low and high speeds, and (iv) misses out on insights that proper-velocity[15,16] offers e.g. into the lightspeed-limit and relativistic velocity-addition. The single map-frame approach avoids these problems, with the result that elements of it are finding their way into texts on all levels.

Some of these path interconnections are illustrated in Figure 1. The layout is designed around a metric equation start, but one can also start with Lorentz transforms since the implication between them runs both ways. However the distance to your favorite proofs as well as to the most interesting applications does depend on the starting point. Although there are other ways to draw the connections and weigh destinations, this map's author might for example argue that the average path to other zones is $20/12 \simeq 1.7$ steps with a metric equation start as compared to $25/12 \simeq 2.1$ steps for a Lorentz transform start.

As teachers we should probably choose a path through space-time for students which draws strength from our prior training and acquired insight into both paths, as well as the path's connection to the past and future of students in each given course. For instance, with introductory physics students it's quite easy to tell students (even if the book doesn't) that time passes differently on different clocks so that, unless otherwise stated, time will be measured on a set of synchronized "map-clocks" affixed to the yardsticks used to measure position.

IV. MULTIPLICITY-BASED THERMODYNAMICS

The question here is: Do I start by introducing temperature in historical units and the zeroeth law while saving entropy to the end, or do I start with choice-multiplicity and entropy so that the assumptions behind the ideal gas law, equipartition, and mass action are explicit from day one? Senior undergraduate texts almost all now do the latter, while only a small number of introductory texts

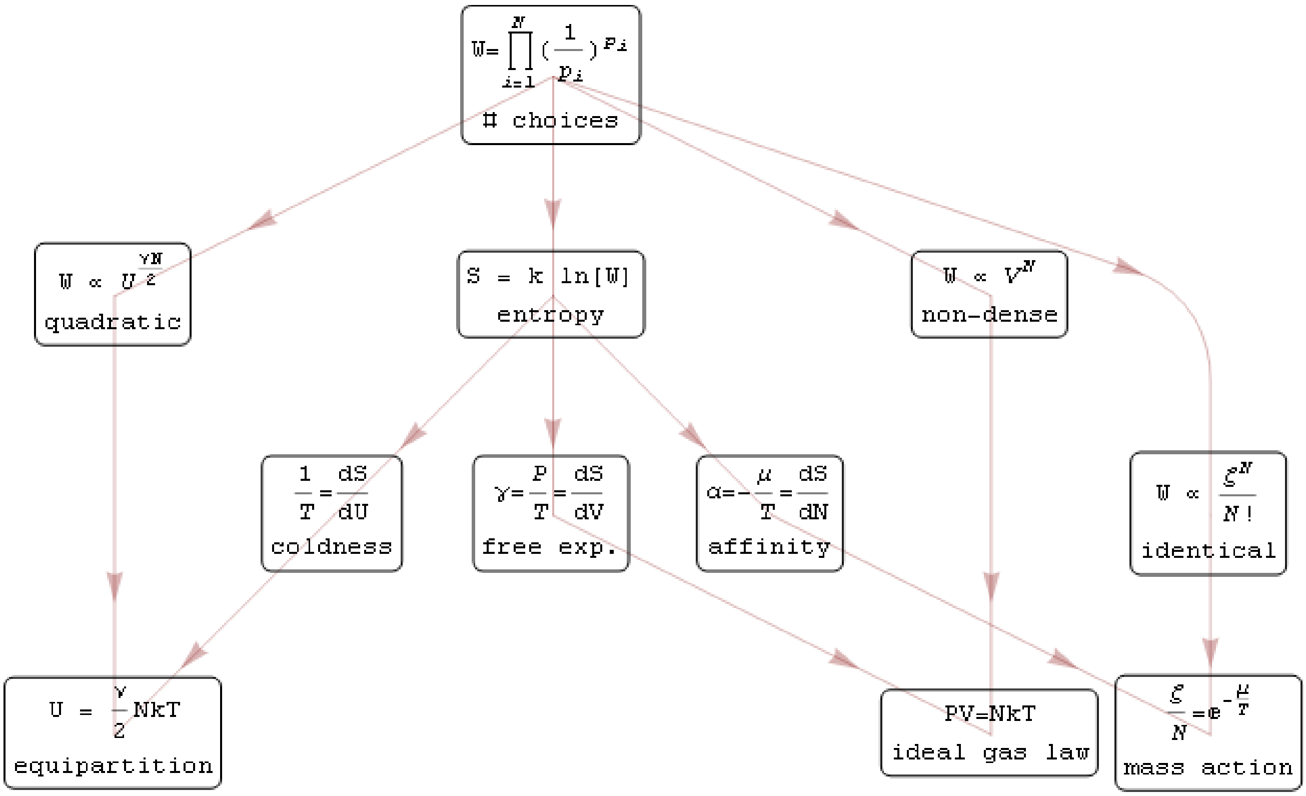

FIG. 2: Choice multiplicity ⇒ gas law, equipartition & mass action.

have made the switch so far.

The simplest axiomatic path to: (a) the ideal gas equation, (b) equipartition, and (c) the law of mass action is likely Boltzmann's choice-multiplicity

$$W = \prod_{i=1}^{N} \left(\frac{1}{p_i}\right)^{p_i}, \tag{1}$$

from which entropy $S = k\ln[W]$ and its derivatives may be defined. This choice-multiplicity, of course, is just the dimensionless count which underlies the familiar use of both information units and Joules per degree Kelvin[12].

Historical approaches introduce these consequences as empirical and/or informally-useful relationships, without clear definition of their underlying mechanism and assumptions and typically with discussions of entropy/multiplicity (the horse) following these consequences (the cart). Such approaches do not provide insight into: (i) the quantitative limitations of these concepts, or (ii) strategies for moving beyond those limitations e.g to systems in which subsystem correlations cannot be ignored.

Another reason to introduce multiplicity first is that the laws of thermodynamics (short of two physical postulates) follow therefrom as well. The zeroth law follows from the fact that the largest number of states is available when the uncertainty slopes of two subsystems (reciprocal temperatures for the energy observable) equilibrate as a conserved quantity is shared.

The first law, oft described as a statement of energy conservation, in fact arises from maximum entropy inference as a relation between ordered and disordered changes in any observable, whether they are conserved on transfer between subsystems or not. Likewise for the second law, whose physics actually comes not from statistical inference but from the assumption that mutual information available on the state of an isolated system will not increase over time.

Finally the very definition of reciprocal temperature as an uncertainty slope will convince many that the change in state-uncertainty about any finite system, per unit change in energy, is likely to be finite. Hence reciprocal-temperature's infinity (the absolute-zero of temperature) is likely inaccessible. This natural definition of temperature has the added advantage that it prohibits one from approaching absolute-zero from negative or positive directions, and shows that the negative absolute-zero approachable e.g. by spin systems with a population inversion is as far away from positive absolute-zero as you can get.

Examples of the power in this recasting of familiar rules include many senior undergraduate thermal physics texts, like those by Kittel & Kroemer[17], Dan Schroeder[18], and Claude Garrod[19] (who refers to reciprocal temperature[20] as *coldness*), Tom Moore & Dan Schroeder's AJP paper[21], Tom Moore's introductory physics Unit T[6], etc.

V. COLLATERAL CONNECTIONS

We've now covered two paradigm-shifts which have a well-defined place in the physics curriculum. The approach taken with respect to them in a given class should inform itself to both teacher & student backgrounds, as well as to *course objectives*. Connection to larger *program objectives* may also figure into the choice of one's algorithmic path. For example, let's look to see if the second paradigm-shift makes contact with other developments of interest to physics students as well.

To do this we step back from uncertainties to probability measures, and then forward from uncertainties to correlation measures, to show how the second simplification also allows physics to make contact with a number of other lively disciplines. Because of the physics in between, out-of-discipline students may never hear about these connections if they aren't at least mentioned in one of their physics classes.

A. surprisals

Recall that information units can be introduced by the statement that # choices equals $2^{\#\text{bits}}$. Also very small probabilities p can be put into everyday terms as the **surprisal**[22] $s = n$ bits of tossing n coins all heads up since $p = 1/2^{\#\text{bits}}$, with the added advantage that surprisals add whenever their probabilities multiply. Evidence in bits[23] for a true-false proposition can similarly be written as $e[p] = s[1-p] - s[p]$, where surprisal is $s[p] = \ln_2[1/p]$.

All of these applications rely on the fact that probabilities between 0 and 1 can be written as multiplicities $w_p = 1/p$ between 1 and $+\infty$ or as surprisals between 0 and $+\infty$ using information units determined by the constant k in the expression $s_p = k\ln[1/p]$. This surprisal ⇔ multiplicity ⇔ probability inter-conversion is summarized by:

$$0 \leq s_{\text{p}} \equiv k \ln\left[w_{\text{p}}\right] \equiv k \ln\left[\frac{1}{p}\right] \leq \infty \tag{2}$$

where of course the units are bits if $k = 1/\ln[2]$.

B. average surprisals

The treatments of the ideal gas law, equipartition, mass action, and the laws of thermodynamics in the previous section connect to this tradition by defining uncertainty or entropy S as an **average surprisal** e.g. in J/K between 0 and $+\infty$, Boltzmann's multiplicity W between 1 and $+\infty$ as $e^{S/k}$ where k is Boltzmann's constant, and 1/W as a reciprocal-multiplicity between 0 and 1. Their relevance to the thermal side of physics education has been discussed above.

More generally the interconversion for the average surprisal, uncertainty, or entropy associated with predicted probability-set q, as measured by operating probability-set p, can be written:

$$0 \leq S_{\text{p/p}} \leq S_{\text{q/p}} \equiv k \ln\left[W_{\text{q/p}}\right] \equiv k \sum_{i=1}^{N} p_i \ln\left[\frac{1}{q_i}\right] \leq \infty. \tag{3}$$

Although written for a discrete probability-set, the expression is naturally adapted to continuous as well as quantum mechanical probability-sets[24,25]. In this context natural as distinct from historical units for temperature become energy per unit information, and for heat capacity become bits[26].

Note that the upper limit on $S_{\text{p/p}}$ is $\ln_2[N]$. Also the fact that $S_{\text{q/p}} \leq S_{\text{p/p}}$, i.e. that **measurements using the wrong model q are always likely to be more surprised by observational data than those using the operating-model p**, underlies maximum-likelyhood curve-fitting and Bayesian model-selection as well as the positivity of the correlation and thermodynamic availability measures discussed below.

C. net surprisals

The tracking of subsystem correlations has taken a back seat in traditional thermodynamic use of log-probability measures. This is illustrated e.g. by the traditional treatment of subsystem entropies as additive, in effect promising that correlations between e.g. between gas atoms in two volumes separated by a barrier can be safely ignored. More generally, however, subsystem correlations (e.g. between a sent and a received message, or between traits of a parent and of a child) are of central importance. In fact the maximum entropy discussed above is nothing more than minimum KL-divergence with a uniform prior[27], so that physicists expert in its application to analog systems can play a pivotal role informing students who take physics courses about these connections across disciplines.

In particular the foregoing are backdrop to the paradigm-shift which broke out of physics into the wide world of statistical inference in the mid-20th century[28]. We'll touch on only three of the many areas that it's connecting together today, based on their relevance to cross-disciplinary interests of students in physics classes. The specific application areas are: (i) **thermodynamic availability**, (ii) algorithmic **model selection**, and (iii) the **evolution of complexity**. The surprisal ⇔ multiplicity ⇔ probability interconversion for these correlation analyses may be written:

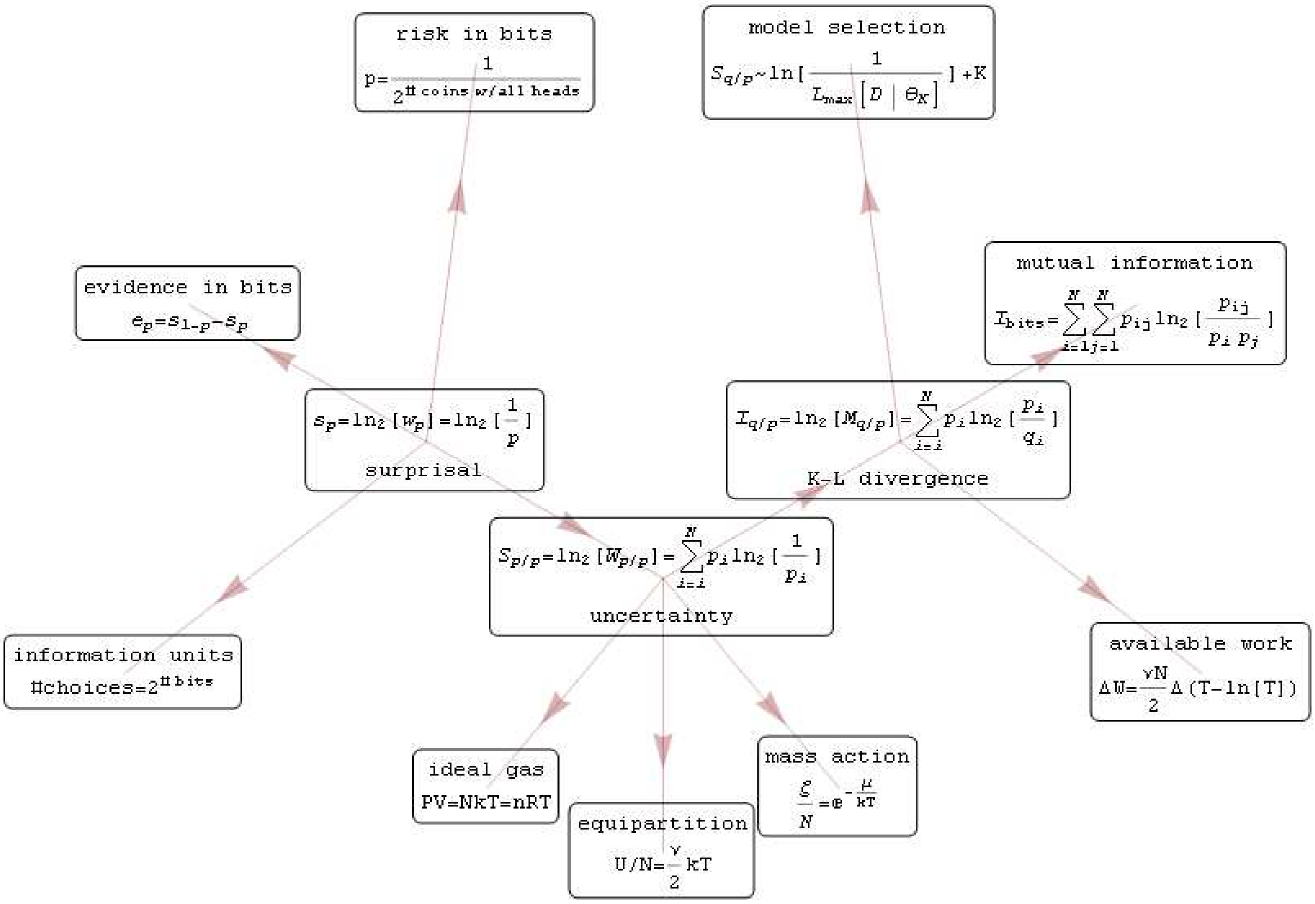

FIG. 3: Cross-disciplinary applications for log-probability measures in statistical inference.

$$0 \leq I_{q/p} \equiv k \ln \left[M_{q/p} \right] \equiv k \sum_{i=1}^{N} p_i \ln \left[\frac{p_i}{q_i} \right] \leq \infty \qquad (4)$$

Log-probability measures are useful for tracking subsystem-correlations in digital as well in analog complex systems. In particular tools based on Kullback-Leibler divergence $I_{q/p} \geq 0$ and the matchup-multiplicities $M_{q/p}$ associated with reference probability-set q have proven useful: (i) to engineers for measuring available-work or *exergy* in thermodynamic systems[29], (ii) to communication scientists and geneticists for studies of: relatedness[30], network structure, & replication fidelity[31,32], and (iii) to behavioral ecologists wanting to select from a set of simple-models the one which is least surprised by experimental data[33,34] from a complex-reality.

In context of this idea-set, the logical schematic in Figure 3 illustrates connections that often go unmentioned between what are now-classical application areas in their specialized fields. It thus suggests that physicists, particularly thanks to their long experience with log-probability measures in analog systems, can play a key role in the cross-disciplinary application of informatics to complex systems.

These multi-moment correlation-measures have 2nd law teeth making them relevant to quantum computing[35], and they enable one to distinguish pair from higher-order correlations making them relevant to the exploration of order-emergence in a wide range of biological systems[36,37]. They may be especially useful in addressing challenges associated with the sustainability of multi-layer complex systems[38].

VI. DISCUSSION

Similar analyses might also help *each of us* decide when it is (and is not) appropriate to spend time in the educational arena e.g. on: (i) geometric-algebra approaches[39–41] to complex numbers & cross-products, (ii) energy[6] & least-action[42] based introductions to mechanics, (iii) vector potential introductions to magnetism[43], (iv) explore-all-paths introductions to quantum mechanics[44]. The approach may even come in handy for mediating differences in research strategy as well, e.g. in deciding how much time to spend (in context of a particular problem) on: (a) CPT approaches to the application of non-Hermitian Hamiltonians[45], (b) molecule-code as distinct from kin-selection models of evolving eusocial or altruistic behavior[4], etc.

VII. CONCLUSIONS

In short both quantitative and schematic considerations of the algorithmic path to key deliverables from **your & your audience's** conceptual starting point may help point you toward approaches that help your students become maximally-informed in minimum time. These may not lessen "the detailed work of content modernization"[5], but they may help provide the process with useful direction taylored to our individual points of reference. What would your concept maps look like in this context?

Acknowledgments

I would like to thank graduate students Bob Collins, Zak Jost, and Pat Sheehan for their participation in a Bayesian informatics development course which helped to nail down some of the inter-connections discussed here.

* pfraundorf@umsl.edu

1 T. S. Kuhn, *The structure of scientific revolutions* (U. Chicago Press, 1970), 2nd ed.
2 M. Gardner, *The New Ambidextrous Universe* (W. H. Freeman, 1990).
3 M. Doebeli, Nature **467**, 661 (2010).
4 M. A. Nowak, C. A. Tarnita, and E. O. Wilson, Nature **466**, 1057 (2010).
5 *How can 20th century simplifications benefit introductory classes?*, AAPT undergraduate education committee summer crackerbarrel (1999), cf. Tom Moore's reference to "the detail work" of content modernization, URL `http://www.umsl.edu/~fraundorfp/aapt99x3.html`.
6 Moore, *Six ideas that shaped physics* (McGraw-Hill, NY, 2003), 2nd ed.
7 P. Fraundorf, *A one-map two-clock approach to teaching relativity in introductory physics*, arXiv:physics/9611011v3[physics.ed-ph] (1996), URL `http://arxiv.org/abs/physics/9611011`.
8 E. Taylor and J. A. Wheeler, *Exploring black holes* (Addison Wesley Longman, 2001), 1st ed.
9 C. G. Adler, Am. J. Phys. **55**, 739 (1987).
10 L. B. Okun, Am. J. Phys. **77**, 430 (2009).
11 T. R. Sandin, Am. J. Phys. **59**, 1032 (1991).
12 G. Oas, *On the use of relativsitic mass in various published works*, arXiv:physics/0504111v1 [physics.ed-ph] (2009), URL `http://arxiv.org/abs/physics/0504111`.
13 A. P. French, *Special relativity*, The M.I.T. Introductory Physics Series (W. W. Norton, New York, 1968), paeg 154: "...acceleration is a quantity of limited and questionable value in special relativity.".
14 T.-P. Cheng, *Relativity, gravitation and cosmology* (Ox, 2005), page 6: in special relativity... "we are still restricted to" ... "inertial frames of reference" and hence no acceleration.
15 F. W. Sears and R. W. Brehme, *Introduction to the theory of relativity* (Addison-Wesley, NY, New York, 1968), section 7-3.
16 W. A. Shurcliff, *Special relativity: The central ideas* (1996), 19 Appleton St, Cambridge MA 02138.
17 C. Kittel and H. Kroemer, *Thermal Physics* (W. H. Freeman, New York, 1980), p. 39.
18 D. V. Schroeder, *An Introduction to Thermal Physics* (Addison-Wesley, San Francisco, 2000).
19 C. Garrod, *Statistical Mechanics and Thermodynamics* (Oxford University Press, New York, 1995).
20 J. Castle, W. Emmenish, R. Henkes, R. Miller, and J. Rayne, *Science by degrees: Temperature from Zero to Zero*, Westinghouse Search Book Series (Walker and Company, New York, 1965).
21 T. A. Moore and D. V. Schroeder, Am. J. Phys. **68**, 1159 (2000).
22 M. Tribus, *Thermostatics and thermodynamics* (D. Van Nostrand Co., Princeton, 1961).
23 E. T. Jaynes and G. L. Bretthorst, *Probability theory: The logic of science* (Cambridge U. Press, 2003).
24 E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
25 E. T. Jaynes, Phys. Rev. **108**, 171 (1957).
26 P. Fraundorf, Am. J. Phys. **71**(11), 1142 (2003).
27 P. C. Gregory, *Bayesian logical data analysis for the physical sciences: A comparative approach with Mathematica support* (Cambridge U. Press, Cambridge UK, 2005).
28 C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, IL, 1949).
29 M. Tribus and E. C. McIrvine, Scientific American **224**(3), 179 (1971).
30 C. H. Bennett, M. Li, and B. Ma, Scientific American **288**(6), 76 (2003).
31 T. M. Cover and J. A. Thomas, *Elements of information theory* (John Wiley and Sons, NY, 2006), 2nd ed.
32 A.-L. Barabasi, *Linked: How everything is connected to everything else and what it means for business, science and everyday life* (Plume/Penguin, 2003).
33 K. P. Burnham and D. R. Anderson, Wildlife Research **28**, 111 (2001).
34 K. P. Burnham and D. R. Anderson, *Model selection and multimodel inference: A practical information-theoretic approach* (Springer Science, NY, 2002), 2nd ed.
35 S. Lloyd, Physical Review A **39**(10), 5378 (1989).
36 E. Schneidman, S. Still, M. J. Berry II, and W. Bialek, Phys. Rev. Lett. **91**, 238701 (2003).
37 M. Harper, *Information geometry and evolutionary game theory*, arXiv:0911.1383v1 [cs.IT] (2009), URL `http://arxiv.org/abs/0911.1383`.
38 P. Fraundorf, Complexity **13**(6), 29 (2008).
39 D. Hestenes, Am. J. Phys. **71**, 105 (2003).
40 D. Hestenes, Am. J. Phys. **71**(7), 691 (2003).
41 C. J. L. Doran and A. N. Lasenby, *Geometric algebra for physicists* (Cambridge U. Press, 2003).

[42] T. A. Moore, Am. J. Phys. **72**(4), 522 (2004).
[43] E. J. Konopinski, Am. J. Phys. **46**(5), 499 (1978).
[44] E. F. Taylor, S. Vokos, J. M. O'Meara, and N. S. Thornber, Computers in Physics **12**, 190 (1998).
[45] C. M. Bender, Reports on Progress in Physics **70**, 947 (2007).